On the quantification of complexity and diversity from phenotypes to ecosystems

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Zachary Harrison Marion

December 2016

To Mom, Dad, and Loki—the best dog in the world.

Acknowledgements

I am greatly indebted to the many individuals who have directly and indirectly contributed to my growth as a person—both professionally and personally—during my tenure at the University of Tennessee. My advisor, Ben Fitzpatrick, has been invaluable in helping me to think critically about all aspects of my research, even in the face of overwhelming frustration. I am indebted to him for taking a chance on me after a tumultuous stent at Georgia Tech. I would also like to thank him for the attention and interest he has given me and my work despite very little overlap in our research interests.

I am also as much—if not more—indebted to Jim Fordyce who, without question, has influenced my personal and scientific life more than anyone else. From the moment I harassed him into hiring me to do undergraduate research (despite having not completed his Ph.D. at the time), Jim has taught me how to be a scientist. I have learned more over beers with him than any class, seminar, or workshop I could ever have taken. Not only is he a great collaborator, Jimmy is also one of my closest friends. He has been influential in my post-doc offer with Matt Forister as well as my upcoming job as a lecturer on Bayesian statistics. Jim is responsible for teaching me about the perks and pitfalls of rabbit-holes, and both he and Ben indulged me as a statistical rabbit-hole took over my dissertation post-candidacy. Both also taught me the valuable lesson of not letting the perfect get in the way of the good.

I would like to thank my committee members Nathan Sanders and Shawn Campagna for their support throughout my doctoral career. Also thanks are in order for Brian O'Meara and Susan Kalisz. I am extremely grateful for the generous financial support provided by the EEB

department and an NSF doctoral dissertation improvement grant (DEB-1405887) that funded my research.

Personal thanks are in order for all of my friends, collaborators, and colleagues who made life bearable these last few years. I am indebted to (in no particular order) Sharon Clemmensen, Sam Borstein, Lacy Chick, Leigh Moorhead, Rachel Fovargue, Rachel Wooliver, Tyson Paulson, Chelsea Miller, Hailee Korotkin, Todd Pearson, Liam Muller, Jeremy Beaulieu, Katie Kulhavy, Christopher Peterson, and so many more. Thanks also to Bert White and Prom for keeping me sane.

Finally, I would like to say thanks to my parents for supporting me and encouraging me all these years. And thanks to my dog Loki, the best field and officemate a man could have!

"[Perfection] is the enemy of the good."

–Orlando Pescetti

"So...newts"

–A very wise man

Abstract

A cornerstone of ecology and evolution is comparing and explaining the complexity of natural systems, be they genomes, phenotypes, communities, or entire ecosystems. These comparisons and explanations then beget questions about how complexity should be quantified in theory and estimated in practice. Here I embrace diversity partitioning using Hill or effective numbers to move the empirical side of the field regarding the quantification of biological complexity.

First, at the level of phenotypes, I show that traditional multivariate analyses ignore individual complexity and provide relatively abstract representations of variation among individuals. I then suggest using well-known diversity indices from community ecology to describe phenotypic complexity as the diversity of distinct subsidiary components of a trait. I show how total trait diversity can be partitioned into within-individual complexity (alpha diversity) and between-individual components (beta diversity) within a hierarchical framework.

Second, I use simulations to demonstrate that naïve measures of standardized beta diversity such as turnover or local/regional dissimilarity are biased estimators when the number of sampled units (e.g., quadrats) is less than the "true" number of communities in a system (if it exists). I then propose using average pairwise dissimilarities and show that this measure is unbiased regardless of the number of sample units. Moreover, the measure is intuitively interpreted as the average proportional change in composition as one moves from one sample to the next.

Finally, I apply a hierarchical Bayesian approach to the estimation of species abundances within and among samples, communities, or regions. This strategy accommodates difficult problems of bias and uncertainty in the estimation of the diversity of the underlying communities while providing integrated estimates of uncertainty. Moreover, multilevel hierarchies are possible.

We can then use model comparison to determine whether patches/communities/habitats within regions or sets are distinct subcommunities of a metapopulation, or whether they are "arbitrary" distinctions from one contiguous system.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Ecologists and evolutionary biologists have always been focused on the complexity and diversity of natural systems (e.g., genomes, phenotypes, communities, or ecosystems) for a myriad of reasons. Many classical theories have concerned species richness (MacArthur and Wilson, 1967; Pianka, 1974) or the abundance distribution of species (Simpson, 1949; Preston, 1962*a*). Other ecological questions have revolved around temporal changes in diversity (Menge and Sutherland, 1976; Stegen et al., 2013), the relationships between complexity and environment (Loreau et al., 2001; Ricklefs, 2006; Chao et al., 2016), or the dependence of diversity on area (MacArthur and Wilson, 1967). With the onset of the -omics revolution and "big data", key questions are now being asked about the complexity and diversity of microbial ecosystems, complex phenotypic traits, or interaction networks (e.g, Torsvik and Øvreås, 2002; Ihmels et al., 2007; Ings et al., 2009; Marion et al., 2015*c*). However, relating the concept of diversity to measurable quantities has been challenging. There has been some disagreement about what fundamental parameters should be studied (e.g., Ellison, 2010), and there has been very little attention to practical challenges of estimating and comparing parameters using samples from the real world.

1.1 The importance of effective numbers

Over the last several years, one of the greatest advances in the quantification of diversity has been the mathematical unification of many commonly used diversity indices through Hill or effective

numbers. Originally described by Hill (1973), they were reintroduced to ecologists when Jost (2006) derived strong mathematical proofs for the measures. These equations transform measures of entropy or probability of identity into metrics that are easily interpretable as the equivalent number of species present in a community if all species had equal abundances. Specifically, diversity metrics such as richness, Shannon's entropy (Shannon, 1948), or Simpson's measure of identity (Simpson, 1949) relate to each other through a parameter q, known as the diversity order. Where the different metrics differ is in their treatment of unevenness. When $q = 0$, only presence or absence is important; when $q = 1$, species are weighted equally by their relative abundances. And when $q \geq 2$, rare species begin to be downweighted in importance (Jost, 2006; Chao et al., 2012). Not only did Hill numbers unite these seemingly disparate indices, they provided an intuitive framework to both describe the number of species and their relative abundance distribution.

Since their reintroduction by Jost (2006), effective numbers have been rapidly adopted by ecologists (Ellison, 2010) when partitioning diversity into within (α) vs among-community (β) components (Jost, 2007). Moreover, they have been extended and improved on. For example, effective numbers have been extended to account for phylogenetic or functional trait relatedness (Chao et al., 2014), and recently they have been unified (Chao and Chiu, 2016) with a variance partitioning approach for beta diversity advocated by Legendre and De Cáceres (2013).

1.2 Practical estimation of diversity

Although the theoretical underpinnings of effective numbers and diversity in general has been compelling, practical implementation and analysis of diversity using real data have been challenging. One of the biggest challenges concerns the distinction between describing the data and describing the "true" underlying ecological system of interest. This is because empiricists almost always use samples (e.g., quadrats, transects) to describe the real world, and samples have uncertainty. Because of the limitations of time and money, ecologists will rarely account for every individual in a community. Thus samples will usually underestimate the true diversity of a system (in part because rare species are missed (Gotelli and Colwell, 2001), and in part because samples

tend to be more heterogeneous—i.e., less even—than the underlying true distribution). The problem of rarity and sampling is not unique to ecology (Paninski, 2003; Hausser and Strimmer, 2009), but will be particularly insidious for many fundamental questions and theories in which ecologists are interested.

For example, there is a well-known latitudinal gradient in diversity as one moves from the species-poor poles to the species-rich tropics (Pianka, 1974; Kraft et al., 2011). However, in the tropics, sampling effects are often magnified because there are more species overall there, and many of them will be of low abundance at any one site. Thus, not only are more rare species missed for any one sample, but it will take more time and effort to collect an equivalent number of samples because processing any one sample takes more time. Unfortunately, despite these potential sources of biases, many ecologists ignore the uncertainty in sampling, or use corrections that are post-hoc and describe the data rather than the underlying system (Gotelli and Colwell, 2001; Chao et al., 2013, 2015, 2016). When making comparisons among systems (which is what ecologists often want to do), this adds a level of incommensurability that is unappreciated.

My dissertation advances the field of diversity quantification by providing new avenues of quantification from phenotypes to ecosystems. First, I show how the phenotypic complexity of an organism can be conceptualized as a component of diversity, providing an intuitive way to analyze phenotypic diversity and differentiation. Second, I show how standardized estimates of heterogeneity among communities (i.e., beta diversity) are confounded with sample size and then provide an intuitive solution in the form of average pairwise turnover. And last, I show how hierarchical Bayesian models can be used to model the relative abundances of species within and across samples, and then use those relative abundances to decompose the diversity of the "true" underlying ecological system. Moreover, this framework provides an integrated avenue for making comparisons about differences between and among multiple assemblages.

1.3 Phenotypic complexity as diversity

Many, if not most, components of an organism's phenotype are better viewed as the expression of multiple correlated traits, rather than independent attributes. These components function as

multivariate complexes, where the subsidiary parts operate and evolve together. Examples of such phenotypes include behavioral syndromes (Sih et al., 2004), biomechanical systems (Wainwright, 2007), chemical defenses (Jones and Firn, 1991), and metabolic pathways (Hamberger and Bak, 2013). Phenotypic complexity might characterize individual cells, organ systems, organisms, social groups, or ecosystems. In chapter 2, I apply diversity partitioning—an approach traditionally used to describe community complexity—as a general framework for the quantification of of phenotypic complexity.

I first discuss some of the more traditional avenues of analysis for such data. These include collapsing the abundances into one total amount or abundance, ANOVA and MANOVA approaches, ordination, and cluster analysis (Legendre and Legendre, 2012). While these strategies all describe different aspects of the phenotypes in their own way, they often black-box the complexity through their simplification and dimensionality reduction.

Second, I show how we can use Hill numbers from community ecology (Hill, 1973; Jost, 2006; Chao et al., 2014) to describe phenotypic complexity as the effective number of components within an idealized individual. Using profile plots, I illustrate how we can calculate diversity for various orders of q to get an idea of both how many components there are and the abundance distribution comprising the different components. Then, using diversity partitioning (Whittaker, 1960; Jost, 2007), we can then separate the complexity into the within-group component (α)—the average effective number of components found in a typical individual, and the among-group component (β)—the effective number of completely distinct combinations or mixtures of those components found across individuals.

Finally, I provide three case studies to highlight how complexity as diversity provides unique and intuitive descriptions of the data without obscuring some of the more relevent aspects of the complexity. These include multivariate chemical defenses (chemotypes) and extended microbial phenotypes of different tree species. In all cases, new information was gained, suggesting that this method is—at the very least—a useful compliment to many of the more traditional approaches.

1.4 Average pairwise dissimilarity as a solution to the sample size bias of standardized measures of beta diversity

Beta diversity is an important and commonly used metric in ecology to quantify differences in composition among phenotypes, communities, or ecosystems (Whittaker, 1960; Jost, 2007; Kraft et al., 2011; Marion et al., 2015*c*). Because beta diversity is dependent on N, the total number of units within a system, beta is often converted to related indices such as turnover or local or regional differentiation (Jost, 2007; Chao and Chiu, 2016) when making comparisons among systems with different sample sizes.

However, when the N units are samples, problems can arise. First, within a unit, unless all individuals are exhaustively sampled, rare species will be missed, resulting in an underestimate of alpha diversity (Gotelli and Colwell, 2001; Hausser and Strimmer, 2009; Marcon et al., 2015). Second, sampling fewer than the "true" number of units underestimates the total diversity of the system (γ). A third and underappreciated issue is that for most ecological data, discrete sample units such as transects or quadrats do not correspond to natural subdivisions of the system. Instead, usually these systems are continuous environments that are not easily discretized. For these kinds of systems, there is no "true" N and no "true" alpha or beta diversity. While these metrics still have value for ecologists, descriptive statistics such as these will be peculiar to the sampling design unless care is taken.

In chapter 3, I use simple conceptual examples and a simulation study to first show that standardized measures of beta diversity such as turnover or local/regional differentiation (Jost, 2007; Chao and Chiu, 2016)—assumed to remove the effect of N—nonetheless are still biased estimates of sample size. Specifically, I show that N-community measures of dissimilarity—depending on the order q and dissimilarity metric used—tend to over- or underestimate the true diversity of a system when less than the true N units are sampled. The degree of over- or underestimation is highly dependent on the number of samples or survey units.

Second, I show how these same dissimilarity metrics still have utility by using average pairwise versions of these measures. For any effective number, we take the average of a matrix of all pairwise dissimilarities. Again using simulations, I show that these pairwise measures give consistent and unbiased answers regardless of sample size. Moreover, average pairwise dissimilarities are intuitive measures of heterogeneity because they estimate the expected difference in composition between a random pair of survey or sample units. This approach solves the problem of incommensurability across regions or study systems, allowing researchers to ask broad questions without risking erroneous inference arising from sampling design.

1.5 A complete statistical framework for partitioning and testing hypotheses about diversity

Most ecological datasets consist of samples, and as mentioned previously, samples bring uncertainty with them. However, historically this uncertainty has been ignored, and point estimates of diversity are taken as known fixed measurements (e.g., Magurran et al., 2010; Kraft et al., 2011). When uncertainty is addressed, most methods have used bootstrapping in one of two ways (Efron, 1982; Efron and Tibshirani, 1986). In the first, parametric bootstrapping assumes a distribution such as a multinomial to allocate the total number of individuals within a site (Charney and Record, 2012; Marcon et al., 2015). The latter form of bootstrapping resamples sites nonparametrically. Unfortunately, both of these methods assume that attributes of the system are fixed and known quantities: the total number of individuals within a site in the former, and the total number of sites in the latter. In both cases, the uncertainty is limited by the sample, and thus these methods describe the data rather than the underlying system.

In chapter 4, I advocate for a formal statistical framework in the form of a hierarchical Bayesian model. In this method, we model the abundance of each species at each site as multinomially distributed. From this model I get a posterior distribution of the relative abundances of each species at each site $\vec{\theta}$, upon which we can apply diversity partitioning. I also get a vector of hyperparameters ($\vec{\Theta}$) describing the average relative abundance of species across the assemblage.

I also simulate sample data at each iteration using the current posterior $\theta's$, and differentiate between these posterior $\theta's$ and simulated sample data as describing the underlying system vs. describing the data, respectively.

Using simulations, I show that this method does at least as good at as naïve estimates at predicting the true underlying diversities, and sometimes better than bias corrected measures. Moreover, as information increases (i.e., more complete sampling) estimates improve dramatically because information and uncertainty is shared across sites with the prior. For example, rare species missed at one site are partially accounted for if sampled at another site, improving the overall estimation of diversity. I then demonstrate that we can use model selection via WAIC (Watanabe, 2010) or leave-one-out cross validation (Vehtari et al., 2016*b*) to ask whether it is better to model communities as a single assemblage or a metacommunity comprised of distinct subassemblages. Finally, I illustrate the utility of the method with a real-world example of orabatid mite communities distributed across a habitat complexity gradient. Using the integrated Bayesian framework, I am able to reveal novel and biologically relevant patterns in the data that have not been previously discovered.

Together, this framework has three important benefits. First, the posterior parameter estimates describe the abundances of the true underlying assemblage rather than describing the data, independent of sampling design. Second, information sharing implicitly mitigates problems of unequal sample sizes and sample effort. Third and finally, this approach provides natural and intuitive envelopes of uncertainty that propagates across hierarchical levels without information loss. Practically, this means that ecologists now have an integrated statistical framework in which to make comparisons and contrasts across study systems and sampling designs.

Chapter 2

Extending the concept of diversity partitioning to characterize phenotypic complexity

This chapter is a lightly revised version of a paper by the same name published in The American Naturalist and co-authored with James A. Fordyce and Benjamin M. Fitzpatrick.

Marion, Z.H., J.A. Fordyce, & B.M. Fitzpatrick (2015). Extending the concept of diversity partitioning to characterize phenotypic complexity. *The American Naturalist*. 186(3): 348–361.

Abstract

Most components of an organism's phenotype can be viewed as the expression of multiple traits. Many of these traits operate as complexes, where multiple subsidiary parts function and evolve together. As trait complexity increases, so does the challenge of describing complexity in intuitive, biologically meaningful ways. Traditional multivariate analyses ignore the phenomenon of individual complexity and provide relatively abstract representations of variation among individuals. We suggest adopting well-known diversity indices from community ecology to describe phenotypic complexity as the diversity of distinct subsidiary components of a trait. Using a hierarchical framework, we illustrate how total trait diversity can be partitioned into within-individual complexity (alpha diversity) and between-individual components (beta diversity). This approach complements traditional multivariate analyses. The key innovations are (i) addition of individual complexity within the same framework as between-individual variation, and (ii) a group-wise partitioning approach that complements traditional level-wise partitioning of diversity. The complexity-as-diversity approach has potential application in many fields, including physiological ecology, ecological and community genomics, and transcriptomics. We demonstrate the utility of this complexity-as-diversity approach with examples from chemical and microbial ecology. The examples illustrate biologically significant differences in complexity and diversity that standard analyses would not reveal.

2.1 Introduction

Most phenotypic components are comprised of multiple traits collectively operating as an integrated unit (Pigliucci and Preston, 2004; Swallow and Garland, 2005). Phenotypic complexity is an intuitive concept, but can be difficult to define (e.g., Horgan, 1995). "Complex" is generally defined as consisting of interconnected or subordinate parts (Swallow and Garland, 2005); more complex phenotypes have more interconnected parts than less complex phenotypes. Examples of complex phenotypes include behavioral syndromes (Sih et al., 2004), biomechanical systems (Wainwright, 2007), chemical defenses (Jones and Firn, 1991), and metabolic pathways (Hamberger and Bak, 2013). Complex phenotypes might characterize individual cells, organ systems, organisms, social groups, or ecosystems.

Here we apply an existing approach to a new conceptual arena by using the diversity partitioning analyses of community ecology as a general framework for quantifying phenotypic complexity. We illustrate how trait diversity can be partitioned into distinct within- vs. among-group components of variation at multiple hierarchical levels, all the way down to individual complexity. For a given hierarchical level or grouping, each diversity estimate (within-group, among-group, or pooled total) describes a unique aspect of the phenotype in an intuitive and biologically meaningful way. We demonstrate the appeal of the complexity-as-diversity approach with examples of multivariate chemical defenses and of soil microbial traits that are important in nutrient and energy flow.

For some aspects of an organism, it is natural to express complexity as the diversity of subordinate parts. In the interest of clarity, we use "phenotype" to refer to a multivariate set of traits and, generally, not as the total set of all traits comprising an organism (e.g., Lynch and Walsh, 1998). The framework we present here is particularly suited for analyzing complex phenotypes whose constituent traits can be intuitively conceived as counts or amounts. Obvious examples include cell or tissue types, gene expression patterns, ecological metabolomics, chemical defenses, and the proportion of time devoted to various behaviors. We view it as complementary to more traditional multivariate approaches, including those derived from quantitative genetics theory.

2.1.1 Traditional analytical approaches and their shortcomings

As phenotypic complexity increases, so does the challenge of analyzing and summarizing that complexity in intuitive yet biologically relevant ways. For example, many chemically mediated phenotypes consist of several—often disparate—compounds that vary both in molecular identity and quantity (e.g., Howard and Blomquist, 2005; Mithöfer and Boland, 2012). In some cases, hundreds of different molecules of varying concentrations may be contributing to the overall phenotypic, or even ecosystem, function (Tumlinson, 2014). Traditionally, these components of the phenotypes have been analyzed as total amounts or concentrations, thereby ignoring their underlying complexity (see Schofield et al., 2001; Salminen and Karonen, 2011, for reviews).

Another way of assessing complex phenotypes has been to conduct a MANOVA and subsequently use post-hoc univariate tests on each subunit's abundance separately (e.g., Fordyce and Malcolm, 2000). These methods are more nuanced than the simple summation techniques described above, but they have their drawbacks. First, as the number of subsidiary traits comprising a phenotype increases, the statistical power to detect meaningful differences decreases rapidly when one corrects for multiple comparisons (Curran-Everett, 2000). Second, univariate tests ignore covariance among traits, and thus they will not capture the synergistic effects of trait interactions or identify trait redundancy (Rasmann and Agrawal, 2009). Third, and more important, these methods are designed to describe and detect differences in trait means but do little to describe the variance, which often is as much—if not more—of interest to ecologists and evolutionary biologists (Violle et al., 2012).

A common approach used to analyze complex phenotypes is to embrace the multivariate nature of the data and use ordination (e.g., nMDS, PCA, RDA) or cluster analysis. The overall goal is to depict the similarity among objects (e.g., tissue types, individuals, species) in reduced or lower dimensional space (Legendre and Legendre, 2012). Objects with greater similarity are placed closer to each other than objects with less similarity. However, there is a cost of interpretability with these multivariate methods. This is because (dis)similarity metrics remove the direct link between the new components and original variables, making variable contribution more challenging to decipher. Furthermore, these methods begin with dissimilarities or distances

between individuals, sites, or replicates. Therefore they do not address the complexity of individual phenotypes.

The strategies mentioned above are overtly designed to reduce the dimensionality (complexity) of the data. Instead of analyzing phenotypic complexity, the first step is to avoid it, or place it into a black box. Although simplification is almost always a necessary aspect of analysis, simplification can sometimes be achieved in ways that illuminate—rather than obscure—interesting components of complexity. Diversity partitioning is one way to explicitly account for both the qualitative and quantitative aspects of phenotypic complexity and does so while providing easily interpretable and biologically meaningful estimates that describe complexity.

2.1.2 A new conceptual approach: complexity as diversity

Two distinct traditions have approached complexity in a similar way. Studies of the evolution of organismal complexity have used the number of part types to represent complexity at a given level of organization. For example, the number of genes in a genome (Kuo et al., 2009), the number of kinds of intracellular structures (McShea, 2000), and the number of cell types (McShea, 1996; Bell and Mooers, 1997) have been used as general measures of organismal complexity. Likewise, ecologists describe ecosystems with few interacting species as "simple", and those with many interacting species as "complex" (May, 1973; Ings et al., 2009). Both traditions point to a strong conceptual connection between diversity and complexity formalized in information theory (Rényi, 1961; Bonchev, 2003), and ecologists have used this formalism to develop a quantitative approach to species diversity (Hill, 1973).

