IDEA AND PERSPECTIVE

Taking species abundance distributions beyond individuals

Hélène Morlon,[1]* Ethan P. White,[2] Rampal S. Etienne,[3,1] Jessica L. Green,[1,4] Annette Ostling,[5] David Alonso,[3] Brian J. Enquist,[6,4,7] Fangliang He,[8] Allen Hurlbert,[9] Anne E. Magurran,[10] Brian A. Maurer,[11] Brian J. McGill,[12] Han Olff,[3] David Storch[13,14] and Tommaso Zillio[8]

Abstract

The species abundance distribution (SAD) is one of the few universal patterns in ecology. Research on this fundamental distribution has primarily focused on the study of numerical counts, irrespective of the traits of individuals. Here we show that considering a set of Generalized Species Abundance Distributions (GSADs) encompassing several abundance measures, such as numerical abundance, biomass and resource use, can provide novel insights into the structure of ecological communities and the forces that organize them. We use a taxonomically diverse combination of macroecological data sets to investigate the similarities and differences between GSADs. We then use probability theory to explore, under parsimonious assumptions, theoretical linkages among them. Our study suggests that examining different GSADs simultaneously in natural systems may help with assessing determinants of community structure. Broadening SADs to encompass multiple abundance measures opens novel perspectives in biodiversity research and warrants future empirical and theoretical developments.

Keywords

Species abundance distribution, body-size, size distribution, size–density relationship, biomass, energy use, size–energy relationship, resource partitioning, macroecology, metabolic theory.

Ecology Letters (2009) **12**: 488–501

INTRODUCTION

The species abundance distribution (SAD) describes the commonness and rarity of species in ecological communities. The SAD is one of the oldest patterns in ecology (Fisher *et al.* 1943; Preston 1948; MacArthur 1957), and remains at the core of ecological research (McGill *et al.* 2007; Alonso *et al.* 2008; Dornelas & Connolly 2008). Repeatable patterns in the shape of the SAD across taxa and environments (the hollow curve, with few dominant and many rare species) suggest that

[1]*Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR, USA*
[2]*Department of Biology and the Ecology Center, Utah State University, Logan, UT, USA*
[3]*Community and Conservation Ecology Group, University of Groningen