The principle behind this connection is that we can measure the complexity of a system or structure by the amount of information needed to describe it (Shannon, 1948; Kolmogorov, 1965), just as the diversity of a set can be measured by the amount of information needed to describe its composition. We can characterize a low-diversity system by measuring or identifying just a few of its elements; looking at a new element provides little new information. For a high-diversity system, each new element might provide a lot of new unexpected information about the whole. For example we do not learn much by finding a pine in a pine forest; finding a rare

orchid might add a lot to our description of the community, but it happens rarely. This principle underlies Shannon's H (Shannon, 1948), which is the average information gained by randomly sampling and identifying one item. If all types are rare, each new sample provides a lot of unique information; the average (H) is large. If the system is mostly one type, each new sample is likely to be more of the same; the average information is small.

The same principle applies to organismal phenotypes. Consider the thought experiment of randomly sampling cells, mRNA, or pheromone molecules from an individual. If repeated sampling from the same individual gives the same tissue type or transcript or chemical compound, we conclude that the organism's phenotype is relatively simple. If new samples are often previously unobserved tissue types, transcripts, or compounds, we conclude that the individual's phenotype is relatively complex. For example, Iason et al. (2005) used diversity indices to summarize the diversity and the equitability of terpenes in pine trees. However, they did not comment on the generality of the approach or the potential for hierarchical partitioning.

Here we apply the familiar conceptual framework of community diversity to describe multivariate phenotypes as the diversity—or effective number—of distinct subsidiary traits making up an individual's phenotype. We then illustrate how that diversity can be partitioned into distinct within- vs. among-group components of variation at multiple hierarchical levels. For a given hierarchical level or grouping, each diversity estimate (within-group, among-group, or pooled total) describes a unique aspect of phenotypic complexity in an intuitive and biologically meaningful way. We demonstrate the appeal of the complexity-as-diversity approach with examples of multivariate chemical defenses and of phenotypes of soil microbes associated with plant rhizospheres.

We emphasize that the complexity-as-diversity approach should be viewed as complementary to traditional multivariate methods as each is likely to expose different aspects and patterns of a particular dataset. As described here, the approach is entirely descriptive. We are not inventing any new calculations; rather we are promoting the use of established frameworks to analyze novel kinds of data.

2.1.3 Interpreting effective numbers

The main idea of this paper (complexity as diversity) is not tied to a particular computational program or mathematical representation of diversity. To make the discussion more concrete and provide methodological background for the examples, we outline one specific approach based on methods that have recently become popular in community ecology (Hill, 1973; Jost, 2007) in addition to physics and economics (Rényi, 1961; Patil and Taillie, 1982; Tsallis, 2001).

Ecologists typically quantify species diversity using indices such as richness (the number of species), average information (Shannon entropy), or the probability of identity (Simpson concentration). With simple transformations (Rényi, 1961; Hill, 1973), these indices all express the diversity of a community (D) as numbers equivalents (Patil and Taillie, 1982) or the "effective number" of species (Macarthur, 1965). Numbers equivalents describe the number of equally likely or common components (e.g., expressed genes, chemical compounds, species) needed to obtain a given value of that index. Numbers equivalents have intuitive behaviors and interpretations because they are in biologically relevant units of complexity with "doubling properties" (Hill, 1973). For example, say a chemical ecologist has two equally diverse yet chemically distinct (no compounds shared) samples, each with effective compound number X. If those two samples were combined, the chemical diversity of the combined sample should become $2X$ (Jost, 2007). When diversities are expressed as numbers equivalents, their magnitudes have simple interpretations as the number of equally common phenotypic components. Likewise, a numbers equivalent describing the beta diversity between these two samples should be two, because they are two completely distinct sets of chemicals. Thus, a numbers-equivalent approach immediately provides a biologically interpretable measure of complexity and variation, unlike most multivariate or distance-based approaches.

Many commonly used diversity indices have numbers equivalents; where they differ is in their sensitivity to unequal amounts of the different subunits, indicated by the diversity order q (Hill, 1973; Jost, 2006). The phenotypic diversity (D) of order q when there are K traits (analogous to the S species of communities) is defined for $q \neq 1$ as

$$ {}^{q}D = \left(\sum_{i=1}^{K} p_i^q \right)^{1/(1-q)} \tag{2.1} $$

where p_i is the relative amount of the ith trait (Jost, 2006). The phenotypic diversity is undefined for $q = 1$, but its limit exists and equals the exponential of Shannon's information index:

$$ {}^{1}D = \lim_{q \to 1} {}^{q}D = \exp\left(-\sum_{i=1}^{K} p_i \log p_i \right). \tag{2.2} $$

Diversity of order 0 (richness) is completely insensitive to quantitative variation; only the presence or absence of a phenotypic component is considered. When $q = 1$ (exponential of Shannon's H), each component is weighted by its frequency or relative abundance. For values of $q > 1$ (e.g., inverse Simpson concentration when $q = 2$), more abundant phenotypic components become increasingly influential. Because all of the different orders are in units of effective numbers, we can use the differential sensitivity to unequal abundances to graphically assess phenotypic unevenness as a function of order q (Hill, 1973; Jost et al., 2010).

In diversity profile plots, phenotypes with completely equal abundances of each subsidiary component will have the same effective number for any order q; such a profile appears as a horizontal line in the plot (i.e, example I in Figure 2.1). Sharply descending curves (example III) indicate high unevenness: e.g., chemical phenotypes consisting of a few compounds of high concentration and many of low abundance, or tissues with a few highly expressed genes and several genes with minimal expression. Effective numbers, when combined with diversity profile plots, facilitate intuitive comparisons of differences in the magnitude of qualitative variation and the distribution of quantitative unevenness for complex phenotypes.

Characterizing complexity as diversity provides familiar and intuitive measures based on information theory. When used properly, these can be compared fairly across systems (Jost, 2006, 2007). And, this approach leads naturally to hierarchical analysis; the complexity of individual phenotypes can be understood in the same framework as the diversity among individuals in a population, differentiation between populations, and so forth.

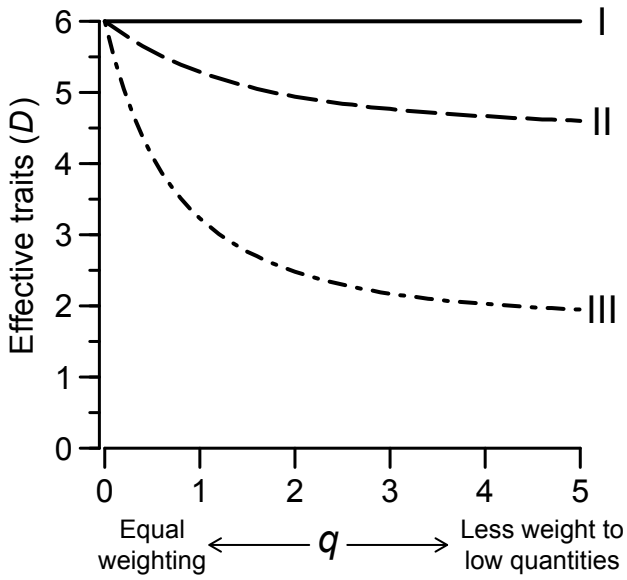


Figure 2.1: Diversity profile plot illustrating the relationship between effective phenotype diversity (D) and diversity order (q). In this example, six traits have been evaluated. There are three samples (I, II, and III), each differing only in the evenness of subsidiary trait quantity. When $q = 0$, traits are weighted equally regardless of differences in trait abundance, and each phenotype has six effective traits at this order. As the diversity order increases, traits with higher abundances become more influential, and less weight is given to traits of low quantity. In example I, the relative amounts of each component are equal, and the diversity profile is a horizontal line ($D = 6$ for all q). III is dominated by two traits comprising $\approx$ 60% and 20%, respectively, of the total abundance and so the profile curve descends sharply before leveling out at ≈ 2 effective traits. II is intermediate between I and III; the curve indicates mild disparity among components.

2.2 Phenotypic complexity and hierarchical diversity partitioning

Diversity or number of part types at a single level does not perfectly capture notions of organism-level or ecosystem-level complexity because complexity of part types is often heterogeneously distributed within and across organisms, lineages, or landscapes. Approaches that quantify the number of components or the hierarchical organization of parts are more satisfying. Community ecologists have a long tradition of partitioning a region's total species diversity into two components: a measure of the average within-location or within-community diversity (alpha) and a measure of the among-location or among-community differences in species composition (beta; Whittaker, 1972; Lande, 1996; Jost, 2007).

Just as communities are partitioned into within-group vs. among-group components, we can do the same for complex phenotypes. Alpha diversity (${}^{q}\alpha$) is the within-individual component. When using effective numbers, α diversity can be interpreted as the effective number of individual phenotypic elements in a sample: e.g., the effective number of genes transcribed within a tissue, or the effective number of deterrent compounds that make up a plant's anti-herbivore defenses. Gamma diversity (${}^{q}\gamma$) is the total effective number of constituent parts of a pooled group (e.g., an experimental plot, population, or species pool).

The among-group diversity component—beta diversity—has two different interpretations depending on whether partitioning is additive (Lande, 1996; Veech et al., 2002) or multiplicative (Jost, 2007; Chao et al., 2012). In an additive partitioning framework, additive beta, or "diversity excess," is calculated as ${}^{q}\beta^{+} = {}^{q}\gamma - {}^{q}\alpha$ and quantifies the effective number of group-level phenotypic components not found within a typical sample (Chao et al., 2012). Additive beta is in units of the "effective number of parts" as alpha and gamma are, and thus it reflects the magnitude of the difference between groups or samples. For example, if a population of ten plants is chemically defended against herbivores with a gamma diversity of ten effective defensive compounds and an alpha diversity of seven effective defensive compounds per plant on average, under the additive partitioning framework, $\beta^{+} = 10 - 7 = 3$ effective defensive compounds. That

is, effectively, three defensive compounds are in the population in excess of those present in a typical plant.

In the context of multiplicative partitioning, beta diversity (${}^q\beta = {}^q\gamma/{}^q\alpha$) is interpreted as the effective number of completely distinct phenotypic combinations present within a hierarchical level and estimates the extent of differentiation among phenotypes (Jost, 2006, 2007): e.g., the effective number of completely distinct chemotypes, or unique chemical cocktails, that an organism might use to defend against natural enemies within a population. Using the hypothetical plant defense example above, multiplicative beta would be $\beta = 10/7 \approx 1.43$ effective chemotypes; i.e., there would be 1.43 unique combinations of defensive chemicals in that population.

As defined above, γ and β are dependent on the number of individuals (N). Estimates based on a sample will tend to underestimate the true population γ and β. Moreover, comparisons among study systems would be confounded by differences in N. To provide a standardized estimator of turnover (${}^qT$; Jost, 2007), β can be transformed to

$$ {}^qT = \frac{{}^q\beta - 1}{N - 1} = \frac{{}^q\beta^+}{\alpha(N-1)}. \tag{2.3} $$

This differentiation measure expresses beta diversity as a fraction of its maximum possible value given N; it is zero when samples are identical and one when every phenotypic combination is distinct. In the hypothetical plant example, the turnover is approximately 0.048. Note that this is a global measure that can be sensitive to sampling. See Chao et al. (2012, 2014) for an in-depth treatment of measures of similarity and differentiation.

2.2.1 Level-wise vs. group-wise partitioning

Diversity partitioning has been extended to more than two levels (Crist et al., 2003). The general strategy for hierarchical partitioning is illustrated in Figure 2.2. The key feature of adopting this approach for phenotypes is that the complexity (α diversity) of the individual organism or structure is being expressed in the same framework as the variation among individuals and among groups of individuals (e.g., populations, regions, or experimental groups). This emphasizes interpretation of

the variation in a dataset in terms of partitioning or turnover (Anderson et al., 2011; Chao et al., 2012) of numbers of phenotypic elements, rather than more abstract multivariate distances.

We describe the approach proposed by Crist et al. (2003) as "level-wise" partitioning. Suppose we have $i = 1, 2, 3, ...m$ sampling (e.g., individuals, demes, regions, ... continents). The total diversity is partitioned at the highest level as $\gamma = \beta_m^+ + \alpha_m$. The average diversity within subunits is further partitioned as $\gamma = \beta_m^+ + \beta_{m-1}^+ + \alpha_{m-1}$, and so forth:

$$\gamma = \alpha_1 + \sum_{i=1}^{m} \beta_i^+. \tag{2.4}$$

Similarly for multiplicative partitioning,

$$\gamma = \alpha_1 \prod_{i=1}^{m} \beta_i. \tag{2.5}$$

This level-wise partitioning approach gives one β, one β^+, and one α for each level (Crist et al., 2003). Level-wise α_i is calculated from the average diversity within samples at hierarchical level i,

$$\alpha_i = D(\overline{H}_i), \tag{2.6}$$

where $\overline{H}_i$ is the diversity index (e.g., richness, Shannon index, etc.) at level i, and $D(H)$ is the function that converts the index into an effective number (Jost, 2006). $\overline{H}$ can be estimated in several ways depending on the diversity order (q) and the goals and assumptions of a given study. Often a simple average among subunits will be most appropriate and intuitive. Unbalanced sampling effort and some research questions might justify alternative weighting schemes (e.g., Chiu et al., 2014), but results can be difficult to interpret in the context of information theory when $q \neq 1$ (for discussion, see Jost, 2006, 2007; Chao et al., 2012).

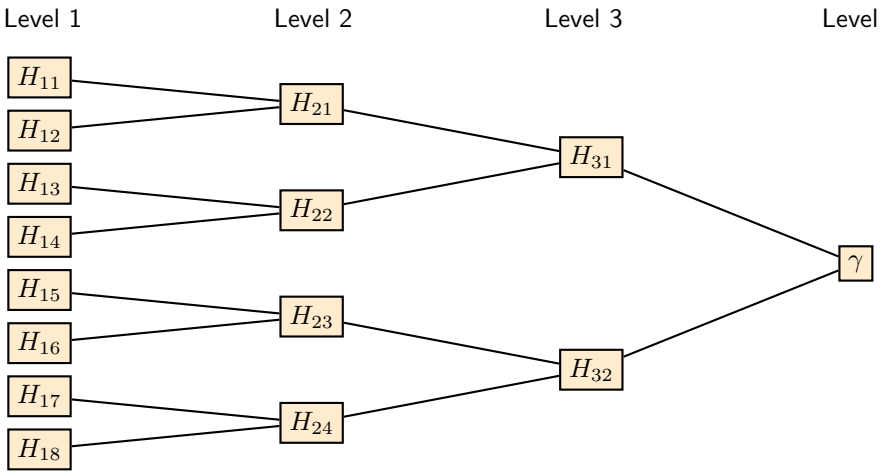
As a complementary alternative to level-wise partitioning, we propose group-wise partitioning, which summarizes diversity for each group at each hierarchical level. Instead of a single α_i and β_i at level i, we will have separate estimates for each of the n_{i+1} groups at the next level up. For example, if individuals were sampled from four populations, we would have four estimates

of mean phenotypic complexity and four estimates of among-individual diversity. This approach facilitates comparisons among groups at each hierarchical level, whereas level-wise partitioning emphasizes comparisons among levels in the hierarchy. Both partitioning designs are illustrated in Figure 2.2.

The partitioning scheme above, adopted from community ecology (Lande, 1996; Veech et al., 2002; Jost, 2007), is entirely general with respect to the order q. Previous approaches to phenotypes or functional traits have decomposed diversity at specified levels. However, the lowest level of organization or hierarchy is usually the individual organism. In contrast, our approach decomposes phenotypic complexity at all hierarchical levels, including the within-individual level; the complexity of the individual phenotype is quantified as the diversity of phenotypic elements. The approach can be implemented for a variety of diversity indices, and should not be taken as specific to the analysis of effective numbers. Proper estimation of diversity indices is an active area of development (Bonachela et al., 2008; Marcon et al., 2014; Zhang et al., 2014). Our implementation of this approach in the empirical examples below employs equal weighting and standard calculations of effective numbers.

2.3 Empirical examples

We present three case studies to illustrate how this method can be applied to different questions in ecology and evolutionary biology and provide biologically meaningful interpretations. Two of these studies are from research on chemically mediated phenotypes. The first asks how damage by a specialist herbivore affects complexity and diversity of plant defenses (Fordyce and Malcolm, 2000). The second asks whether changes in chemical diversity in toads can be explained as phenotypic plasticity. The third, a reanalysis of the data of Schweitzer et al. (2008), takes individual complexity into account when asking about the relative importance of tree genotype and species for associated soil microbial communities. All data for these examples are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.7vh21 (Marion et al., 2015*b*).



Level-wise partitioning

$$\alpha_1 = D\left(\tfrac{1}{8}\sum_{a=1}^{8} H_{1a}\right) \qquad \alpha_2 = D\left(\tfrac{1}{4}\sum_{b=1}^{4} H_{2b}\right) \qquad \alpha_3 = D\left(\tfrac{1}{2}\sum_{c=1}^{2} H_{3c}\right)$$
$$\beta_1^+ = \alpha_2 - \alpha_1 \qquad \beta_2^+ = \alpha_3 - \alpha_2 \qquad \beta_3^+ = \gamma - \alpha_3$$
$$\beta_1 = \alpha_2/\alpha_1 \qquad \beta_2 = \alpha_3/\alpha_2 \qquad \beta_3 = \gamma/\alpha_3$$

Group-wise partitioning

a indexes individuals in subpopulation b, b indexes subpopulation in population c

$$\alpha_{1b} = D\left(\tfrac{1}{2}\sum_{a} H_{1a}\right) \qquad \alpha_{2c} = D\left(\tfrac{1}{2}\sum_{b} H_{2b}\right) \qquad \alpha_3 = D\left(\tfrac{1}{2}\sum_{c=1}^{2} H_{3c}\right)$$
$$\beta_{1b}^+ = D(H_{2b}) - \alpha_{1b} \qquad \beta_{2c}^+ = D(H_{3c}) - \alpha_{2c} \qquad \beta_3^+ = \gamma - \alpha_3$$
$$\beta_{1b} = D(H_{2b})/\alpha_{1b} \qquad \beta_{2c} = D(H_{3c})/\alpha_{2c} \qquad \beta_3 = \gamma/\alpha_3$$

Figure 2.2: Hierarchical diversity partitioning for a hypothetical three-level design. This is a generalized hierarchy where levels one, two, three, and four might represent individuals, subpopulations, populations, and total. H_{ij} is any diversity index calculated for subunit j, of level i. $D(H)$ is the function that converts the index into an effective number (Jost, 2006), or the identity function if one wishes to work with H instead of effective numbers. In the figure, j is indexed differently for each level i: i.e., for level $i = 1$, j is indexed by a ($j \in a$); at level $i = 2$, $j \in b$; at level $i = 3$, $j \in c$. The α's are average within-level (or group) diversities, and γ is total diversity. Partitioning can be additive or multiplicative. Additive between-subunit diversity (β^+; "diversity excess") quantifies diversity not found within a typical sample. Multiplicative between-subunit diversity (β) represents the effective number of distinct configurations found within a level. Partitioning can be by level, where a single α_i and β_i characterizes each level, averaging across subunits with each rank. We also suggest group-wise partitioning, where separate α_{ij} and β_{ij} components are calculated for each grouping. In this example, there are four group-wise β_{1j}'s at the first level and two β_{2j}'s at the second level. For an expanded generalization, see online appendix A.

We use hierarchical bootstrapping to approximate the uncertainty around the diversity estimates (Efron, 1982). In each iteration, the appropriate subgroups contained within a level are re-sampled, the groups within those subgroups are resampled, and so forth. Confidence intervals approximated through bootstrapping should be interpreted with some caution, as bootstrapping might underestimate uncertainty because bootstrap replicates are not independent random samples of the population (Schenker, 1985; Efron and Tibshirani, 1986; Dixon et al., 1987). This might be especially relevant at lower orders of q, as the importance of rare aspects of multivariate phenotypes are emphasized. This is a general limitation of bootstrapping and is not unique to our approach. Although bootstrapping might not be ideal, it is frequently used for diversity analyses (Gotelli and Colwell, 2010; Marcon et al., 2012), and we include it here to provide an approximation of uncertainty.

For the analyses, we used functions from the *vegan* (Oksanen et al., 2012) and *vegetarian* (Charney and Record, 2012) packages in R (R Development Core Team, 2014). We implemented the group-wise partitioning approach described above using the CRAN package *hierDiversity* (Marion et al., 2015*a*).

2.3.1 Inter-tissue induction of cardenolides in common milkweeds (Asclepius syriaca)

Most milkweeds (*Asclepius spp.*; Apocynaceae) employ a complex blend of cardiotoxic steroids called cardenolides that putatively act to deter generalist consumers. However, specialist herbivores employ behavioral or physiological strategies to overcome or circumvent these defenses (Malcolm and Brower, 1989; Malcolm, 1992). In an observational experiment with common milkweed (*A. syriaca*), Fordyce and Malcolm (2000) examined the qualitative and quantitative variation in cardenolides among five plant tissue types (leaves, stem epidermis, cortex, vascular cylinder, and pith) and whether these defenses differed between plants with or without larvae of the specialist weevil, *Rhyssomatus lineaticollis*. Larval weevils specialize on milkweed pith tissue, which—the authors hypothesized—was a strategy of spatially avoiding cardenolides.

Contrary to their hypothesis, Fordyce and Malcolm (2000) found that the pith tissue had the second-highest cardenolide concentrations after the vascular tissue, and thus weevils were not spatially avoiding cardenolides. Further, they found that damaged plants had 33% lower cardenolide concentrations overall, and that there was a qualitative shift towards more lipophilic compounds in damaged plants. Because adult weevils feed on leaves prior to oviposition, and damaged ramets had lower cardenolide concentrations in total, Fordyce and Malcolm (2000) speculated that *R. lineaticollis* females might be manipulating host-plant chemistry for the benefit of their offspring or choosing plants with lower concentrations.

To address the questions of whether damaged or undamaged plants were more chemically complex and whether they differed in how that complexity was partitioned among tissues, we reanalyzed the dataset using the multiplicative diversity partitioning methodology described above (Figure 2.3). Within individuals, damaged plants had 1–2 more compounds per tissue than undamaged plants (α). This trend was most pronounced in the vascular tissue and cortex and least apparent in the pith and epidermis. Yet the chemical compositions among tissues in undamaged plants ($^1\beta$ = 2.11, 95% CI [2.01–2.21]) were more different than the tissues of plants with signs of weevil damage ($^1\beta$ =1.61, [1.54–1.68]; Figure 2.3A). There was almost twice as much chemical turnover among tissues of undamaged plants compared to tissues from damaged plants (0.3 [0.27–0.33] vs. 0.16 [0.14–0.18] for $q = 1$; Figure 2.3C).

At the among-individual within damage-treatment level, damaged plants had about one compound more than undamaged plants, and about two more effective compounds overall. Interestingly, for the highest diversity orders, damaged plants had slightly higher beta diversity as well (Figure 2.3B), although this was largely driven by chemical differences in vascular tissue composition. This suggests that damaged individuals did not differ in low-abundance compounds but did vary in the composition of the few compounds of high concentration, with about 10% turnover in chemical phenotype among individuals (Figure 2.3D).

Fordyce and Malcolm (2000) were able to infer differences in chemistry between damaged and undamaged plants but their characterization of these differences remained incomplete. Using the diversity partitioning approach, we add three biological insights. First, tissues from damaged plants tended to be more similar to one another than the tissues of undamaged plants (fig. 2.3A &

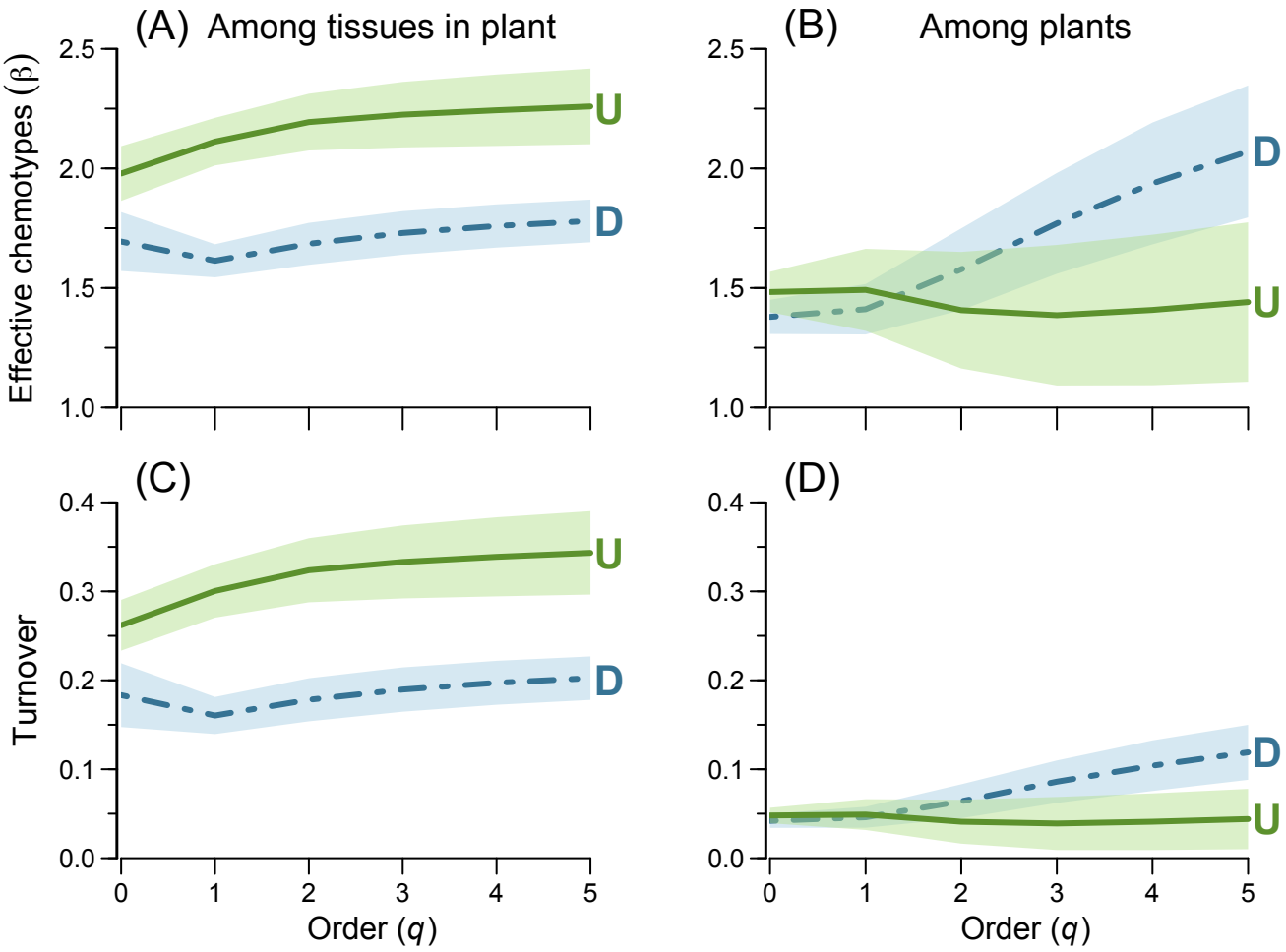


Figure 2.3: Diversity profile plots of chemotypic beta diversity (A & C averaged among tissues within individual milkweed ramets and (B & D) among ramets of *A. syriaca* with or without herbivore damage (Fordyce and Malcolm, 2000). Damaged plants (blue dashed lines) were identified by the presence of characteristic oviposition scars produced by the specialist weevil *R. lineaticollis*; undamaged plants (green solid lines) lacked such scars. For diversity orders $q =$ 0–5, chemotypic variation among tissues or ramets is expressed as (A & B) the effective number of chemotypes or unique chemical blends present or (C & D) the effective compound turnover. Shaded confidence intervals for each treatment are 2 SE. Errors for (B & D) were calculated via hierarchical bootstrapping because there is only one among-plant observation for each treatment.

C). Although they could not quantify it, Fordyce and Malcolm (2000) hypothesized that females may assess/manipulate plant chemistry prior to oviposition, narrowing the chemotypic uncertainty of cardenolide composition for their offspring. The diversity-as-complexity approach provides evidence that chemical variation among tissues is lower in weevil-damaged plants, supporting their hypothesis. Second, this change in complexity was accompanied by a shift in partitioning; among-individual variation was increased as damaged plants appeared to become more heterogeneous in chemical phenotype (Figure 2.3B & D). Thus, in a patch with ramets infected by *R. lineaticollis*, the associated increase in ramet chemical heterogeneity might affect the colonization and assembly of subsequent herbivores in the community. Finally, damaged plants were more chemically complex in that they expressed a greater effective number of cardenolides in their tissues. The increase in the effective number of cardenolides found in damaged plants might explain why plants that are colonized by weevils early in the season suffer less cumulative herbivory later (Van Zandt and Agrawal, 2004) if later herbivores are unable to tolerate the increased cardenolide diversity.

2.3.2 Experimental induction of chemical defenses in American toads (Bufo/Anaxyrus americanus)

Many amphibians display remarkable plasticity in their responses to environmental conditions (Newman, 1992; Relyea, 2001), and the effects of natural enemy cues on larval amphibian plasticity have been well-documented (see Benard, 2004; Relyea, 2007, for reviews). Yet, to our knowledge, the plasticity of amphibian chemical defenses has been studied only a handful of times (e.g., Benard and Fordyce, 2003; Hagman et al., 2009; Hayes et al., 2009), which is surprising considering that—among vertebrates—amphibians are most notable in their diversity of putative chemical defenses (Daly, 1995).

We applied our partitioning method to a subset of previously unpublished data from a study that assessed the plasticity of chemical defenses (bufadienolides) in adult American toads (*Bufo americanus*) following repeated expression of the parotoid glands. This study tested the hypothesis that toads facultatively alter the composition of their defensive secretions in response

to frequent "attacks". The complexity-as-diversity approach allows us to ask whether the diversity of compounds is affected by repeated expression.

Parotoid secretions were collected on pre-weighed filter paper after manual expression by compressing each gland between two fingers. Following an initial sampling, toads were randomly assigned to one of three expression-frequency treatments ($N = 42$)—frequent expression (5x), medium expression (1x), or low expression (0x)—and sampled over five months. At the experiment's conclusion, toad glands were expressed one final time for comparison, followed by high-performance liquid chromatography (HPLC) of the secretions. Here we compare the baseline chemotypes to the final chemotypes among treatments.

The traditional ordination-based hypothesis test (distance-based redundancy analysis (dbRDA); Legendre and Anderson, 1999) might support the hypothesis that parotoid expression led to statistically significant chemotypic differences among treatments ($F_{2,35} = 3.473$, $p = 0.015$; Figure 2.4A). However, total concentrations of bufadienolides were significantly lower in the high expression treatment ($F_{2,35} = 25.212$, $p < 0.001$; Figure 2.4B), suggesting an alternative explanation. A simple depletion of bufadienolides from repeatedly expressed glands might generate differences as a side-effect. As rare compounds become lost (or undetectable), the distribution of relative abundances can become distorted, analogous to a bottleneck effect on genetic diversity. The diversity partitioning approach provides an intuitive way to evaluate this alternative hypothesis.

Diversity profile plots illustrate three key impacts of repeated parotoid expression on the diversity of bufadienolides. First, alpha diversity (chemical complexity of a given toad's secretion) decreases (Figure 2.4C), as we might expect from depletion. Second, beta diversity increases (Figure 2.4D), as expected by independent random sampling (like genetic drift) as different toads lose compounds by chance. Finally, the way alpha diversity decreases and beta increases differs for different diversity orders of q. As with a genetic bottleneck effect (Luikart et al., 1998), rare compounds are lost rapidly (changing diversities of order $q = 0$) without dramatically shifting the relative abundances of more common compounds (diversities of higher order change less). Equivalently, in community ecology, it is generally understood that richness ($q = 0$) is much more

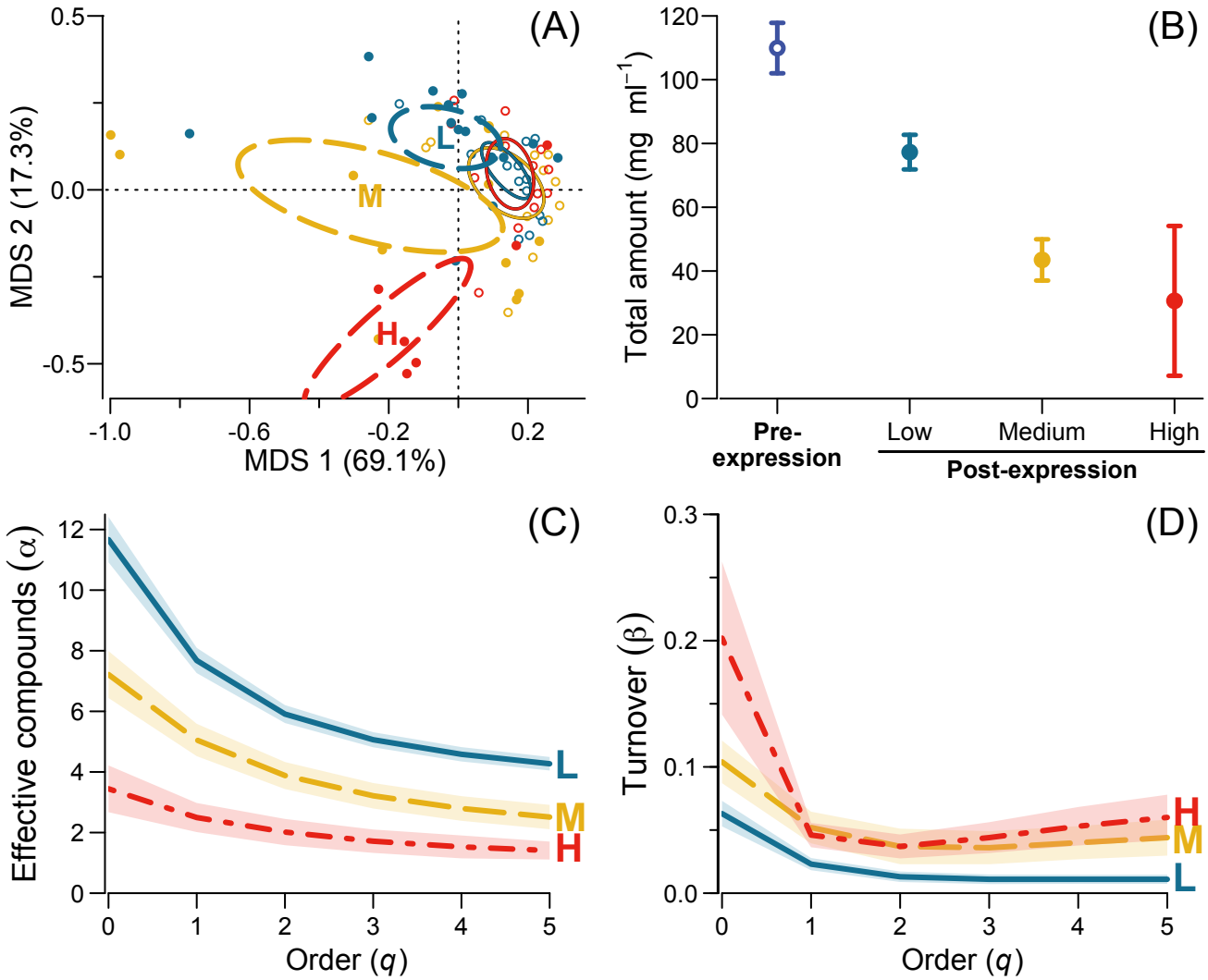


Figure 2.4: Multiple methods for quantifying the bufadienolide complexity of American toad (*B. americanus*) defenses due to different frequencies of parotoid expression. Following an initial expression, adults were assigned to low (blue), medium (yellow), or high (red) parotoid expression treatments. (A) Ordination using horn dissimilarities ($q = 1$) shows significant chemotypic differences among post-expression treatments (dbRDA: $p = 0.01$; filled circles, dashed lines) but not pre-expression treatments ($p = 0.91$); open circles, solid lines). Ellipses are 2 SE around treatment centroids. (B) Mean amounts of total bufadienolides pre- (open) and post-expression (filled). Pre-expression treatments were statistically similar ($p = 0.249$) and were therefore pooled. Error bars are 2 SE. Diversity profiles for (C) the average effective number of compounds and (D) effective chemical turnover among individuals following expression. Confidence intervals are 2 SE and were calculated using hierarchical bootstrapping.

sensitive to sampling than Shannon's H ($q = 1$) or Simpson's λ ($q = 2$; Colwell and Coddington, 1994; Chao et al., 2005).

We further supported the depletion hypothesis by examining the impact of sampling on the ordination test. First, simply raising the detection threshold for compounds in the low expression treatment to mimic the effects of lowered concentration (setting all compounds with relative abundance < 0.15 to zero and recalculating relative abundance of the remaining compounds) made the difference disappear ($F_{2,35} = 1.339$, $p = 0.305$). Second, we compared the original low expression data to the recalculated low expression data and found high statistical support for a difference ($F_{1,28} = 5.4$, $p = 0.006$).

In this example, analyzing the chemical complexity of parotoid secretions as diversity within and between toads provides a clearer view of experimentally induced changes than that offered by the more abstract results of multivariate ordination. This is because the complexity is parsed into interpretable and biologically relevant estimates. The differences illustrated in Figure 2.4A might be taken as evidence for the idea that toads produce different chemical phenotypes as a plastic response to repeated "attacks." However, borrowing an insight from population genetics (i.e., the random sampling of genetic drift) regarding the effects of depletion on diversity profiles, together with evidence of depletion in Figure 2.4B, leads us to infer that the chemical differences between toads exposed to different treatment levels reflect random drift owing to depletion of bufadienolides in frequently expressed parotoid glands.

2.3.3 Plant genotypic variation and soil microbial phenotypic diversity

Gene expression or metabolic markers can also be used to assess phenotypic complexity and diversity in variety of contexts. Schweitzer et al. (2008) used phospholipid fatty acid (PLFA) biomarkers to assess rhizosphere microbial community composition as a complex extended phenotype of *Populus* trees, asking whether PLFA profiles consistently differed among tree genotypes in a common garden. With 19 PLFAs assayed, they chose to summarize the PLFA phenotype using a one-dimensional nMDS ordination to provide a single number (the nMDS score) for each tree. They then performed univariate ANOVAs with the nMDS score as the

dependent variable to quantify variation among genotypes. We reanalyzed a subset of their data to illustrate how diversity partitioning can offer additional insights relative to ordination-based approaches that summarize complexity with abstract multivariate scores.

The dataset includes 3–4 clones each of four genotypes of *Populus fremontii* and five genotypes of *Populus angustifolia* (Figure 2.5). For ordination, we computed dissimilarities for $q = 0$ (Sorenson distance for presence-absence), $q = 1$ (Horn distance for information), and $q = 2$ (Morisita-Horn distance for probability of identity). We used permutational MANOVA on distances (Anderson, 2001) and dbRDA (Legendre and Anderson, 1999) to evaluate how well the dissimilarity matrices for different levels of q were accounted for by clonal genotype within each species. Depending on the analysis, genotype (within species) was estimated to account for 30-60% of the variation among trees. This is consistent with the results of Schweitzer et al. (2008).

We then used diversity partitioning to express the variation among genotypes and species in terms of the complexity of the individual PLFA profiles. This approach agrees with the distance-based ordination analyses in the sense that genotype appears to account for a large fraction of the additive beta diversity (Table 2.1). However, beta diversity is itself only a tiny fraction relative to the complexity of each individual (Figure 2.6). In terms of turnover, trees of a single genotype differed from each other by 2%, and genotypes differed from other genotypes by only 1%. Moreover, nested randomization tests did not support rejecting the null hypothesis that variation (beta diversity) between trees of the same genotype is equivalent to variation between trees of different genotypes or species.

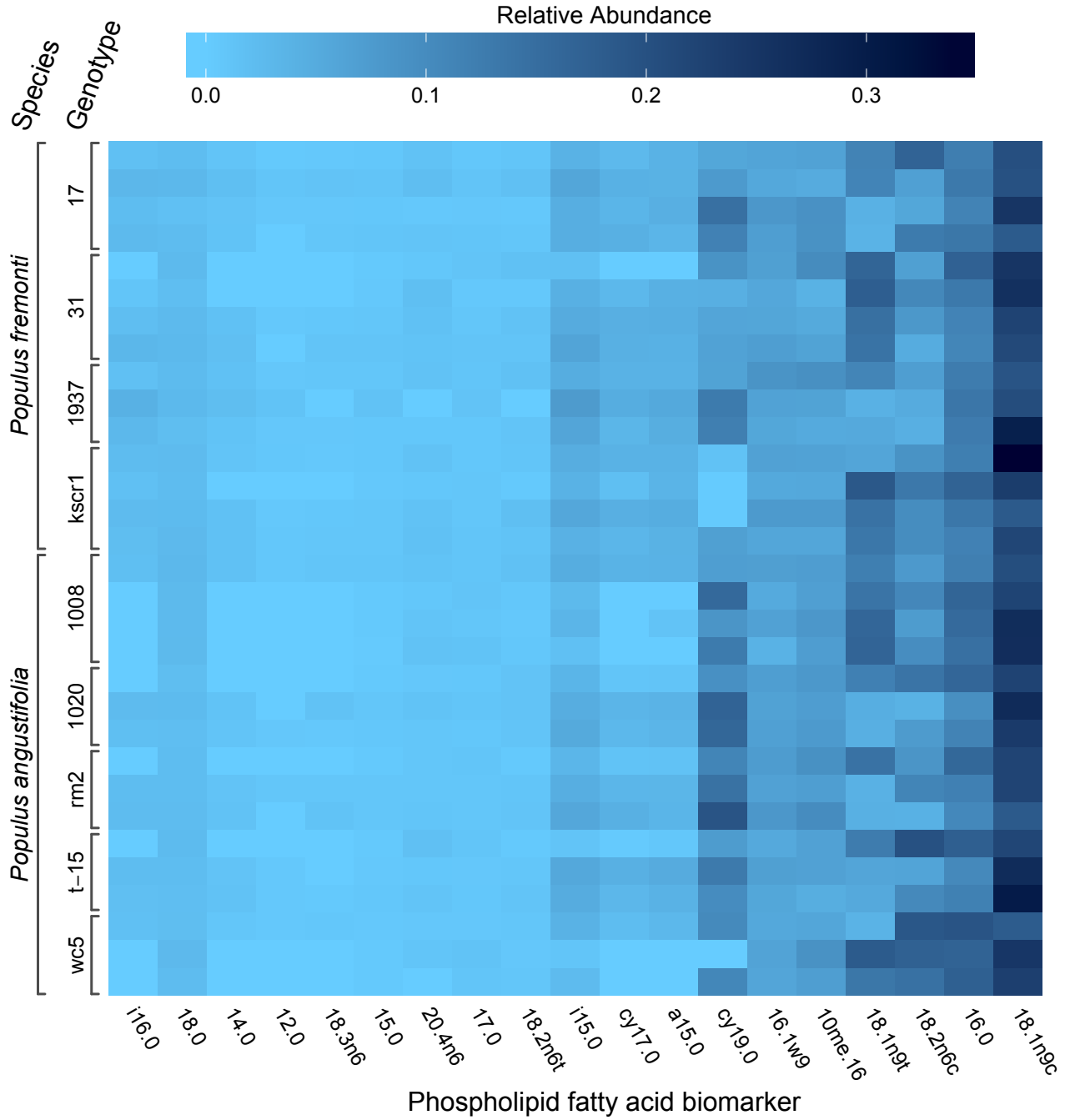


Figure 2.5: Heat map visualization of the relative abundances of each phospholipid fatty acid biomarker (Schweitzer et al., 2008) among replicated genotypes of two cottonwood species.

Table 2.1: Results from alternative methods of partitioning PFLA complexity within and between two *Populus* species. Distance-based approaches (dbRDA & perMANOVA) ignore the within-tree diversity (complexity) therefore the percent variance explained is only over the between-tree variation. For all estimates we show results for diversity of order $q = 2$. For additive level-wise partitioning Simpson's λ & percent variance explained is given. For level-wise and group-wise multiplicative partitioning we give effective numbers with turnover in parentheses. Under group-wise partitioning, estimates of trees-within-genotypes and within-tree diversity are averages.

			Level-wise		Group-wise	
	dbRDA	PerMANOVA	Additive	Multiplicative	*P. angustifolia*	*P. fremontii*
Between spp.	7.88%	12.11%	< 0.01 (0.14%)	1.01 (0.01)	1.01 (0.01)	
Genotypes in spp.	37.09%	45.69%	< 0.01 (0.49%)	1.03 (< 0.01)	1.04 (0.01)	1.03 (0.01)
Trees in genotypes	55.04%	42.20%	0.01 (0.79%)	1.05 (< 0.01)	1.05 (0.02)	1.05 (0.02)
Within trees (α)			0.87 (98.6%)	8.01	7.46	8.35

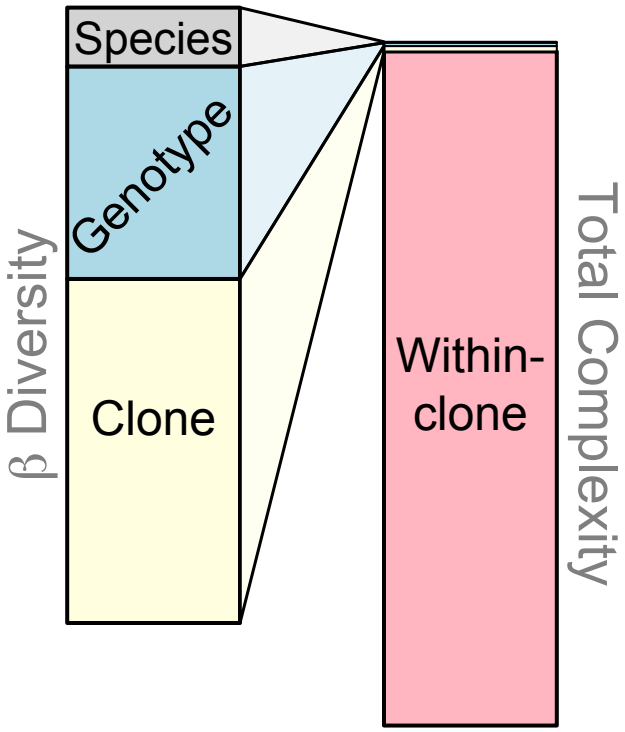


Figure 2.6: Diagram of the relative proportions of phospholipid fatty acid biomarker variation explained using additive diversity partitioning among *Populus* species, genotypes, and individuals from Schweitzer et al. (2008). Although genotype explains a large fraction of the beta or among-plant diversity, it explains $< 2\%$ of the total PFLA complexity present.

In this example, a pattern elucidated by ordination found to be statistically significant is shown to be biologically subtle in relation to the complexity of individual samples (Table 2.1, Figure 2.5). Variation within individuals dwarfs the variation among individuals, genotypes, or species (Figure 2.6). Ordination-based measures start at the between-individual level (dissimilarity) and entirely ignore the within-individual level of biological complexity, which our approach captures (i.e., α diversity). Figure 2.6 is consistent with the idea that the structure and function of microbial communities needs to be understood at a finer spatial scale (Ettema and Wardle, 2002). Our analysis of this dataset also emphasizes the wide range of biological phenomena that can be treated as "complex" phenotypes.

2.4 Conclusion

Virtually all phenotypes function and evolve as integrated multivariate systems that exhibit both qualitative and quantitative variation. Yet many of the most common analytical approaches typically applied to complex phenotypic data give results that reduce complexity at the cost of biological intuition. Our adoption of diversity partitioning provides a novel framework for quantifying complex traits by hierarchically separating the variation into biologically meaningful within- and among-group components of interest. Our case studies highlight the utility of our approach by uncovering aspects of variation in the data that other complementary methods missed. Differences between damaged and undamaged milkweeds (Fordyce and Malcolm, 2000) were more thoroughly characterized as differences in partitioning of chemical complexity among plant tissues. Changes in the secretions of repeatedly "attacked" toads were interpreted as effects of a physiological bottleneck rather than an adaptive induced response. Dissimilarity among rhizosphere communities associated with different cottonwood genotypes and species (Schweitzer et al., 2008) was shown to be small when compared to the phenotypic complexity within each sample. In all of these examples, insight was added by analyzing the complexity of individual phenotypes as an integral component of the functional ecological diversity of the study system.

The complexity-as-diversity approach can be extended to include more information about differences among constituent elements. In community ecology, functional or phylogenetic dissimilarity has been incorporated as a distance, e.g., Rao's (1982) quadratic entropy (Mouchet et al., 2010; Chiu et al., 2014). It would be straightforward to include, say, differences in polarity between compounds in our analyses of milkweed or toad defensive chemistry. Additionally, our approach outlined here used relative amounts as opposed to absolute amounts or abundances. Recently, Chiu et al. (2014) have extended diversity partitioning to incorporate absolute abundances, which may be of more interest depending on the question. Other extensions to this approach might be possible.

The framework outlined here is particularly appropriate for analyzing complex phenotypes whose constituent variables can be intuitively conceived as relative amounts, as in the examples

given and in Iason et al. (2005). Many other phenotypes are clearly appropriate for a complexity-as-diversity approach. Diversity of cell or tissue types might have regular scaling relationships with size, abundance, and species diversity (Bell and Mooers, 1997). Diversity and dissimilarity of gene expression patterns might be related to functional complexity and modularity of phenotypes. Hierarchical analysis can be used in ecological metabolomics (Sardans et al., 2011; Jones et al., 2013) to quantify the contributions of cells, organisms, populations, etc. to the physiological complexity of ecosystems. Behavioral ecologists might quantify behavioral complexity and diversity of populations and communities by analysis of time spent doing various activities or numbers of times each behavior is performed. It is less obvious why one would use this approach on something like limb morphology, though there might be value in summarizing the unevenness of relative bone lengths or the relative contributions of several skeletal elements to a given movement. We think the complexity-as-diversity approach sufficiently differs from previous analyses that it might inspire entirely new questions or hypotheses about the ecology and evolution of phenotypes.

2.5 Acknowledgments

We thank J. Beaulieu, S. Campangna, J. Pruitt, N. Sanders, D. Simberloff, R. Zenni, and the HOFF group for helpful comments and discussion. J. Schweitzer kindly provided the *Populus* dataset and insightful discussion. We also thank E. MacDonald for the toad dataset. This research was funded by the National Science Foundation (DDIG award no. DEB-1405887 to Z. H. Marion and DEB-0614223 & 1050947 to James A. Fordyce) and the Department of Ecology & Evolutionary Biology at the University of Tennessee (Z. H. Marion).

2.6 Supporting information

2.6.1 Expanded description of the hierarchical partitioning scheme

Here we provide an expanded description of hierarchical diversity partitioning (e.g., Lande, 1996; Veech et al., 2002; Crist et al., 2003). To be consistent with the main text, we use the specific

4-level sampling design depicted in Figure 2.2 of the main text. Figure 2.7 is reproduced below with an alternate notation to that used in the main text.

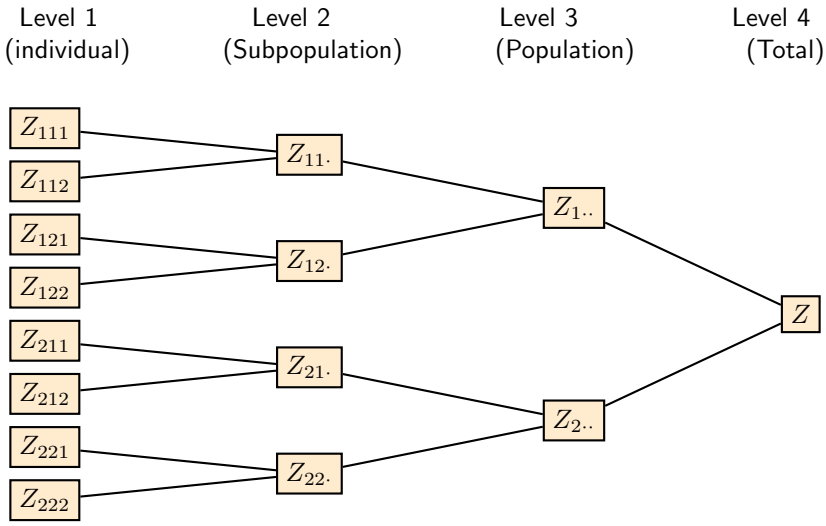


Figure 2.7: Hypothetical nested sampling scheme from Figure 2.2 of the main text with alternative notation. Here, we represent the full dataset as Z. The data from individual k in subpopulation j in population i is Z_{ijk}.

Aside from the group-wise partitioning scheme, the calculations described herein were adopted directly from community ecology. More thorough background and discussion can be found in several recent publications (Jost, 2007; Chao et al., 2012). We implemented our approach using the R (R Development Core Team, 2014) package *hierDiversity* (Marion et al., 2015*a*).

Imagine we have a dataset Z of sizes or amounts of phenotypic elements (say concentrations of L different enzymes) from a set of individuals sampled from subpopulations that are grouped in populations (Figure 2.7). That is, z_{ijkl} is the amount of enzyme l found in individual k from subpopulation j of population i. The data for all L enzymes from that individual is represented as Z_{ijk}. To compute diversity indices and numbers equivalents, we first convert the raw data to

relative amounts p_{ijkl} and compute averages at each level.

$$\begin{aligned}
\text{individual } k\text{:} \quad & p_{ijkl} = \frac{z_{ijkl}}{\sum_{l=1}^{L} z_{ijkl}} \\
\text{subpopulation } j \text{ mean:} \quad & \bar{p}_{ij\cdot l} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} p_{ijkl} \\
\text{population } i \text{ mean:} \quad & \bar{\bar{p}}_{i\cdot\cdot l} = \frac{1}{n_i} \sum_{j=1} \bar{p}_{ij\cdot l} \\
\text{grand mean:} \quad & \bar{\bar{\bar{p}}}_{\cdot\cdot\cdot l} = \frac{1}{n} \sum_{i=1}^{n} \bar{\bar{p}}_{i\cdot\cdot l}
\end{aligned} \tag{A1}$$

where n is the number of subunits at level 3 (in this example, $n = 2$ populations), n_i is the number of level 2 subunits within the i-th subunit of level three (here, $n_i = 2$ subpopulations within each population), and n_{ij} is the number of level 1 subunits within level 2 subunit j of level 3 subunit i (here, $n_{ij} = 2$ individuals within each subpopulation). Note that averaging could be weighted differently (e.g., we could average over all individuals at each level), but this can lead to nonsensical results (Jost, 2007).

Now compute the diversity indices ${}^qH$ and Hill numbers for each partition of the data to represent the effective number of enzymes expressed.

$$\begin{aligned}
\text{complexity of individual } k\text{:} \quad & {}^qD_{ijk} = {}^qD({}^qH_{ijk}) \quad \text{e.g.,} \quad {}^2H_{ijk} = \sum_{l=1}^{L} p_{ijkl}^2 \\
\text{diversity of subpopulation } j\text{:} \quad & {}^qD_{ij\cdot} = {}^qD({}^qH_{ij\cdot}) \quad \text{e.g.,} \quad {}^2H_{ij\cdot} = \sum_{l=1}^{L} \bar{p}_{ij\cdot l}^{\,2} \\
\text{diversity of population } i\text{:} \quad & {}^qD_{i\cdot\cdot} = {}^qD({}^qH_{i\cdot\cdot}) \quad \text{e.g.,} \quad {}^2H_{i\cdot\cdot} = \sum_{l=1}^{L} \bar{\bar{p}}_{i\cdot\cdot l}^{\,2} \\
\text{total diversity:} \quad & {}^qD = {}^qD({}^qH_{\cdots}) = {}^q\gamma \quad \text{e.g.,} \quad {}^2H_{\cdots} = \sum_{l=1}^{L} \bar{\bar{\bar{p}}}_{\cdot\cdot\cdot l}^{\,2}
\end{aligned} \tag{A2}$$

Level-wise partitioning

Following Crist et al. (2003), additive level-wise partitioning of the total diversity is as follows.

Alpha diversity within populations

$$ {}^{q}\alpha_3 = {}^{q}D\left(\frac{1}{n}\sum_{i=1}^{n}{}^{q}H_{i\cdot\cdot}\right) \tag{A3} $$

Diversity excess between populations

$$ {}^{q}\beta_3^{+} = {}^{q}\gamma - {}^{q}\alpha_3 \tag{A4} $$

Alpha diversity within subpopulations (average over the entire dataset)

$$ {}^{q}\alpha_2 = {}^{q}D\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}{}^{q}H_{ij\cdot}\right)\right] \tag{A5} $$

Average diversity excess between subpopulations, within populations

$$ {}^{q}\beta_2^{+} = {}^{q}\alpha_3 - {}^{q}\alpha_2 \tag{A6} $$

Average complexity of individuals (average over the entire dataset)

$$ {}^{q}\alpha_1 = {}^{q}D\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\left(\sum_{k=1}^{n_{ij}}{}^{q}H_{ijk}\right)\right]\right) \tag{A7} $$

Average diversity between individuals within subpopulations

$$ {}^{q}\beta_1^{+} = {}^{q}\alpha_2 - {}^{q}\alpha_1 \tag{A8} $$

Note the nested averaging illustrated in equations 5 and 7 ensures equal weighting at each hierarchical level. In principle, other weighting schemes could be used. For example, Crist et al. (2003) suggested weighting by the proportion of individuals found in each subunit. However, Jost (2006, 2007) warns that the information theory for weighted analysis has not been resolved for non-Shannon diversity indices (Jost, 2006, 2007).

Multiplicative β and turnover ${}^{q}T$ can be calculated at any level (Jost, 2007; Chao et al., 2012)

$$ {}^{q}\beta_i = \frac{{}^{q}\alpha_i + {}^{q}\beta_i^{+}}{{}^{q}\alpha_i} \tag{A9} $$

and

$$ \begin{aligned} {}^{q}T_i &= \frac{{}^{q}\beta_i^{+}}{{}^{q}\alpha_i(N-1)} \\ &= \frac{{}^{q}\beta_i - 1}{N-1} \end{aligned} \tag{A10} $$

Group-wise partitioning

Level-wise partitioning compares complexity and diversity across hierarchical levels. Sometimes, investigators might be more interested in comparing complexity and diversity among groups at the same level in a sampling scheme or experimental design. For this purpose, we propose group-wise partitioning. Group-wise partitioning provides separate α and β components for each subunit in a hierarchy. For the example in Figure A1, the calculations are as follows.

Alpha diversity within populations

$$ {}^{q}\alpha_{\cdot\cdot} = {}^{q}D\left(\frac{1}{n}\sum_{i=1}^{n}{}^{q}H_{i\cdot}\right) \tag{A11} $$

Diversity excess between populations

$$ {}^{q}\beta^{+}_{\cdot\cdot\cdot} = {}^{q}\gamma - {}^{q}\alpha_{\cdot\cdot\cdot} \tag{A12} $$

Alpha diversity of subpopulations within each population

$$ \begin{aligned} {}^{q}\alpha_{1\cdot\cdot} &= {}^{q}D\left(\frac{1}{n_1}\sum_{j=1}^{n_1} {}^{q}H_{1j}\right) \\ {}^{q}\alpha_{2\cdot\cdot} &= {}^{q}D\left(\frac{1}{n_2}\sum_{j=1}^{n_2} {}^{q}H_{2j}\right) \end{aligned} \tag{A13} $$

Diversity excess between subpopulations within each population

$$ \begin{aligned} {}^{q}\beta^{+}_{1\cdot\cdot} &= {}^{q}D({}^{q}H_{1\cdot\cdot}) - {}^{q}\alpha_{1\cdot\cdot} \\ {}^{q}\beta^{+}_{2\cdot\cdot} &= {}^{q}D({}^{q}H_{2\cdot\cdot}) - {}^{q}\alpha_{2\cdot\cdot} \end{aligned} \tag{A14} $$

Average complexity of individuals within each subpopulation

$$ \begin{aligned} {}^{q}\alpha_{11\cdot} &= {}^{q}D\left(\frac{1}{n_{11}}\sum_{k=1}^{n_{11}} {}^{q}H_{11k}\right) \\ {}^{q}\alpha_{12\cdot} &= {}^{q}D\left(\frac{1}{n_{12}}\sum_{k=1}^{n_{12}} {}^{q}H_{12k}\right) \\ {}^{q}\alpha_{21\cdot} &= {}^{q}D\left(\frac{1}{n_{21}}\sum_{k=1}^{n_{21}} {}^{q}H_{21k}\right) \\ {}^{q}\alpha_{22\cdot} &= {}^{q}D\left(\frac{1}{n_{22}}\sum_{k=1}^{n_{22}} {}^{q}H_{22k}\right) \end{aligned} \tag{A15} $$

Diversity between individuals within each subpopulation

$$
\begin{aligned}
{}^{q}\beta^{+}_{11\cdot} &= {}^{q}D({}^{q}H_{11\cdot}) - {}^{q}\alpha_{11\cdot} \\
{}^{q}\beta^{+}_{12\cdot} &= {}^{q}D({}^{q}H_{12\cdot}) - {}^{q}\alpha_{12\cdot} \\
{}^{q}\beta^{+}_{21\cdot} &= {}^{q}D({}^{q}H_{21\cdot}) - {}^{q}\alpha_{21\cdot} \\
{}^{q}\beta^{+}_{22\cdot} &= {}^{q}D({}^{q}H_{22\cdot}) - {}^{q}\alpha_{22\cdot}
\end{aligned}
\tag{A16}
$$

Multiplicative β and turnover can be calculated for each subunit as before.

Chapter 3

Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover

This chapter is in revision at Ecology and co-authored with James A. Fordyce and Benjamin M. Fitzpatrick.

Abstract

Beta diversity is an important metric in ecology quantifying differentiation or disparity in composition among communities, ecosystems, or phenotypes. To compare systems with different sizes (N, number of units within a system), beta diversity is often converted to related indices such as turnover or local/regional differentiation. Here we use simulations to demonstrate that these are naïve measures of dissimilarity because they depend on sample size and design. We show that when N is the number of sampled units (e.g., quadrats) rather than the "true" number of communities in the system (if such exists), these differentiation measures are biased estimators. We propose using average pairwise dissimilarity as an intuitive solution. That is, instead of attempting to estimate an N-community measure, we advocate estimating the expected dissimilarity between any random pairs of communities (or sampling units)—especially when the "true" N is unknown or undefined. Fortunately, measures of pairwise dissimilarity or overlap have been used in ecology for decades, and their properties are well known. Using the same simulations, we show that average pairwise metrics give consistent and unbiased estimates regardless of the number of survey units sampled. We advocate pairwise dissimilarity as a general standardization to ensure commensurability of different study systems.

3.1 Introduction

Diversity decomposition is an important technique for describing qualitative and quantitative aspects of variation within and among communities, ecosystems, and even phenotypes (Whittaker, 1972; Lande, 1996; Jost, 2007; Marion et al., 2015*c*). Under this paradigm, the total diversity (gamma diversity, D_γ) of a set of units is partitioned into two components: a component describing the average diversity of items (such as species) within a unit (alpha diversity, D_α) and a component (beta diversity, D_β) describing the variability among units in their composition of items (Whittaker, 1972; Anderson et al., 2011; Chao et al., 2014).

Here, we describe an underappreciated source of confusion regarding the among-unit (β) component and advocate a simple solution general to any field of study concerned with diversity (e.g., Strong et al., 1998; Yeo and Burge, 2004; Stinson, 2005; Hausser and Strimmer, 2009). We suggest using well-known and intuitive pairwise measures of dissimilarity rather than set-wise (or N-community) measures that are related to the size of the set. This is because set-wise measures can not be easily estimated from a sample nor compared across sets differing in size. For this paper we will focus on the sets and units of interest in community ecology: diversity refers to species diversity; "unit" refers to a community or survey unit (e.g., patch, transect, or quadrat); and "set" refers to a group of communities or survey units (e.g., a region or study area with samples that comprise the set).

Although there is debate about how to quantify diversity (e.g., as species richness, Shannon's H, Simpson's D, etc.), there is growing consensus that a family of metrics based on Hill numbers (Hill, 1973; Jost, 2006) are generally appropriate for ecology (see Ellison, 2010). Hill number diversities (${}^qD$) vary in the parameter q, which determines sensitivity to abundances. For example, when $q = 0$ then ${}^0D_\alpha$ depends only on presence and absence (i.e., it is the average richness or number of species present within a unit). When $q = 1$, ${}^1D_\alpha$ depends somewhat on abundance and is equal to the exponential of Shannon entropy. And when $q = 2$, then ${}^2D_\alpha$ depends more

strongly on the most abundant species and is equal to the inverse of Simpson's concentration (the probability that two randomly sampled individuals are the same species).

Beta diversity (D_β) is often of particular ecological interest because it describes differences in composition among communities, groups, or survey units (e.g., among habitat patches or along gradients in space or time). Depending on the goal of a study, the total diversity can be decomposed using either multiplicative ($\gamma = \alpha \times \beta$) or additive ($\gamma = \alpha + \beta^+$) partitioning (Ellison, 2010; Anderson et al., 2011; Chao et al., 2012). β^+ and β can both be converted to the same class of overlap/dissimilarity metrics (see below; Chao et al., 2014; Chao and Chiu, 2016). Using these overlap/dissimilarity indices, Chao and Chiu (2016) recently described a simple mathematical unification of diversity decomposition with the variance partitioning approach (Whittaker, 1972; Legendre and Legendre, 2012; Legendre and De Cáceres, 2013). Given these developments, fields of study concerned with diversity are close to having a conceptually and quantitatively consistent system for describing and analyzing diversity.

Two features of real data complicate practical execution of diversity decomposition. First, a system's diversity is intrinsically related to its size (N, the number of units in the set), making it difficult to interpret comparisons between systems with different N (Chao et al., 2008; Jost et al., 2010). For example, two archipelagos might consist of different numbers of islands. If the average diversity per island (α) was the same between the two archipelagos, and the average difference between islands was same between the two, the archipelago with the greater N must have greater γ diversity and β diversity.

Second, real data are usually comprised of samples raising questions about sample-based estimates of parameters or descriptive statistics intended to describe actual systems. The diversity of a sample is known to underestimate the diversity of a unit (α) the sample was taken from because rare species are often missed. A number of workable corrections are available for α to estimate unseen species (e.g., Chao et al., 2006, 2013, 2014, 2015). However, currently there is no generalized solution, and this remains an active area of research across disciplines (Hausser and

Strimmer, 2009; Chao and Jost, 2015; Marcon et al., 2015). An additional and underappreciated issue is that sampling fewer than the 'true' number of units underestimates the total diversity of the set (γ). Together, these issues conspire to make the relationship between sample β diversity and true β diversity unclear.

For many kinds of ecological data, discrete survey units (e.g., transects or quadrats) often do not correspond to natural subdivisions of the system. For example, the famous Barro Colorado Island tree data (Condit et al., 2002) are comprised of 50 plots within a single continuous tropical forest. The Doubs fish dataset (Verneaux, 1973; Verneaux et al., 2003) used as an example by Borcard et al. (2011) consists of 30 sample sites along a continuous river. The coral data used as an example for beta diversity analysis by Anderson et al. (2011) and Chao and Chiu (2016) come from transects across contiguous reef flats (Brown et al., 1990; Warwick et al., 1990). The bulk of Whittaker's 1960 data from his seminal paper introducing the term "beta diversity" consists of samples across arbitrary subdivisions of various gradients. For these kinds of study systems that are not naturally organized as units within a set—unlike the archipelago example above—there is no "true" N, and no "true" alpha or beta diversity. Nevertheless, ecologists use diversity decomposition to describe the diversity and spatial heterogeneity of these systems (e.g., Kraft et al., 2011). Unfortunately, naive calculation of alpha and beta diversities results in descriptive statistics that are peculiar to the sampling design.

The remainder of this paper presents an intuitive and general solution to the problems with calculating β diversity. The approach is applicable regardless of the particular bias correction used for α diversity. For systems intrinsically organized as units within a set, average pairwise dissimilarity is a natural and intuitive description of the among-unit component of diversity. For continuous systems with arbitrary survey units, pairwise dissimilarity remains dependent on the grain of the sampling design, but we argue that it better represents heterogeneity than an N-community measure that also varies with N.

Three subtly different transformations of β have been proposed for comparing systems with different N. Jost (2007) suggested "turnover rate per sample:"

$$T_{q,N} = ({}^{q}D_{\beta} - 1)/(N - 1). \tag{3.1}$$

Chao and Chiu (2016) prefer the following dissimilarity metrics because they also relate to the variance partitioning framework. "Local" or "Sørensen-type" dissimilarity is

$$1 - C_{q,N} = 1 - \frac{(1/{}^{q}D_{\beta})^{q-1} - (1/N)^{q-1}}{1 - (1/N)^{q-1}} \tag{3.2a}$$

and represents the effective average proportion of species within a unit that are not ubiquitous. Note that when $q = 0$, this measure is equal to Jost's turnover ($1 - C_{0,N} = T_{0,N} = \frac{1}{N-1}\frac{{}^{0}D_{\beta}^{+}}{{}^{0}D_{\alpha}}$). "Regional" or "Jaccard-type" dissimilarity,

$$1 - U_{q,N} = 1 - \frac{(1/{}^{q}D_{\beta})^{1-q} - (1/N)^{1-q}}{1 - (1/N)^{1-q}}, \tag{3.2b}$$

is intended to describe the effective average proportion of species that are not ubiquitous (Chao and Chiu, 2016). When $q = 0$, this measure reduces to $1 - U_{0,N} = \frac{N(1-1/{}^{0}D_{\beta})}{N-1} = \frac{N}{N-1}\frac{{}^{0}D_{\beta}^{+}}{{}^{0}D_{\gamma}}$. Both the local and regional measures are identical for $q = 1$ and are calculated as:

$$1 - C_{1,N} = 1 - U_{1,N} = \frac{\log({}^{1}D_{\beta})}{\log(N)} \tag{3.2c}$$

(Chao and Chiu, 2016). All three of these measures scale ${}^{q}D_{\beta}$ between zero and one. These metrics take the value of zero when all communities are identical and the value of one when all communities are entirely distinct.

Yet each of these N-community dissimilarity measures is influenced by both ${}^{q}D_{\beta}$ and N. This is not a sampling issue: an N_1-community measure is not the same theoretical quantity as an N_2-community measure unless $N_1 = N_2$. For example, even if we had perfect comprehensive

information with which to calculate (not estimate) *N*-community dissimilarities, one cannot meaningfully compare a true 137-island dissimilarity for Hawaii with a true 7,107-island community dissimilarity for the Philippines or a true 33-island community dissimilarity for New Zealand.

It follows that when N is a number of *sampled* communities (or sampling units) rather than a *true* number of communities in a set, eq. 3.1 and 3.2 are biased estimators of the "true" values (if such exist). Moreover, we find it misleading to characterize eq. 3.1 as a rate per sample because it is not a consistent estimator of an expected difference between two samples (i.e., setting N=2). Rather, eq. 3.1 and 3.2 define set-wide (or N-community) measures that are not consistently estimated from a sample of communities or survey units.

Many investigators have recognized that similarity-based measures of composition are biased by sample completeness *within* samples due to finite sampling (i.e., missing rare species that were unobserved but actually present). Several corrections have been proposed, but none are perfect (e.g., Wolda, 1981; Chao and Shen, 2003; Chao et al., 2006; Hausser and Strimmer, 2009; Marcon et al., 2015). In any case, those corrections account only for within-sample bias. Beta diversity, turnover, and N-community dissimilarity measures are also affected by the number of samples taken.

Here we use simple simulations to demonstrate the sample-size dependence inherent in measures of dissimilarity. We advocate a solution in the form of *average pairwise* dissimilarity to avoid confusion owing to unequal sample sizes. This approach simply uses pairwise dissimilarity instead of N-community dissimilarity. For any Hill number, we take the average of a matrix of pairwise dissimilarities calculated according to eq. 3.1 and 3.2. Average pairwise dissimilarity is an intuitive measure of heterogeneity because it estimates the expected difference between a random pair of survey units. For example, the pairwise Sørensen-type dissimilarity represents the effective proportion of species in one unit that will not be found in a second unit. We show that average pairwise metrics give consistent answers for any sample size.

3.1.1 Conceptual examples

To illustrate some of the problems with interpreting eq. 3.1 and 3.2 as general descriptors of beta diversity, consider a simple presence-absence table where one species is uniquely missing from each site.

species	A	B	C	D	E
site 1	0	1	1	1	1
site 2	1	0	1	1	1
site 3	1	1	0	1	1
site 4	1	1	1	0	1
site 5	1	1	1	1	0

The richness of each site (α diversity, $q = 0$) is four species and the richness of the system (γ diversity) is five species, so multiplicative β diversity is $\gamma/\alpha = 5/4 = 1.25$. In this case, these numbers do not change when fewer than five sites are sampled (any two sites captures the full five species present in the system). Therefore Sørensen dissimilarity and Jost's turnover change according to $\frac{1}{N-1}$ and Jaccard dissimilarity changes according to $\frac{N}{N-1}$. For any sample of two sites eq. 3.1 and 3.2 give pairwise turnover (Sørensen dissimilarity): $T_{0,2} = 1 - C_{0,2} = \frac{1.25-1}{2-1} = \frac{1}{4}$ and pairwise regional (Jaccard) dissimilarity: $1 - U_{0,2} = \frac{2(1-1/1.25)}{2-1} = \frac{2}{5} = 0.4$. As we increase the number of sites sampled Sørensen dissimilarity has a larger denominator, so it decreases:

$$T_{0,3} = \frac{1}{8}$$

$$T_{0,4} = \frac{1}{12}$$

$$T_{0.5} = \frac{1}{16}$$

Likewise, $N/(N-1)$ decreases with increasing N, so Jaccard dissimilarity decreases with number of sites sampled:

$$1 - U_{0,3} = 0.300$$

$$1 - U_{0,4} = 0.2\overline{66}$$

$$1 - U_{0,5} = 0.250$$

Clearly, dissimilarity metrics calculated from a sample of $n < N$ sites do not estimate the same quantities as dissimilarity metrics calculated for the full N sites in the system. Moreover, only in the pairwise case ($n = 2$) do the intuitive interpretations of dissimilarity make sense. The Sørensen dissimilarity ($T_{0,2} = 1/4$) indicates that one out the four species present within a site is not expected to be present in a second site, and the Jaccard dissimilarity ($1 - U_{0,2} = 2/5$) indicates that two out of the five species present in a pair of sites are not present in both sites.

The presence-absence matrix above makes the calculations transparent, but it is a rather idealized scenario. As a more generalizable case, consider an example with Lucky Charms cereal with four kinds of marshmallows, each which is equally abundant (relative frequencies all $= 1/4$). If we draw a random sample of three marshmallows, getting all four kinds is impossible. We might get three of a kind (with probability $4 \times \left(\frac{1}{4}\right)^3 = \frac{1}{16}$), we might get three distinct kinds (with probability $\frac{3}{8}$), or we might draw two of one kind and one of another (the most likely outcome, with probability $= \frac{9}{16}$; see online appendix S1 for more detail). As a result, the expected richness of a sample (α diversity with $q = 0$) is about 2.3 distinct kinds of marshmallow. If you draw a second independent random sample, the expected pooled richness (γ diversity) for the pair (six marshmallows total) is about 3.3 unique kinds. Therefore the expected pairwise multiplicative β diversity is about 1.4, expected turnover (Sørensen dissimilarity, $T_{0,2}$) is 0.4, and expected Jaccard dissimilarity is about 0.571. If you draw a third independent random sample of three marshmallows, expected γ richness increases to about 3.7 and as α remains the same, expected β increases to 1.6, expected Sørensen/turnover decreases to 0.3, and expected Jaccard dissimilarity decreases to about 0.563. The limits as the number of independent samples approaches infinity are $\gamma = 4$ kinds of marshmallows overall, $\alpha \approx 2.3$ kinds per sample, and $\beta \approx 1.73$. Expected ∞-sample Sørensen dissimilarity approaches zero (division by infinity in equation 1) and the expected

∞-sample Jaccard dissimilarity approaches 0.42. Again, the N-sample dissimilarity statistics are expected to change as a function of the sampling design, making them difficult to interpret as fundamental measures of heterogeneity of the underlying system. The lessons of the Lucky Charms example apply to any study in which increasing the number of survey units continuously increases the probability of encompassing the full γ diversity of the system.

3.2 Methods

To help clarify the issue inherent in applying N-community dissimilarities to samples, we used the Barro Colorado Island (BCI) dataset (Condit et al., 2002) in the *vegan* R package (Oksanen et al., 2016; R Core Team, 2016). This dataset includes 50 plots with counts of 225 tree species. For this example, we assume the dataset describes the "true" tree species diversity of the island and that the system is truly organized as a set of 50 units. We simulated a sampling process by randomly drawing a subset quadrats with different sampling intensities (N = 5, 10, 20, 30, 40). For each sampling intensity, we performed 1,000 replications and compared the N-community measures of dissimilarity (turnover [eq. 3.1], local [eq. 3.2a], and regional [eq. 3.2b]) against the "true" dissimilarity measures of the full dataset for diversity orders q = 0, 1, 2.

To illustrate the value of the pairwise approach, we also calculated the average of the pairwise metrics for each subset for the same measures of dissimilarity as above and compared them to the averages for the full dataset. All of the code is provided as an online supplement (Data S1–S2).

3.3 Results and Discussion

Figure 3.1 illustrates how the N-community measures of dissimilarity vary with N. If using a sample of N_{sample} units to estimate a community-wide measure for a system containing N_{true} units, the N_{sample} community dissimilarity tends to over- or underestimate the N_{true} community

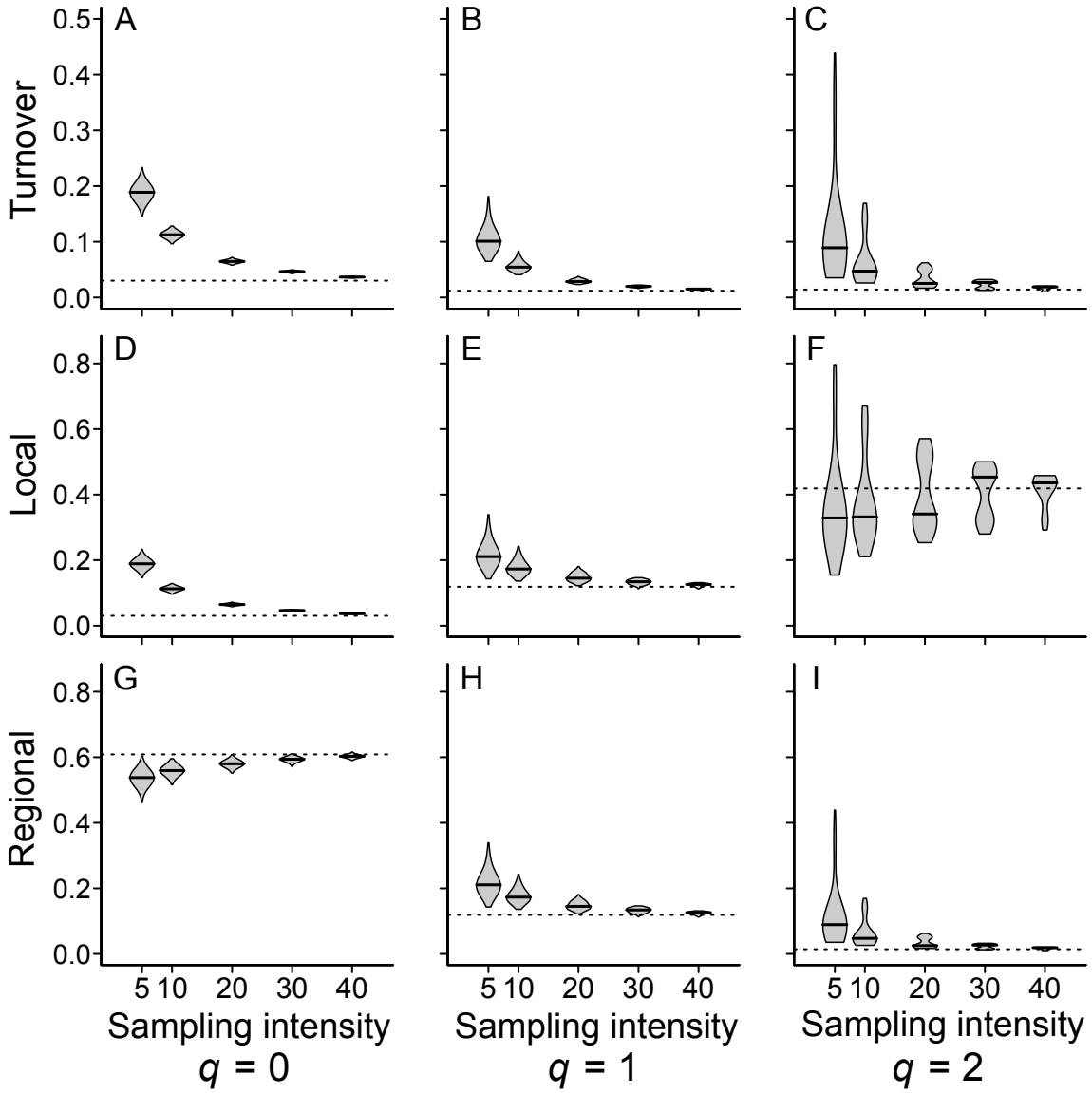


Figure 3.1: Violin plots of the relationship between sample size and three commonly used N-community dissimilarity metrics—turnover (A–C), Sørensen-type (local; D–F), and Jaccard-type (regional G–I)—for three orders of diversity (q). We resampled the BCI dataset 1000 times at various sampling intensities (i.e., 5, 10, 20, 30, 40) and compared the distribution of dissimilarities to the "true" dissimilarity of the entire dataset ($N = 50$; dotted horizontal lines). The plots show the median (black bar) and kernel density estimator of the resampled distribution of dissimilarity metrics.

dissimilarity depending on the metric and the order q. The degree of over- or underestimation is highly dependent on the number of samples or survey units. As sample completeness increases all three metrics begin to converge on the "true" community dissimilarity. In contrast, regardless of sample size, the expected value of average pairwise dissimilarity does a good job of estimating the average pairwise dissimilarity for the full dataset with no systematic bias (fig. 3.2). As with the N-community measures, sampling variance increases with decreased sampling intensity.

For the purposes of this example, we assumed that the 50 plots were true subunits of the BCI system. But like most systems, BCI is not actually composed of discrete units; it is a contiguous tropical forest. The plots in this dataset (Condit et al., 2002) are a property of the sampling design, not a property of the system. In systems like this, if beta diversity is intended to represent spatial heterogeneity, conclusions regarding spatial heterogeneity from N-sample dissimilarities will be affected by the number of sampled survey units in addition to their size and the distance between them. In contrast, the pairwise approach is only sensitive to the latter.

Pairwise approaches have been used for decades in population genetics and ecology, and have sound mathematical foundations (e.g., Wright, 1943; Bray and Curtis, 1957; Morisita, 1959; Horn, 1966; Rogers, 1972; Watterson, 1975; Wolda, 1981; Nei, 1987). The average pairwise turnover or dissimilarity has an intuitive interpretation as the average proportional change in composition from sample to sample. A pairwise dissimilarity matrix can be analyzed in various ways, such as ordination (Legendre and Legendre, 2012), visualizing the distribution or variance in pairwise differences (e.g, Slatkin and Hudson, 1991; Rogers and Harpending, 1992; Schneider and Excoffier, 1999), or regressing pairwise distance against geographic or temporal distance (e.g, Wright, 1943; Smouse et al., 1986; Rousset, 1997; Koizumi et al., 2006). Indeed, many biologists are explicitly interested in spatial turnover (Nekola and White, 1999; Anderson et al., 2011; Legendre and Gauthier, 2014)—i.e., expected community dissimilarity between sites along gradients or across landscapes.

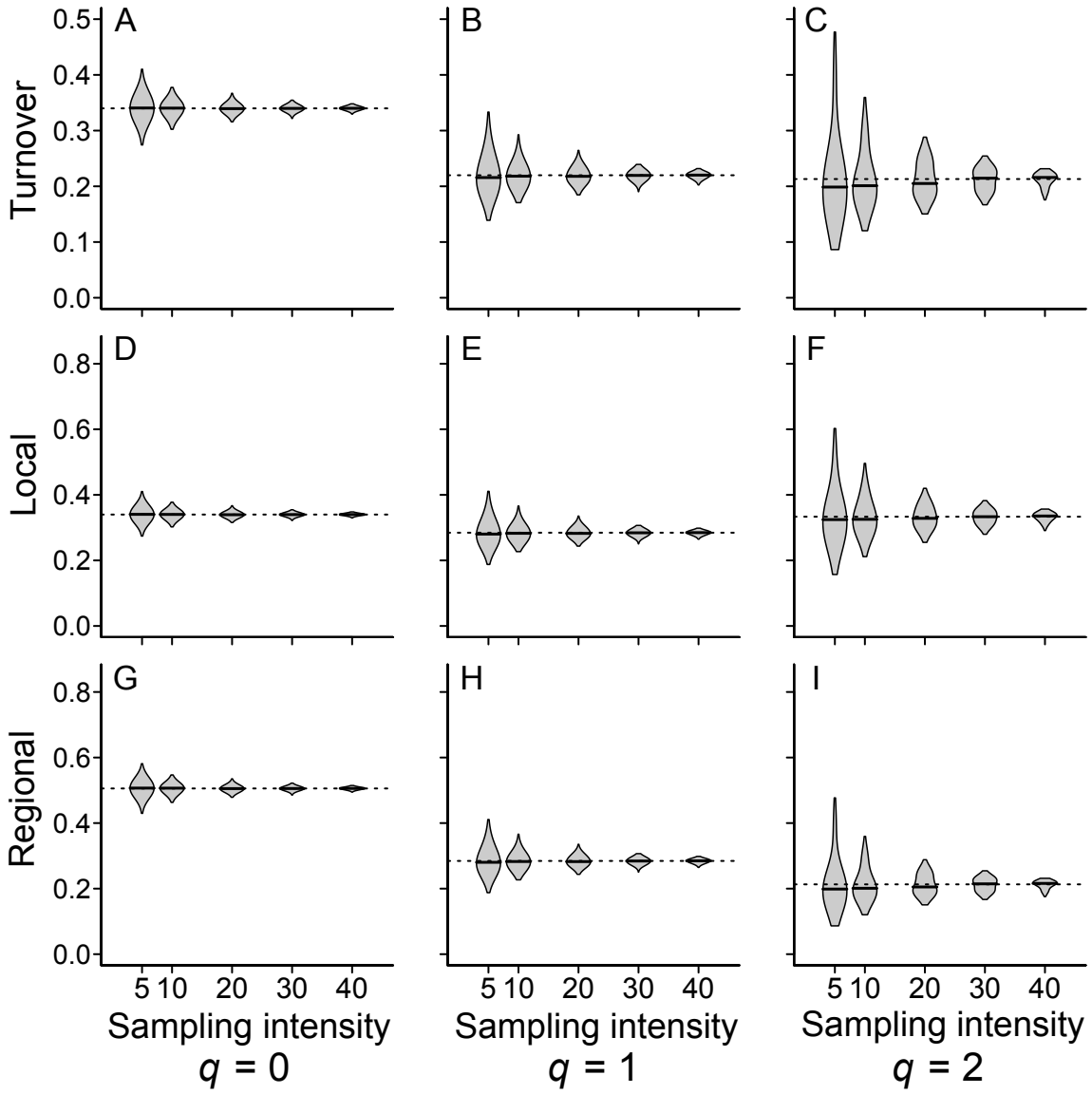


Figure 3.2: Violin plots of the relationship between sample size (N = 5, 10, 20, 30, 40) and the average pairwise (2-community) dissimilarity metrics—turnover (A–C), Sørensen-type (local; D–F), and Jaccard-type (regional G–I)—for q = 0, 1, & 2. The dissimilaries from the same resampled communities in Figure 3.1 were compared to the average pairwise dissimilarity of the full BCI dataset (N = 50; dotted horizontal lines). The plots show the median (black bar) and kernel density estimator of the resampled distribution of dissimilarity metrics.

We acknowledge that average pairwise dissimilarities do not capture all of the information in the data. For example, Baselga (2013) illustrated how systems with the same average pairwise dissimilarity might differ in other aspects of heterogeneity—e.g., different overall gamma diversities or among-pair variances. But, as we have shown, simply using N-community measures can be misleading. N-community measures can be compared only if N is standardized. In principle, any N could be used as a standard, but we question how intuitive standardizing $N > 2$ would be. A rarefaction curve of beta (similar to fig. 3.1) would address the issues raised here and by Baselga (2013). Nevertheless, we find standardizing $N = 2$ to be intuitive because average pairwise dissimilarity estimates the expected dissimilarity between one sample and the next—generally how we interpret turnover.

Here we have illustrated that average pairwise dissimilarity is an easily interpretable description of among sample heterogeneity compared to N-community measures, which are not comparable among systems and vary systematically with sampling intensity. Pairwise dissimilarity among samples is intuitive regardless of the grain at which heterogeneity is described, whether it be replicate samples from a single community, along an environmental gradient, or at larger scales, among vastly different communities. The pairwise approach is valid regardless of one's choice of bias correction for within-sample alpha diversity. This intuitive framework for describing disparity has a long history and broad applicability in community ecology and beyond.

Acknowledgments

We thank N. Sanders, N. Gotelli, C. Nice, T. Paulson, R. Wooliver and two anonymous reviewers for discussion and comments. This research was funded by the National Science Foundation (DDIG award no. DEB-1405887 to Z. H. Marion and DEB-0614223 & DEB-1050947 to James A. Fordyce) and the Department of Ecology & Evolutionary Biology at the University of Tennessee (Z. H. Marion).

3.4 Supporting information

3.4.1 Appendix S1: Extended notes on the Lucky Charms example

To obtain expected richness (number of distinct kinds of marshmallows) in a random sample of three, consider the three possible outcomes.

1. Three of a kind (richness = 1): The probability of getting three of one particular kind (say, pink hearts) is $\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$. There are four ways to get three of a kind, so the probability is $4 \times \left(\frac{1}{4}\right)^3 = \frac{1}{16}$.

2. Three unique kinds (richness = 3): This is the probability that the second is not the same kind as the first (e.g., the probability of not getting a yellow moon = 3/4) times the probability that the third is not the same kind as either the second or first (e.g., 1/2 of the marshmallows are neither yellow moons nor orange stars). $\frac{3}{4} \times \frac{1}{2} = \frac{3}{8}$.

3. Two of a kind (richness = 2): By subtraction the probability is $\frac{9}{16}$.

The expected richness (sum over each outcome times its probability) is

$$\mathsf{E}(\alpha) = 1 \times \frac{1}{16} + 2 \times \frac{9}{16} + 3 \times \frac{3}{8} = 2.3125;$$

i.e., about 2.3 distinct kinds of marshmallow per sample.

To get the expected γ richness for two independent random samples of marshmallows, we could work out probabilities of all possible outcomes (one, two, three, or four kinds) for a random sample of six, but it is slightly simpler to work out the probability that each kind is present in the pool and estimate the expected richness (number of presences) as a binomial(n, p) random variable with $p =$ probability a kind is present, and $n =$ the number of kinds (four). First, given the three possible outcomes (above) for a random sample of three marshmallows, the probability

that any particular kind is absent from a single sample is

$$\left(\frac{3}{4}\times\frac{1}{16}\right)+\left(\frac{1}{2}\times\frac{9}{16}\right)+\left(\frac{1}{4}\times\frac{3}{8}\right)=\frac{27}{64}=0.421975$$

so the probability that any particular kind is present in two independent samples is 1 - (the probability that it is absent from $both$) $= 1-(27/64)^2 = 0.822$, and the expected pooled richness is then 4 times this probability of presence:

$$\mathsf{E}(\gamma)=4\times\left[1-\left(\frac{27}{64}\right)^2\right]=3.288086$$

If you draw a third sample, the probability of absence becomes $(27/64)^3$, leading to an expected γ of

$$\mathsf{E}(\gamma_{N=3})=4\times\left[1-\left(\frac{27}{64}\right)^3\right]=3.69961$$

Increasing the number of samples increases the exponent and the term in square brackets get closer and closer to 1.00, hence the limit as N approaches infinity of $\mathsf{E}(\gamma_{N=\infty})=4$.

Chapter 4

A hierarchical Bayesian model to partition diversity while accounting for incomplete sampling

This chapter is coauthored with James A. Fordyce and Benjamin M. Fitzpatrick to be submitted to *Methods in Ecology and Evolution*.

Abstract

The use of Hill or effective numbers to partition the total diversity (γ) of species within a region into within- (α) vs. among-sample (β) components is becoming increasingly popular for describing community diversity as well as multivariate traits. Using this approach, the effective number of species is dependent on how differential abundances are weighted. Yet despite the conceptual progress, estimating and partitioning diversity based on real data remains a challenge because of finite sampling limitations, and a general understanding of the challenges and pitfalls of estimating desired quantities from real data has lagged behind.

Here we apply a hierarchical Bayesian approach to the estimation of species abundances within and among samples, communities, and regions. We then use posterior predictive samples and the underlying parameter estimates of the abundances themselves to partition diversity for any order of q. Moreover, this framework allows for multilevel hierarchies and ecological hypothesis testing using Bayesian model selection. Using simulations, we demonstrate how this approach mitigates problems of finite sampling due to unequal sample size or sampling effort. Finally, we provide an empirical example to highlight how the method can be applied in practice.

4.1 Introduction

One cornerstone of ecology is comparing and explaining the genomic (e.g., Ihmels et al., 2007), phenotypic (e.g., Marion et al., 2015*c*), or community diversity of natural systems (e.g, Macarthur, 1965; Patil and Taillie, 1982; Torsvik and Øvreås, 2002). This naturally leads to important questions about how diversity should be quantified in theory and estimated in practice. Many classical ecological theories address species richness (MacArthur and Wilson, 1967; Pianka, 1974). Others incorporate the concepts of evenness or probability of identity (Simpson, 1949; Preston, 1962*a*,*b*; Hubbell, 2001)—analogous to homozygosity in population genetics. Key questions have to do with changes in diversity over time (Menge and Sutherland, 1976; Stegen et al., 2013; Legendre and Gauthier, 2014), dependence of diversity on area (Preston, 1962*a*; Whittaker et al., 2001; Kraft et al., 2011), relationships between diversity and environmental factors (Loreau et al., 2001; Ricklefs, 2006; Chao et al., 2016), and how diversity is partitioned within and between communities (Whittaker, 1960; Anderson et al., 2011; Chao et al., 2014). There has been substantial progress regarding theoretical metrics and their decomposition into within- and among community components (Legendre and De Cáceres, 2013; Chao et al., 2014). Yet, a general understanding of the challenges and pitfalls of estimating desired quantities from real data has lagged behind. Here we attempt to move the empirical side of the field forward with a Bayesian approach that helps address difficult problems of bias and uncertainty in estimating, partitioning, and comparing diversity.

Recently, there has been a general consensus that metrics based on Hill numbers (Hill, 1973; Jost, 2006) are intuitive and appropriate for ecological questions (see Ellison, 2010). Hill number diversities (${}^qD$) are mathematically related metrics that vary in their sensitivity to differential abundances according to the parameter q. When $q = 0$, only presences or absences matter (${}^0D$ = average richness or number of species present within a unit). When $q = 1$ (${}^1D$ = exponential of Shannon entropy), species are equally weighted by their abundances. When $q \geq 2$

(^2D = inverse of Simpson's concentration), then $^qD_\alpha$ depends increasingly strongly on the most abundant species; rare species have less influence.

However, despite this conceptual progress, estimating and partitioning diversity based on real data remains a challenge because of finite sampling limitations. First, it is well established that the diversity of a sample tends to underestimate the true diversity of the underlying unit from which the sample was taken (Gotelli and Colwell, 2001; Paninski, 2003; Hausser and Strimmer, 2009; Chao et al., 2014). This arises in part because rare species are often missed unless all individuals within a patch or community are found and identified. But, Hill numbers are still systematically underestimated even when all species are accounted for because they are concave functions of proportions (a consequence of Jensen's Inequality, see Paninski, 2003). Authors in a variety of fields have proposed bias corrections, but no general solution is currently known (Miller, 1955; Chao and Shen, 2003; Hausser and Strimmer, 2009; Chao et al., 2013).

Second, when partitioning diversity among sampling units (such as islands, ponds, or other kinds of discrete patches), the among-sample diversity component (β diversity) and measures of overlap or turnover derived from it are dependent on the number of sampling units. This means the β diversity of an archipelago of 33 islands cannot be compared with the β diversity of an archipelago of 42 islands without first standardizing to a fixed number of islands. Likewise, the β diversity of any set of habitat patches (such as an archipelago) cannot be estimated from a subsample of the patches. Elsewhere, we have suggested that the average (or expected) pairwise turnover is the most intuitive way to express the among-unit component of diversity in a manner that can be compared across systems and studies (see chapter 3).

Third, many ecological datasets are comprised of discrete samples—such as transects or quadrats—from contiguous systems. For example, the Barro Colorado Island tree data of Condit et al. (2002) consist of 50 plots within a single continuous tropical forest. The attributes of such plots are peculiar to the sampling design, not intrinsic characteristics of the contiguous

system. Nonetheless, ecologists use diversity decomposition estimates from discrete samples to learn about the diversity and heterogeneity of the underlying natural system of interest.

Finally, although multiple samples are often collected, the uncertainty associated with sampling is often ignored and only a single point estimate of diversity is reported. When error is considered, the uncertainty intervals around the point estimate are usually generated through resampling or parametric bootstrapping (Efron, 1982; Efron and Tibshirani, 1986). Parametric bootstrapping (using the multinomial distribution, for example) is often within-sample. This assumes the total abundance of a site is fixed or known with certainty. Alternatively, nonparametric bootstrapping usually relies on resampling of sites, assuming the samples are fixed or true. In both instances, uncertainty intervals are limited by sample size and effort. Thus, ecologists end up describing the sample diversity rather than diversity of the underlying system of actual interest. Another challenge is that ecologists are usually interested in comparisons—e.g., is this island different from that island, or is this mountain range different from that mountain range. This is a fundamentally statistical problem. Quantifying the uncertainty around diversity estimates is a start, but—to date—an integrative statistical approach has been lacking.

Here, we present a framework for statistical analysis that addresses all of these challenges. We advocate using hierarchical Bayesian models to explicitly model the uncertainty in species abundances. Our approach can accommodate problems of finite sampling such as unequal sample sizes or sample effort. Moreover, multilevel hierarchies are possible. We can then use model comparison through WAIC, leave-one-out/K-fold cross-validation (Watanabe, 2010; Vehtari et al., 2016*b*) to assess whether patches/communities/habitats within regions or sets are distinct subcommunities of a metapopulation or whether they are "arbitrary" distinctions from one contiguous system.

4.2 Methods

4.2.1 Conceptual overview

Our modeling framework begins with a sample $\times$ species matrix where elements are species abundances. We then estimate the species abundance vector of each sample using a Bayesian model. As a general overview of how the Bayesian model works (Gelman et al., 2013):

1. Initial parameter values are proposed: for each sample, we have a set of multinomial probabilities corresponding to each species. Parameter $\theta_{i,k}$ is the probability an individual found in site i is classified as species k.

2. The likelihood of the data given the parameters is multiplied by the prior probability of the parameters.

3. This gives us a posterior probability estimate of the multinomial $\theta's$—interpreted as the relative abundances of species for an observation. For a given iteration, the parameters are then saved or rejected according to the Markov Chain algorithm.

4. New parameter values are proposed, and the process repeats itself for a set number of iterations.

5. Because the first iterations constitute the adaptive phase of the sampler, these "warm-up" or "burn-in" estimates are discarded (in our case, the first half of each chain). The remaining iterations comprise the posterior distribution of relative abundances. Because information is shared among observations through the prior, we can interpret the multinomial $\theta's$ as estimates of the true underlying relative abundances.

6. For each iteration of the posterior, the multinomial $\theta's$ are used post-hoc to estimate and partition the diversity. In principle, the posterior $\hat{\theta}'s$ can be plugged into formulae for Hill numbers or any diversity metric with an ad-hoc bias correction or not. Because we have

a posterior distribution of relative abundances, we also have a posterior distribution of diversities that incorporate uncertainty.

The multinomial $\theta's$ are probabilities of occurrence for a species. Even though a species may never be observed, that probability will never be truly zero. In other words, this model assumes we have a comprehensive list of species in the true community, so we assume the true richness is known. The problem of estimating richness has been addressed in several ways (Colwell and Coddington, 1994; Gotelli and Colwell, 2001, 2010). Many of these can be incorporated with our approach in an *ad hoc* manner (see Discussion), but simultaneous estimation of richness (i.e., the "support size" of a multinomial distribution) and relative abundances (the event probability parameters of the multinomial distribution) requires more assumptions and/or more information than the framework described here. Investigators in ecology often use rarefaction to characterize expected richness in a sample (Chao et al., 2014; Chao and Jost, 2015). Here, we use posterior predictive simulations to estimate expected richness (or other metrics) given a sampling scheme as follows:

7. For each iteration, we simulate sample data from the posterior distribution of multinomial $\theta's$.

8. The simulated sample data is saved and then used to estimate diversity as well.

Thus we make a distinction between estimation of the underlying community diversity (i.e., the real world) and estimation of the sample diversity (i.e., the collected data).

4.2.2 The Bayesian model

Consider a dataset of N samples (e.g., quadrats or transects), with S species from 1, 2, ..., S in total among samples. Let $\mathbf{Y} = [y_{i,k}] \geq 0$ be an $N \times S$ community abundance matrix:

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,S} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N,1} & y_{N,2} & \cdots & y_{N,S} \end{bmatrix} \tag{4.1}$$

where $y_{i,k}$ is the count of species k in the i-th sample, and $\sum \vec{y_i} = \sum_{k=1}^{S} y_{i,k} = n_i$. We assume that the vector of species abundances for observation i follow a multinomial distribution:

$$\vec{y_i} \sim \text{Multinomial}(\vec{\theta_i}, n_i) \tag{4.2}$$

where n_i is the total abundance of individuals for observation i and the parameter vector $\vec{\theta_i}$ is a k-simplex (i.e., a k-length vector with non-negative values that sum to one) describing the probability of species occurrences for observation i. Perhaps more intuitively, we can interpret $\theta_{i,k}$ as the relative abundance of species k in observation i.

For each $\vec{\theta_i}$, we use softmax normal priors with mean μ_k and common standard deviation σ. The softmax transformation takes a vector (α_i) with unconstrained (i.e., $-\infty$ to $+\infty$) values and constrains it to a k-simplex. Therefore the model is

$$\begin{aligned} \vec{\theta_i} &= \text{softmax}(\vec{\alpha_i}) \\ &= \frac{\exp(\alpha_{i,k})}{\sum_{k=1}^{S} \exp(\alpha_{i,k})} \end{aligned} \tag{4.3}$$

$$\vec{\alpha_i} \sim \text{Normal}(\vec{\mu}, \sigma). \tag{4.4}$$

The softmax function is many-to-one, which can lead to problems of nonidentifiability for the unconstrained $\alpha's$. However, only $K-1$ parameters are actually necessary to parameterize a K-simplex because the Kth is one minus the sum of the other $K-1$ parameters. Therefore we

mitigate the issue of non-identifiability by fixing the first component of α_i to zero. We assume the prior vector of means ($\vec{\mu}$) follow a normal distribution, and the prior standard deviation σ is half-Cauchy distributed:

$$\vec{\mu} \sim \text{Normal}(0, 5) \tag{4.5}$$

$$\sigma \sim \text{Cauchy}^{+}(0, 2.5). \tag{4.6}$$

For computational efficiency we use a non-centered parameterization of the normal distribution (Papaspiliopoulos et al., 2007), and the priors for both μ and σ are weakly informative as recommended by Gelman et al. (2008).

The untransformed priors for the multinomial $\theta's$ are not immediately interpretable because they are normally distributed between $-\infty$ and $+\infty$ and, because of nonidentifiability, are a $K-1$ length vector because the first component of each $\vec{\alpha_i}$ is fixed at zero. However, the $\mu's$ estimated for an assemblage can be easily converted to relative abundances by using the softmax transformation—i.e., $\Theta = \text{softmax}(\vec{\mu})$. As before, fixing the first component to zero prior to the transformation results in the full K length simplex. Transforming the μ hyperparameters into Θ can be thought of as the average relative abundance for each species across all sites in the assemblage, which is often of value (see below).

The main parameters of interest are the posteriors for the multinomial $\theta's$ because these are estimates of the "true" relative abundances for each observation. However, because θ is a vector of probabilities, they will never be zero even though a species may truly be absent (although θ may be very small). This poses a problem for estimating diversities when q is close to zero (e.g., richness). We therefore use posterior predictive simulation to simulate sample data given the posterior parameter estimates at each iteration of the Markov chain. The average abundance of each observation n_i is modeled as Poisson-distributed according to a single λ describing the

mean abundance across samples:

$$n_i \sim \text{Poisson}(\lambda). \tag{4.7}$$

λ is then given a weakly informative half-Cauchy distribution:

$$\lambda_i \sim \text{Cauchy}^{+}(0, 25). \tag{4.8}$$

Then, for each iteration, we used predictive simulation to generate a "new" total standardized abundance $\tilde{n}$ from the posterior estimate of the Poisson λ. This, along with the posterior estimates of $\vec{\theta}_i$, generates a new abundance vector $\tilde{y}_i$:

$$\tilde{n} \sim \text{Poisson}(\lambda) \tag{4.9}$$

$$\tilde{y}_i \sim \text{Multinomial}(\vec{\theta}_i, \tilde{n}). \tag{4.10}$$

We use a single λ to model the total abundance for the posterior predictive simulations as a way to standardize sampling effort among samples. Alternatively, one could use a fixed n instead (e.g., the largest or smallest abundance in the dataset).

For all models in this paper, posterior probabilities for model parameters were estimated using Hamiltonian Monte Carlo (HMC) sampling in the Stan programming language (Carpenter et al., 2016) using the *RStan* Stan Development Team (2016) interface. Four HMC chains were used with 5,000 iterations each. The first 2,500 iterations per chain were adaptive and discarded as warm-up resulting in 10,000 posterior iterations overall. We used several diagnostic tests to confirm that each model had reached a stationary distribution, including visual examination of the HMC chain history, calculation of effective sample size (ESS), and the Gelman-Rubin convergence diagnostic ($\hat{R}$; Gelman and Rubin, 1992; Brooks and Gelman, 1998). In particular, model convergence was assessed by inspecting the diagnostics of the log-posterior density.

4.2.3 Statistical inference: are assemblages distinct?

Often ecologists are interested in assessing whether a community consists of a single contiguous assemblage or a patchwork of distinct subassemblages. Bayesian model selection provides a natural extension to explicitly test such hypotheses by comparing complete-pooling and no-pooling models. A complete-pooling model has a single prior distribution, consistent with the hypothesis that the community of interest is comprised of one contiguous assemblage. In contrast, a no-pooling model assigns a separate prior distribution to each hypothesized subassemblage.

The strengths of the different hypotheses can then be quantified by comparing models using WAIC (the widely applicable or Watanabe-Akaike information criterion; Watanabe, 2010; Gelman et al., 2014), leave-one-out cross-validation (LOO), or k-fold cross-validation (Vehtari et al., 2016*b*). For the purposes of this paper, we use WAIC because its interpretation is similar to AIC—widely used for model comparison in frequentist statistics. WAIC is an improvement over the deviance information criterion (DIC; Spiegelhalter et al., 2002) in that it is fully Bayesian and uses the entire posterior distribution. Thus WAIC provides an estimate of uncertainty for comparison of predictive errors between two (or more) models (Vehtari et al., 2016*b*).

We implemented model comparison by saving the posterior distribution of log-likelihoods from both complete and no-pooling models. We then calculated the difference in WAIC values between models and the predictive errors of the difference using the *loo* package (Vehtari et al., 2016*a*). Like AIC or DIC, there are no firm criteria for determining when one model is unambiguously "better" than another. However, we can use the standard errors that WAIC and LOO provide as strength of the evidence. If the error in predictive difference between models does not overlap zero, we can consider that as strong evidence in favor of one model over the other.

4.2.4 Statistical comparison of assemblage diversities

If model selection suggests that subassemblages should be modeled separately, one might be interested in estimating diversities at multiple levels: within each subassemblage or regionally among subassemblages. At the within-subassemblage level, alpha, beta (or standardized beta), and gamma estimates can be compared by plotting the diversity profiles of the marginal mean (or median) diversity estimates along with their uncertainty intervals such as the 95% highest density interval (HDI) for a range of q.

At the regional scale, we can estimate posterior α, β, or γ similarly using the $\Theta' s$ after softmax transforming the μ priors after softmax transformation. Each Θ is the average relative abundance within assemblage, and the posterior diversity estimates at this scale represent regional diversity across subassemblages.

4.2.5 Simulation study of model efficacy

We evaluated the performance of our method by simulating artificial datasets using an approach from Cayuela et al. (2015). For each iteration, we generated a log-normal species abundance distribution. The log-normal distribution was chosen because—depending on the parameter values—the distribution is right-skewed (few common species and many rare ones), typical of many ecological communities (Preston, 1948). Then we created a "true" community matrix by populating sites/observations with individuals in proportion to the generated species abundance distribution. The total count of individuals in a site were poisson-distributed around a mean number of individuals in the community (generally $\lambda = 500$), while the total number of sites in the "true" community varied between 30–75. We then created an observed community by sampling individuals from each site with sample effort p—the probability of completely sampling every individual—using a random draw from a binomial distribution.

We then used our Bayesian framework to model the abundances of the sampled community under two scenarios: low (12%) and high (80%) sampling effort. For the low sampling effort, 12% represents a situation where the sample size of individuals is the same as the number of species in the community—a known example where there is a problem of sampling bias and underestimation of the true diversity. For each simulation, we partitioned the total diversity (γ) into α and β using the mean of the posterior simulations for $q = 0$, 1, & 2. We did the same for the mean of the multinomial $\theta's$ for $q = 1$ & 2.

For each simulation, the estimated diversities for the Bayesian model were compared to the diversities of the "true" community by taking the difference (Δ diversity = estimated - true). We also calculated Δ diversity using two different approaches as benchmarks. For the first, we used a point ("naïve") estimate of the diversity from the observed community. The second benchmark was similar to the naïve estimate except that we applied a bias-correction (Chao et al., 2013) using the *entropart* R package (Marcon and Herault, 2015).

4.2.6 Simulation study for comparing assemblages

To quantify the performance of our approach at identifying distinct subassemblages, we again simulated communities by modifying code from Cayuela et al. (2015) to assess whether simulated communities were better modeled as one contiguous community or two distinct subassemblages. For each simulated community, the total number of observations N for the combined communities varied randomly between 30–75 as in the simulations described above. The total sample size was then partitioned randomly into two subassemblage sizes with the minimum sample size within a subassemblage fixed at 10. A "true" community was then generated as previously described and then "sampled" with 80% effort to create a true community.

Within this overall simulation framework, we tested three scenarios: 1) one community artificially divided into two subassemblages; 2) two assemblages that differ in species richness but not evenness; 3) two assemblages that differ in evenness but not in species richness. In the

first scenario—representing the null expectation that there is a single contiguous community—we generated one log-normal species abundance distribution comprised of 60 species total and then populated sites in proportion to the relative abundances of those species. We then randomly divided this contiguous community into two as described above. The second scenario was similar to the first except that we generated two log-normal species abundance distributions that were identical except that one had 20 and one had 60 species, respectively. Note that the richness levels refer to the species abundance distributions from which the subassemblages were drawn. The differences in richness among sampled community matrices varied. In the third scenario, the rank abundance distributions of the two subassemblages had a fixed difference in log standard deviation of 1 unit and ranged randomly between 0.095–3.5 on a log scale.

Each simulation scenario was run for 500 iterations each. For each iteration, we calculated the difference in WAIC values between a one-assemblage (complete-pooling) and two-assemblage (no-pooling) model. If the one-assemblage model was more plausible, the WAIC difference were negative; if the two assemblage model was favored, the difference was positive. We then assessed the performance of our approach with three benchmarks from most to least conservative. We asked what was the proportion of iterations per scenario: 1) that unambiguously returned the correct result (i.e., more than one standard error from zero and correct in sign); 2) for which the point estimate of the difference was correct but was within one standard error of zero; 3) for which the point estimate was incorrect but was within one standard error of zero; 4) that unambiguously suggested the incorrect model with more than one standard error from zero and incorrect in sign.

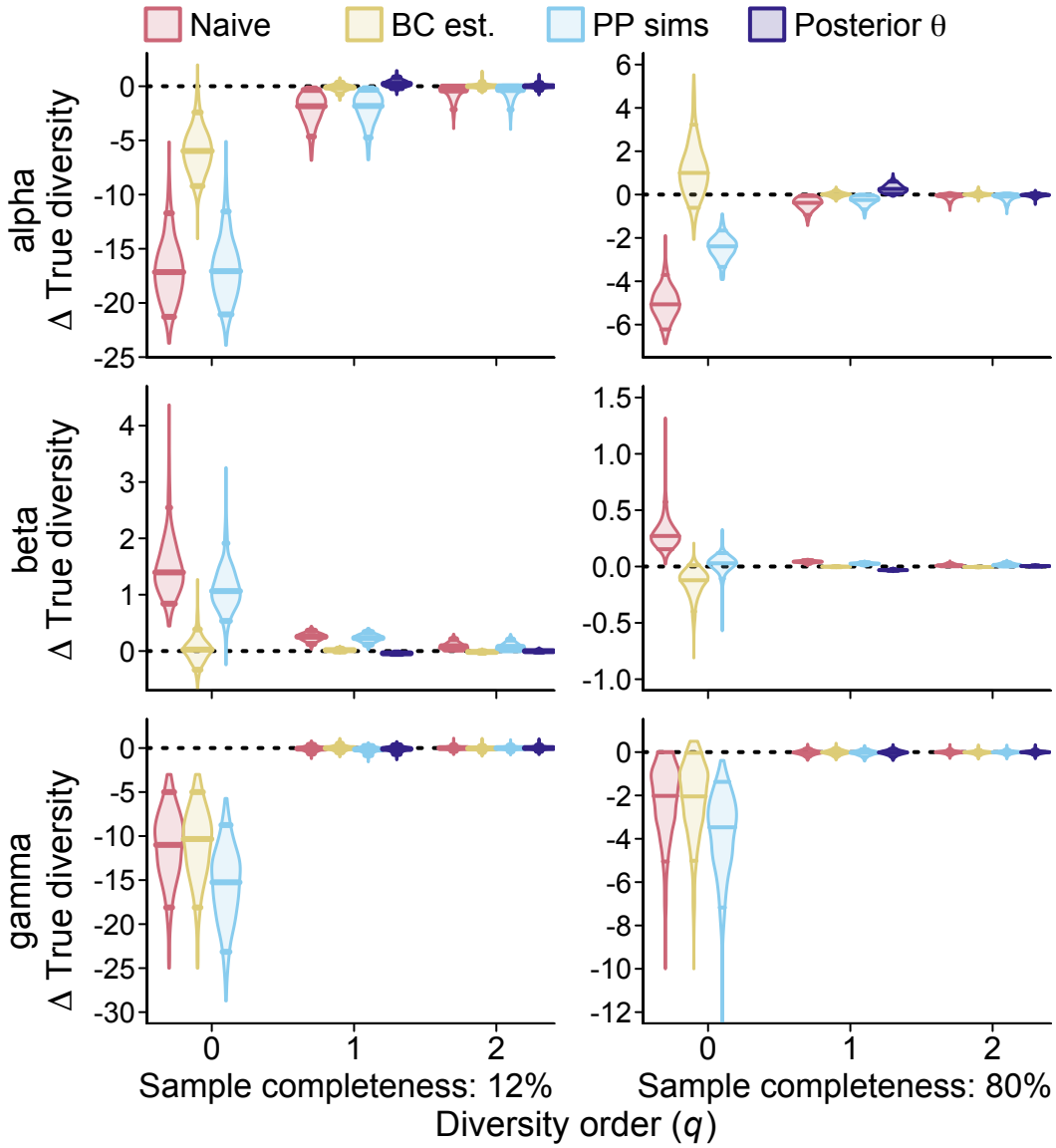


Figure 4.1: Violin plots showing the distribution of deviations from true diversity (Δ True Diversity) under two levels of sampling completeness for $q = 0$, 1, & 2. For each simulation, the mean posterior predictive estimate (blue) and the mean posterior θ (purple) diversity estimates were computed and then subtracted from the true diversity value. We also compared the diversity estimates from the naïve plugin estimator (red) and an estimate that corrects for unseen species (yellow). Δ diversity greater than zero overestimate the true diversity, while Δ less than zero underestimate the true diversity; values at zero (dotted lines) are correct estimates. A sample completeness of 12% means that, on average, for each site, the same number of individuals were sampled as there were total species in the true community. Violin plot lines indicate the mean and upper and lower 95% uncertainty intervals across simulations.

4.3 Results

4.3.1 Model efficacy simulations

Overall, the model performed well at estimating the true diversity (Figure 4.1). When sample completeness was low, such that the total abundance is close to the number of species, the diversity estimates calculated from the posterior predictive data performed about as well as the naïve plug-in estimator. For $q \geq 1$ however, the posterior θ estimates did a much better job, on par with the performance of the bias corrected estimates. As sample completeness increasesd(i.e., 80%), the estimates from the posterior predictive data generally outperformed the naïve estimates due to "information sharing" via the prior. As before, the diversities from the posterior θ's did an excellent job at estimating the true diversities for $q \geq 1$.

4.3.2 Comparing assemblages

Model selection identified the correct model the vast majority of the time (Table 4.1). Across the three scenarios, WAIC strongly supported the incorrect model (i.e., with a standard error of ΔWAIC not overlapping zero) for $\leq$ 1% of the 500 simulations. The WAIC point estimate was consistent with the correct model for $\approx$ 97% (one assemblage and different richness) and 98.4% (different evenness). WAIC strongly supported the correct model for 91–94% of simulations that consisted of two different subassemblages, but had a harder time correctly selecting the null for the one-assemblage community simulations (standard error of ΔWAIC not overlapping zero in $\approx$ 80% of simulations).

4.4 Applications: oribatid mite diversity

As an example, we applied our Bayesian statistical framework to estimate the diversities of oribatid mites using the results of Borcard et al. (1992) and described more fully in Borcard

Table 4.1: Results of simulations testing the ability of WAIC to correctly identify when to model a community as one contiguous assemblage or two distinct subassemblages for three scenarios: 1) a null model of a single community randomly divided into two subcommunities; 2) two assemblages that had different richness but similar evenness; and 3) two assemblages that had similar richness but differed in evenness. For each scenario, we quantified performance as the percent (count in parentheses) out of 500 simulations that were strongly correct, the point estimate was correct, the point estimate was incorrect but error overlaped zero, or were strongly incorrect according to Δ WAIC.

Simulation scenario	Strongly correct	Point estimate correct	Point estimate incorrect	Strongly incorrect
Same community	78.4% (392)	18.4% (92)	3.2% (16)	0.00% (0)
Different richness	91.4% (457)	5.2% (26)	2.4% (12)	1.00% (5)
Different evenness	94.4% (472)	4% (20)	1.2% (6)	0.04% (2)

and Legendre (1994). In 1989, *Sphagnum* moss cores (5 cm in diameter and 7cm deep; n = 70) were sampled from a 10 $\times$ 2.5 m moss and peat blanket floating in Lac Geai at the Station de Biologie des Laurentides, Québec. Samples were stratified by substrate types (e.g., *Sphagnum spp.*, substratum morphology, shrub cover) and sampled proportionally to substratum surface area (Borcard et al., 1992). The result was a 70 sample $\times$ 35 species community matrix. Borcard et al. (1992) used canonical ordination to partition the variation in mite species diversity into environmental and spatial components. The authors found that the environmental variables explained $\approx$ 45% of the variation in community structure, and that the species data and environmental variables had similar spatial structure—possibly due to similar underlying mechanisms. However, the goal of these papers was to tease out the roles of spatial and environmental variation in structuring communities, not to quantify the diversity of the mite community itself.

Here, we focus on one habitat characteristic—a shrub density gradient on a semi-quantitative scale (no shrubs $<$ few shrubs $<$ many shrubs)—to highlight the utility of our statistical approach. We chose this characteristic because environment explained much of the mite community

structure, and shrub density accounted for most of the environmental variation according to a later study (Borcard et al., 2004). Because of the amount of variation in mite community structure density explained, we found it appropriate to use WAIC to ask whether to model the community as one assemblage (complete pooling) or different assemblages along a shrub density gradient (no-pooling).

Bayesian model selection supported the three-community model over the single-community model. The posterior difference in WAIC was positive and had a standard error that did not overlap zero (ΔWAIC: 31.92, 23.08 SE). Thus there was strong support for the model that estimated separate parameters to each shrub density category compared to a model that considered the mite community as one contiguous assemblage.

Therefore, we proceeded to partition diversity within and between samples and shrub categories to address the questions of whether mite assemblages differed in diversity and heterogeneity. We calculated alpha and standardized beta diversity within each shrub-density category using the multinomial $\theta's$ and regionally using the prior $\Theta's$ for q between 1–3. We standardized beta as the average pairwise turnover for comparison among the different shrub density levels because of unequal sample sizes (see chapter 3).

As shrub density increased, so did within-sample mite diversity (Figure 4.2). When posterior estimates of the relative abundances were weighted equally ($q = 1$), samples from areas of *Sphagnum* with the highest shrub density had approximately 10.5 mite species on average [95% HDI: 10.18, 11.06] while samples from areas with few to no shrubs had lower diversity with 8 [7.65, 8.34] and 6.7 [6.26, 7.09] effective species, respectively. At the regional level, samples had 9.68 [8.48, 10.89] effective species on average for $q = 1$, and intermediate estimate across the diversity gradient. Within shrub level and assemblage-wide, the profile plots decreased rapidly as diversity order increases, suggesting that the mite community is comprised of a few common taxa and several species of low abundance.

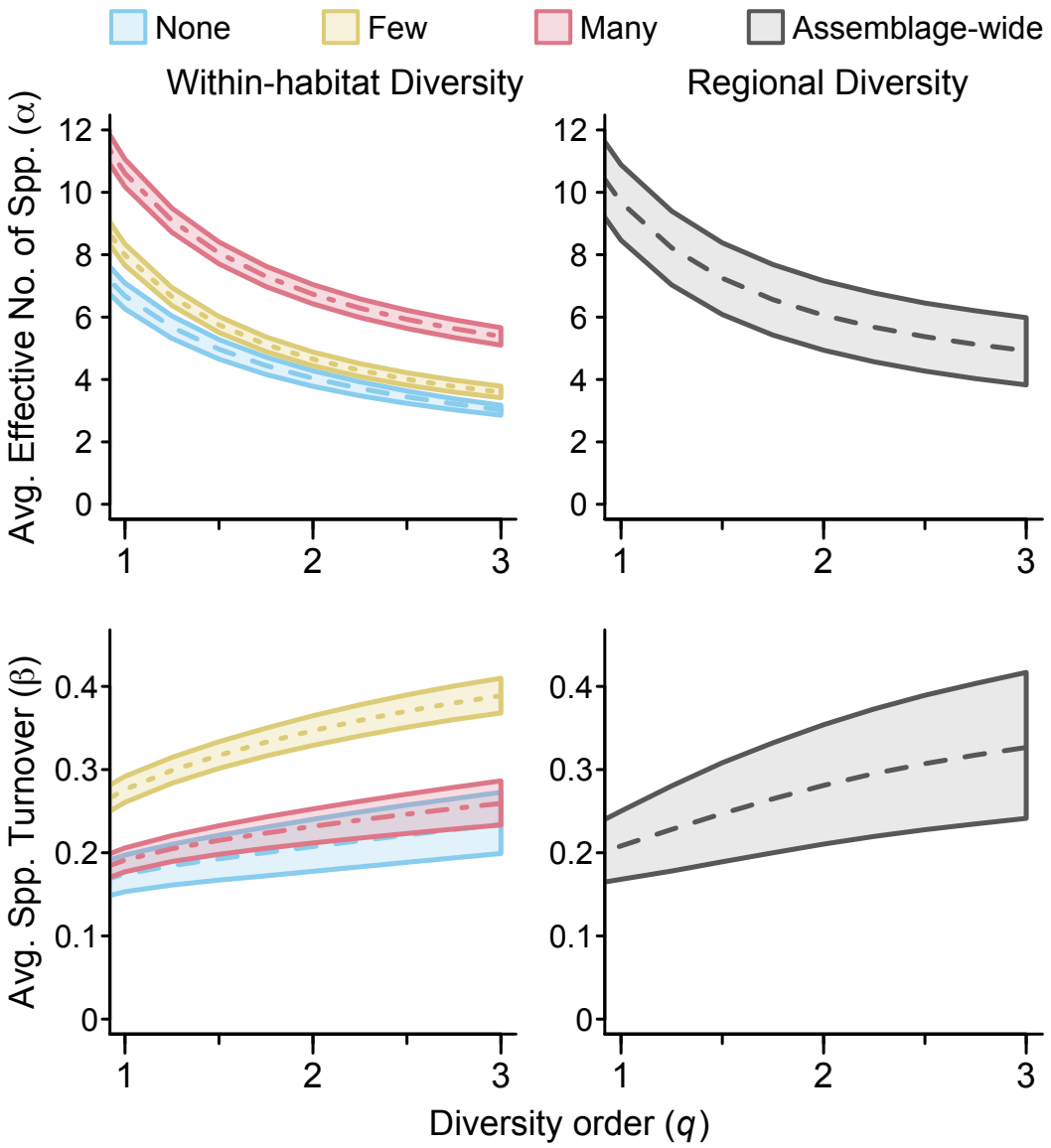


Figure 4.2: Profile plots of α and β diversity of oribatid mites (Borcard et al., 1992; Borcard and Legendre, 1994) for $q = 1$–3. The average effective number of species (top) and average turnover in species composition (bottom) are shown at two scales: within-habitat (left) and regionally (right, in grey). Mites were sampled across a semi-qualitative density gradient: none (blue) $<$ few (yellow) $<$ many (red). Broken lines represent the posterior median diversity estimate, and shaded regions encompass the 95% HDI. Regional beta diversity describes the average turnover in mite composition across the shrub density gradient.

Pairwise turnover revealed more interesting patterns of diversity. The intermediate level of shrub density had almost double the within-habitat heterogeneity ($q = 1$: 0.28 [0.26, 0.29]) compared to the low and high ends of the gradient. Moreover, there was substantial overlap in posterior uncertainty estimates for the no-shrub samples (95% HDI: 0.15, 0.2) and high-density samples (95% HDI: 0.18, 0.21). Despite the overlap in uncertainty, when considering the joint posterior, there was lower average turnover between no-shrub samples and high-density samples for $\approx$ 89% of posterior iterations across orders of q. There was also substantial species turnover across the shrub diversity gradient, especially as rarer mite species decreased in importance—up to $\approx$ 30% [0.24, 0.42] when $q = 3$. This suggests that the identities of the most abundant species (i.e., the rank order of relative abundance) differed among shrub density categories.

Intuitively, these results make sense. It is unsurprising that alpha diversity would increase with shrub density. As the number of shrubs increase, so would habitat complexity, resource input through leaf litter, and the opportunity for multitrophic interactions—each of which could increase diversity (Macarthur, 1965; Menge and Sutherland, 1976). Moreover, the mite community's species-abundance distribution is dominated by a few common species and many rare ones, as are many ecological communities (Preston, 1948). As for beta diversity, The substantial amount of turnover among samples within the intermediate shrub density relative to the low or high densities could be environmentally caused, but we suspect it is simply be a result of the semiquantitative scale used along the density gradient. The "few" category probably encompasses a larger part of the gradient than the "none" or "many" category.

4.5 Discussion and conclusions

Over the last decade there have been major advances in the conceptualization and quantification of diversity within and among communities, regions, and ecosystems. Yet to our knowledge, there has not been a unified framework to estimate that diversity, account for imperfect sampling, and

test ecological questions based on those estimates. Here we offer one avenue to do all of the above under the umbrella of one statistical method using hierarchical Bayesian modeling. Using simulations, we have shown that the model does as well or better at approximating the "true" diversity of a community, especially for $q \geq 1$ (Figure 4.1). Our framework can be used to answer questions and make comparisons about the makeup of assemblages at local and regional scales, as we demonstrated using both simulations (Table 4.1) and an empirical example using a well-known oribatid mite dataset (Figure 4.2; Borcard et al., 1992).

4.5.1 Limitations

The model does have weaknesses, notably the inability to estimate richness. To estimate probabilities of occurrence as parameters of a multinomial distribution, our model takes the "support size" (species richness) as given, and no θ parameter is ever exactly zero. Estimating asymptotic species richness is a long-standing problem in ecology. Current approaches require additional assumptions about the distribution of abundances or more information in the form of replication that may or may not be possible for many ecological study systems. In the future, we hope it will be feasible to simultaneously estimate richness and relative abundances in a Bayesian framework, but the applicability of such an approach might be fundamentally limited to datasets with extraordinary sampling effort. One solution is use posterior predictive simulation to estimate a posterior distribution of sample characteristics, such as the observed sample richness—similar in spirit to traditional rarefaction. However, the diversities calculated from this approach are estimates describing the sample and not necessarily the diversity of the real underlying system, which is ultimately what is of interest. The problem of unseen species is not unique to our approach, nor to ecology in general (Gotelli and Colwell, 2001; Paninski, 2003; Hausser and Strimmer, 2009; Valiant and Valiant, 2011). Until an integrated framework for the estimation of these species is found, we see three kinds of ad-hoc solutions. One would be to use an asymptotic richness estimate to adjust the size of the community matrix (i.e., add all-zero columns) used as

input to our estimation framework. A second alternative would be to apply bias corrections (e.g., Chao and Shen, 2003; Chao et al., 2013; Marcon et al., 2015) to the output of our Bayesian model. A third would be to use posterior simulations or rarefaction to describe expected observed richness given a sampling design. However, these are ad-hoc solutions with their own assumptions and limitations.

A third limitation is that our Bayesian model uses a multinomial likelihood function and is currently limited to discrete data (i.e., counts). However, we hope to develop an extension for continuous data in the near future.

4.5.2 Strengths

Despite these weaknesses, our approach offers a step forward in diversity estimation by providing a cohesive framework for partitioning diversity and testing related ecologically relevant hypotheses. A major strength of Bayesian statistics is that information is shared among samples through the prior. This has three important benefits. First, the posterior $\theta's$ and $\Theta's$ describe the relative abundances for the true underlying assemblage rather than describing the data. This is an important distinction as the data are peculiar to the sampling design and not basic attributes of contiguous systems.

Second, this information sharing implicitly offers a solution to unequal sample size and sampling effort. This means that rare species that might have been missed in one sample are accounted for through the prior from other samples. Such information sharing (at least partially) mitigates the problems associated with finite sampling uncertainty.

A third benefit is that the method provides intuitive and natural error envelopes through the Bayesian posterior rather than relying on bootstrapping, which is fundamentally limited by the sample and challenging to implement hierarchically. Marion et al. (2015*c*) developed a multilevel strategy for diversity partitioning that incorporated hierarchical bootstrapping, but error did not

propagate across levels and thus information was lost. In contrast, uncertainty propagation is implicit for Bayesian models.

The value of this integrated approach is illustrated in our analysis of the mite data. We were able to statistically distinguish mite communities associated with different levels of shrub density, essentially recapitulating a key result of the original studies which had used ordination and randomization tests. In addition, we were able to ask two novel questions: (1) Did the different mite communities vary in α diversity (effective Hill number of species per sample)? (2) Did mite assemblages associated with different habitat categories vary in heterogeneity (pairwise β diversity or average turnover between samples)? The answer to both questions was "yes", and posterior distributions of the diversity components showed that the most heterogeneous habitat category had neither the most nor least α diversity.

4.5.3 Future directions

Although we did not do so here, there is no reason why this method could not be used to estimate phylogenetic or functional diversity (reviewed in Chao et al., 2014) in addition to taxonomic diversity because diversity estimation occurs on the posterior abundance estimates from the Bayesian model. Incorporating phylogenetic or functional trait information would be quite tractable without much computational burden.

Nor are the models necessarily limited to either complete or no pooling. Depending on the question or situation, partial pooling models could be implemented. In partial pooling models, the priors for each subassemblage themselves share a common prior resulting in "shrinkage." This means that, unless there is an abundance of information suggesting subassemblages are distinct, their estimates will be pulled towards each other through the shared hyperprior. It also usually results in improved estimates for subassemblages with low sample sizes. However, partial pooling would require another modeling level and may be more computationally challenging depending on sample sizes and the number of subassemblages modeled. As with the challenge

of estimating richness, fitting more complex models often suffers from nonidentifiability without levels of replication rarely achievable in ecology.

In conclusion, despite the limitations, our statistical framework provides diversity partitioning estimates of the underlying system rather than the data while integrating the ability to test related ecological hypotheses. This perfectly cromulent method will truly embiggen the smallest ecological study.

Acknowledgments

We thank N. Sanders, S. Campagna, and C. Peterson for helpful discussion and comments. This research was funded by the National Science Foundation (DDIG award no. DEB-1405887 to Z. H. Marion and DEB-0614223 & 1050947 to James A. Fordyce) and the Department of Ecology & Evolutionary Biology at the University of Tennessee (Z. H. Marion).

Chapter 5

Conclusions

Ecologists and evolutionary biologists are fascinated by diversity and complexity for any number of reasons. Although the theoretical framework is compelling, the ability to estimate quantities of interest remains a challenge. My dissertation is a step forward in solving these challenges from phenotypes to ecosystems. I have shown that new, intuitive, and ecologically relevant information can be gleaned from phenotypic data by conceptualizing the complexity as components of diversity, which can be done at multiple hierarchical levels with estimates of uncertainty.

I then demonstrated that standardized measures of beta diversity are inherently biased with sample size, systematically over- or underestimating the true diversity of the underlying assemblage. I show how average pairwise measures of heterogeneity offer an intuitive solution to this problem. While they are dependent on the grain at which sampling occurs, they are unbiased and independent of sampling design.

Finally, I extend both of these approaches through the use of a hierarchical Bayesian model. This approach models both the data and—more important—the underlying system of interest. It also implicitly accounts for unequal sampling effort and sample size by allowing uncertainty to propagate without loss of error. Moreover, it allows for comparisons among different assemblages and provides an integrated framework to test ecological hypotheses. Taken together, my work

moves the empirical field forward with methods that can address novel questions and bridge the gap between samples and the actual system of interest, be they traits, communities, or ecosystems.

Bibliography

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32–46. 29

Anderson, M. J., T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, S. P. Harrison, N. J. B. Kraft, J. C. Stegen, and N. G. Swenson. 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecology Letters 14:19–28. 19, 43, 44, 45, 52, 59

Baselga, A. 2013. Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. Ecography 36:124–128. 54

Bell, G., and A. O. Mooers. 1997. Size and complexity among multicellular organisms. Biological Journal of the Linnean Society 60:345–363. 12, 34

Benard, M. F. 2004. Predator-induced phenotypic plasticity in organisms with complex life histories. Annual Review of Ecology Evolution and Systematics 35:651–673. 25

Benard, M. F., and J. A. Fordyce. 2003. Are induced defenses costly? consequences of predator-induced defenses in western toads, *Bufo boreas*. Ecology 84:68–78. 25

Bonachela, J. A., H. Hinrichsen, and M. A. Munoz. 2008. Entropy estimates of small data sets. Journal of Physics A: Mathematical And Theoretical 41:1–9. 20

Bonchev, D. G. 2003. Shannon's information and complexity. In D. H. Rouvray, and D. Bonchev, eds., Complexity in Chemistry: Introduction and Fundamentals, volume 7 of *Mathematical Chemistry Series*, chapter 4, pages 157–187. Taylor & Francis, London. 12

Borcard, D., and P. Legendre. 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). Environmental and Ecological Statistics 1:37–61. xiv, 72, 75

Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826–1832. 74

Borcard, D., F. Gillet, and P. Legendre. 2011. Numerical ecology with R. Use R! Springer, New York. 45

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045–1055. xiv, 72, 73, 75, 77

Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecological Monographs 27:326–349. 52

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational And Graphical Statistics 7:434–455. 66

Brown, B., et al. 1990. Damage and recovery of coral reefs affected by El Niño related seawater warming in the Thousand Islands, Indonesia. Coral Reefs 8:163–170. 45

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. 2016. Stan: a probabilistic programming language. Journal of Statistical Software (in press) . 66

Cayuela, L., N. J. Gotelli, and R. K. Colwell. 2015. Ecological and biogeographic null hypotheses for comparing rarefaction curves. Ecological Monographs 85:437–455. 68, 69

Chao, A., and C.-H. Chiu. 2016. Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. Methods in Ecology and Evolution pages 1–10. 2, 5, 44, 45, 46

Chao, A., and L. Jost. 2015. Estimating diversity and entropy profiles via discovery rates of new species. Methods in Ecology and Evolution 6:873–882. 45, 63

Chao, A., and T.-J. Shen. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environmental and Ecological Statistics 10:429–443. 47, 60, 78

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361–71. 44, 47

Chao, A., L. Jost, S. C. Chiang, Y.-H. Jiang, and R. L. Chazdon. 2008. A two-stage probabilistic approach to multiple-community similarity indices. Biometrics 64:1178–1186. 44

Chao, A., C.-H. Chiu, and T. C. Hsieh. 2012. Proposing a resolution to debates on diversity partitioning. Ecology 93:2037–2051. 2, 17, 18, 19, 35, 38, 44

Chao, A., C.-H. Chiu, and L. Jost. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. Annual Review of Ecology, Evolution, and Systematics 45:297–324. 2, 4, 18, 43, 44, 59, 79

———. 2016. Statistical challenges of evaluating diversity patterns across environmental gradients in mega-diverse communities. Journal of Vegetation Science 27:437–438. 1, 3, 59

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8:148–159. 28

Chao, A., Y. T. Wang, and L. Jost. 2013. Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. Methods in Ecology and Evolution 4:1091–1100. 3, 44, 60, 69, 78

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs 84:45–67. 44, 60, 63

Chao, A., T. C. Hsieh, R. L. Chazdon, R. K. Colwell, and N. J. Gotelli. 2015. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. Ecology 96:1189–1201. 3, 44

Charney, N., and S. Record. 2012. vegetarian: Jost Diversity Measures for Community Data. R package version 1.2. 6, 22

Chiu, C. H., L. Jost, and A. Chao. 2014. Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecological Monographs 84:21–44. 19, 33

Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 345:101–118. 28, 63

Condit, R., N. Pitman, E. G. Leigh, J. Chave, J. Terborgh, R. B. Foster, P. Núnez, S. Aguilar, R. Valencia, G. Villa, et al. 2002. Beta-diversity in tropical forest trees. Science 295:666–669. 45, 50, 52, 60

Crist, T., J. Veech, J. Gering, and K. Summerville. 2003. Partitioning species diversity across landscapes and regions: A hierarchical analysis of alpha, beta, and gamma diversity. American Naturalist 162:734–743. 18, 19, 34, 37, 38

Curran-Everett, D. 2000. Multiple comparisons: philosophies and illustrations. American Journal Of Physiology-Regulatory Integrative And Comparative Physiology 279:R1–R8. 11

Daly, J. W. 1995. The chemistry of poisons in amphibian skin. Proceedings of the National Academy of Sciences of the United States of America 92:9–13. 25

Dixon, P., J. Weiner, T. Mitchell-Olds, and R. Woodley. 1987. Bootstrapping the Gini coefficient of inequality. Ecology 68:1548–1551. 22

Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans, volume 38. Society for Industrial and Applied Mathematics. 6, 22, 61

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1:54–75. 6, 22, 61

Ellison, A. M. 2010. Partitioning diversity. Ecology 91:1962–1963. 1, 2, 43, 44, 59

Ettema, C., and D. Wardle. 2002. Spatial soil ecology. Trends in Ecology & Evolution 17:177–183. 32

Fordyce, J. A., and S. B. Malcolm. 2000. Specialist weevil, *Rhyssomatus lineaticollis*, does not spatially avoid cardenolide defenses of common milkweed by ovipositing into pith tissue. Journal of Chemical Ecology 26:2857–2874. 11, 20, 22, 23, 24, 25, 33

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472. 66

Gelman, A., J. Aleks, P. M. Grazia, and S. Yu-Sung. 2008. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics 2:1360–1383. 65

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. 2013. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. CRC Press. 62

Gelman, A., J. Hwang, and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016. 67

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391. 2, 3, 5, 60, 63, 77

———. 2010. Estimating species richness. In A. Magurran, and B. McGill, eds., Biological Diversity: Frontiers in Measurement and Assessment, chapter 4, pages 39–54. Oxford University Press. 22, 63

Hagman, M., R. A. Hayes, R. J. Capon, and R. Shine. 2009. Alarm cues experienced by cane toad tadpoles affect post-metamorphic morphology and chemical defences. Functional Ecology 23:126–132. 25

Hamberger, B., and S. Bak. 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philosophical Transactions of the Royal Society B-Biological Sciences 368:1–14. 4, 10

Hausser, J., and K. Strimmer. 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. The Journal of Machine Learning Research 10:1469–1484. 3, 5, 43, 44, 47, 60, 77

Hayes, R. A., M. R. Crossland, M. Hagman, R. J. Capon, and R. Shine. 2009. Ontogenetic variation in the chemical defenses of cane toads (*Bufo marinus*): Toxin profiles and effects on predators. Journal of Chemical Ecology 35:391–399. 25

Hill, M. O. 1973. Diversity and evenness: unifying notation and its consequences. Ecology 54:427–432. 2, 4, 12, 14, 15, 43, 59

Horgan, J. 1995. From complexity to perplexity. Scientific American 272:104–109. 10

Horn, H. S. 1966. Measurement of "overlap" in comparative ecological studies. The American Naturalist 100:419–0424. 52

Howard, R. W., and G. J. Blomquist. 2005. Ecological, behavioral, and biochemical aspects of insect hydrocarbons. Annual Review of Entomology 50:371–393. 11

Hubbell, S. P. 2001. The Unified Neutral Theory Of Biodiversity And Biogeography, volume 32. Princeton University Press. 59

Iason, G., J. Lennon, R. Pakeman, V. Thoss, J. Beaton, D. Sim, and D. Elston. 2005. Does chemical composition of individual Scots pine trees determine the biodiversity of their associated ground vegetation? Ecology Letters 8:364–369. 13, 34

Ihmels, J., S. R. Collins, M. Schuldiner, N. J. Krogan, and J. S. Weissman. 2007. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Molecular Systems Biology 3. 1, 59

Ings, T. C., J. M. Montoya, J. Bascompte, N. Bluethgen, L. Brown, C. F. Dormann, F. Edwards, D. Figueroa, U. Jacob, J. I. Jones, R. B. Lauridsen, M. E. Ledger, H. M. Lewis, J. M. Olesen, F. J. F. van Veen, P. H. Warren, and G. Woodward. 2009. Ecological networks—beyond food webs. Journal of Animal Ecology 78:253–269. 1, 12

Jones, C. G., and R. D. Firn. 1991. On the evolution of plant secondary chemical diversity. Philosophical Transactions of the Royal Society B: Biological Sciences 333:273–280. 4, 10

Jones, O. A. H., M. L. Maguire, J. L. Griffin, D. A. Dias, D. J. Spurgeon, and C. Svendsen. 2013. Metabolomics and its use in ecology. Austral Ecology 38:713–720. 34

Jost, L. 2006. Entropy and diversity. Oikos 113:363–375. 2, 4, 14, 15, 18, 19, 21, 38, 43, 59

———. 2007. Partitioning diversity into independent alpha and beta components. Ecology 88:2427–2439. 2, 4, 5, 14, 15, 17, 18, 19, 20, 35, 36, 38, 43, 46

Jost, L., P. DeVries, T. Walla, H. Greeney, A. Chao, and C. Ricotta. 2010. Partitioning diversity for conservation analyses. Diversity and Distributions 16:65–76. 15, 44

Koizumi, I., S. Yamamoto, and K. Maekawa. 2006. Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden charr. Molecular Ecology 15:3175–3189. 52

Kolmogorov, A. N. 1965. Three approaches to quantitative definition of information. Problems of Information Transmission 2:1–7. 12

Kraft, N. J., L. S. Comita, J. M. Chase, N. J. Sanders, N. G. Swenson, T. O. Crist, J. C. Stegen, M. Vellend, B. Boyle, M. J. Anderson, et al. 2011. Disentangling the drivers of β diversity along latitudinal and elevational gradients. Science 333:1755–1758. 3, 5, 6, 45, 59

Kuo, C. H., N. A. Moran, and H. Ochman. 2009. The consequences of genetic drift for bacterial genome complexity. Genome Research 19:1450–1454. 12

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76:5–13. 17, 20, 34, 43

Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecological Monographs 69:1–24. 26, 29

Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology Letters 16:951–963. 2, 44, 59

Legendre, P., and O. Gauthier. 2014. Statistical methods for temporal and space–time analysis of community composition data. Proceedings of the Royal Society of London B: Biological Sciences 281. 52, 59

Legendre, P., and L. Legendre. 2012. Numerical Ecology. 3rd edition. Elsevier B.V., Amsterdam, Holland. 4, 11, 44, 52

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, M. A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D. A. Wardle. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. Science 294:804–808. 1, 59

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity 89:238–247. 26

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, Mass. 10

MacArthur, R., and E. Wilson. 1967. The Theory of Island Biogeography. Landmarks in Biology Series. Princeton University Press. 1, 59

Macarthur, R. H. 1965. Patterns of species diversity. Biological Reviews of the Cambridge Philosophical Society 40:510–533. 14, 59, 76

Magurran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology & Evolution 25:574–582. 6

Malcolm, S. B. 1992. Cardenolide-mediated interactions between plants and herbivores. In G. A. Rosenthal, and M. R. Berenbaum, eds., Herbivores: Their Interaction with Secondary Metabolites, Evolutionary and Ecological Processes, pages 251–296. Academic Press, San Diego, CA. 22

Malcolm, S. B., and L. P. Brower. 1989. Evolutionary and ecological implications of cardenolide sequestration in the monarch butterfly. Experientia 45:284–295. 22

Marcon, E., and B. Herault. 2015. entropart: an R package to measure and partition diversity. Journal of Statistical Software 67:1–26. 69

Marcon, E., Z. Zhang, and B. H'erault. 2015. The decomposition of similarity-based diversity and its bias correction. HAL . 5, 6, 45, 47, 78

Marcon, E., B. Herault, C. Baraloto, and G. Lang. 2012. The decomposition of Shannon's entropy and a confidence interval for beta diversity. Oikos 121:516–522. 22

Marcon, E., I. Scotti, B. Herault, V. Rossi, and G. Lang. 2014. Generalization of the partitioning of Shannon diversity. PLoS One 9:1–8. 20

Marion, Z., J. Fordyce, and B. Fitzpatrick. 2015*a*. hierDiversity: Hierarchical Multiplicative Partitioning of Complex Phenotypes. R package version 0.1. 22, 35

Marion, Z. H., J. A. Fordyce, and B. M. Fitzpatrick. 2015*b*. Data from: Extending the concept of diversity partitioning to characterize phenotypic complexity. Dryad Digital Respository. 20

———. 2015*c*. Extending the concept of diversity partitioning to characterize phenotypic complexity. The American Naturalist 186:348–361. 1, 5, 43, 59, 78

May, R. M. 1973. Stability and complexity in model ecosystems, volume 6. Princeton University Press, Princeton, N.J. 12

McShea, D. W. 1996. Metazoan complexity and evolution: Is there a trend? Evolution 50:477–492. 12

———. 2000. Functional complexity in organisms: parts as proxies. Biology & Philosophy 15:641–668. 12

Menge, B. A., and J. P. Sutherland. 1976. Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. The American Naturalist 110:351–369. 1, 59, 76

Miller, G. 1955. Note on the bias of information estimates. In Information Theory in Psychology II-B, pages 95–100. Free Press. 60

Mithöfer, A., and W. Boland. 2012. Plant defense against herbivores: chemical aspects. Annual Review of Plant Biology 63:431–50. 11

Morisita, M. 1959. Measuring interspecific associaton and similarity between communities. Memoirs of the Faculty of Science, Kyushu University Series E (Biology) 3:65–80. 52

Mouchet, M. A., S. Villeger, N. W. H. Mason, and D. Mouillot. 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Functional Ecology 24:867–876. 33

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press. 52

Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867–878. 52

Newman, R. A. 1992. Adaptive plasticity in amphibian metamorphosis. Bioscience 42:671–678. 25

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2012. vegan: Community Ecology Package. R package version 2.0-4. 22

———. 2016. vegan: Community Ecology Package. R package version 2.3-4. 50

Paninski, L. 2003. Estimation of entropy and mutual information. Neural Computation 15:1191–1253. 3, 60, 77

Papaspiliopoulos, O., G. O. Roberts, and M. Skold. 2007. A general framework for the parametrization of hierarchical models. Statistical Science 22:59–73. 65

Patil, G. P., and C. Taillie. 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77:548–561. doi:{10.2307/2287709}. 14, 59

Pianka, E. R. 1974. Evolutionary Ecology. 1st edition. Harper & Row. 1, 3, 59

Pigliucci, M., and K. Preston. 2004. Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes. Oxford University Press, Oxford. 10

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology 29:254–283. 68, 76

———. 1962*a*. The canonical distribution of commonness and rarity: part I. Ecology 43:185–215. 1, 59

———. 1962*b*. The canonical distribution of commonness and rarity: part II. Ecology 43:410–432. 59

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 50

R Development Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 22, 35

Rao, C. R. 1982. Diversity and dissimilarity coefficients—a unified approach. Theoretical Population Biology 21:24–43. 33

Rasmann, S., and A. A. Agrawal. 2009. Plant defense against herbivory: progress in identifying synergism, redundancy, and antagonism between resistance traits. Current Opinion in Plant Biology 12:473–478. 11

Relyea, R. A. 2001. Morphological and behavioral plasticity of larval anurans in response to different predators. Ecology 82:523–540. 25

———. 2007. Getting out alive: how predators affect the decision to metamorphose. Oecologia 152:389–400. 25

Rényi, A. 1961. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, volume Vol. 1 of *Contributions to the Theory of Statistics*, pages 547–561. University of California Press. 12, 14

Ricklefs, R. E. 2006. Evolutionary diversification and the origin of the diversity-environment relationship. Ecology 87:S3–S13. 1, 59

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9:552–569. 52

Rogers, J. S. 1972. Measures of similarity and genetic distance. In M. R. Wheeler, ed., Studies in Genetics VII, 7213, chapter 4, pages 145–153. University of Texas. 52

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228. 52

Salminen, J., and M. Karonen. 2011. Chemical ecology of tannins and other phenolics: we need a change in approach. Functional Ecology 25:325–338. 11

Sardans, J., J. Penuelas, and A. Rivas-Ubach. 2011. Ecological metabolomics: overview of current developments and future challenges. Chemoecology 21:191–2–25. 34

Schenker, N. 1985. Qualms about bootstrap confidence-intervals. Journal of the American Statistical Association 80:360–361. 22

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152:1079–1089. 52

Schofield, P., D. M. Mbugua, and A. N. Pell. 2001. Analysis of condensed tannins: a review. Animal Feed Science and Technology 91:21–40. 11

Schweitzer, J. A., J. K. Bailey, D. G. Fischer, C. J. LeRoy, E. V. Lonsdorf, T. G. Whitham, and S. C. Hart. 2008. Plant-soil-microorganism interactions: Heritable relationship between plant genotype and associated soil microorganisms. Ecology 89:773–781. xiii, 20, 28, 29, 30, 32, 33

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423. 2, 12, 13

Sih, A., A. Bell, and J. C. Johnson. 2004. Behavioral syndromes: an ecological and evolutionary overview. Trends in Ecology & Evolution 19:372–378. 4, 10

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688. 1, 2, 59

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562. 52

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Biology 35:627–632. 52

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64:583–639. 67

Stan Development Team. 2016. RStan: the R interface to Stan, version 2.12.1. http://mc-stan.org/. 66

Stegen, J. C., A. L. Freestone, T. O. Crist, M. J. Anderson, J. M. Chase, L. S. Comita, H. V. Cornell, K. F. Davies, S. P. Harrison, A. H. Hurlbert, et al. 2013. Stochastic and deterministic drivers of spatial and temporal turnover in breeding bird communities. Global Ecology and Biogeography 22:202–212. 1, 59

Stinson, D. 2005. Cryptography: Theory and Practice, Third Edition. Discrete Mathematics and Its Applications. Taylor & Francis. 43

Strong, S. P., R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek. 1998. Entropy and information in neural spike trains. Physical Review Letters 80:197–200. 43

Swallow, J., and T. Garland. 2005. Selection experiments as a tool in evolutionary and comparative physiology: Insights into complex traits—an introduction to the symposium. Integrative and Comparative Biology 45:387–390. 10

Torsvik, V., and L. Øvreås. 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology 5:240–245. 1, 59

Tsallis, C. 2001. Nonextensive statistical mechanics and thermodynamics: historical background and present status. In Abe, and Okamoto, eds., Nonextensive Statistical Mechanics and its Applications, Lecture Notes in Physics, pages 3–98. Springer-Verlag. 14

Tumlinson, J. H. 2014. The importance of volatile organic compounds in ecosystem functioning. Journal of Chemical Ecology 40:212–213. 11

Valiant, G., and P. Valiant. 2011. Estimating the unseen: an N/log(N)-sample estimator for entropy and support size, shown optimal via new clts. In Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing, STOC 11, pages 685–694. ACM. 77

Van Zandt, P. A., and A. A. Agrawal. 2004. Community-wide impacts of herbivore-induced plant responses in milkweed (*Asclepias syriaca*). Ecology 85:2616–2629. 25

Veech, J. A., K. S. Summerville, T. O. Crist, and J. C. Gering. 2002. The additive partitioning of species diversity: recent revival of an old idea. Oikos 99:3–9. 17, 20, 34

Vehtari, A., A. Gelman, and J. Gabry. 2016*a*. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 0.1.6. 67

———. 2016*b*. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing pages 1–20. 7, 61, 67

Verneaux, J. 1973. Recherches é cologiques sur le r é seau hydrographique du Doubs: essai de biotypologie. Ph.D. thesis, Annales scientifiques de l'Université de Franche-Comté, Biologie Animale. 45

Verneaux, J., A. Schmitt, V. Verneaux, and C. Prouteau. 2003. Benthic insects and fish of the Doubs River system: typological traits and the development of a species continuum in a theoretically extrapolated watercourse. Hydrobiologia 490:63–74. 45

Violle, C., B. J. Enquist, B. J. McGill, L. Jiang, C. H. Albert, C. Hulshof, V. Jung, and J. Messier. 2012. The return of the variance: intraspecific variability in community ecology. Trends in Ecology & Evolution 27:244–52. 11

Wainwright, P. C. 2007. Functional versus morphological diversity in macroevolution. Annual Review of Ecology Evolution and Systematics 38:381–401. 4, 10

Warwick, R., K. Clarke, and Suharsono. 1990. A statistical analysis of coral community responses to the 1982-83 El-Niño in the Thousand Islands, Indonesia. Coral Reefs 8:171–179. 45

Watanabe, S. 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. The Journal of Machine Learning Research 11:3571–3594. 7, 61, 67

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7:256–276. 52

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30:279–338. 4, 5, 45, 59

———. 1972. Evolution and measurement of species diversity. Taxon 21:pp. 213–251. 17, 43, 44

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453–470. 59

Wolda, H. 1981. Similarity indices, sample size, and diversity. Oecologia 50:296–302. 47, 52

Wright, S. 1943. Isolation by distance. Genetics 28:114–138. 52

Yeo, G., and C. B. Burge. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Journal of Computational Biology 11:377–394. 43

Zhang, J., T. O. Crist, and P. Hou. 2014. Partitioning of alpha and beta diversity using hierarchical Bayesian modeling of species distribution and abundance. Environmental and Ecological Statistics 21:611–625. 20

Vita

Zachary Marion was born in Birmingham, AL in 1981. He was raised in Memphis, Tennessee and graduated high school from Craigmont High in 1999. In 2004 he graduated from the University of Tennessee with a Bachelors of Science in Ecology and Evolutionary Biology. Post-graduation, he worked as a laboratory manager for Dr. James Fordyce, continuing research in chemical ecology, both in the field and laboratory. During this time, he also participated in an Antarctic Research expedition (NSF ICEFISH 2004) to study fishes under Dr. Thomas Near. He accepted a graduate position at the Georgia Institute of Technology in 2006, where he was funded by an IGERT training grant from the National Science Foundation. He graduated with a Masters in Science in 2010, upon which he returned to the University of Tennessee to pursue his doctorate under Dr. Benjamin Fitzpatrick. Post-graduation, Zach will stay at the University of Tennessee for one semester as a lecturer teaching a graduate class on Bayesian modeling. He will begin working as a postdoctoral research fellow in the Department of Biology at the University of Nevada, Reno, NV in May 2017.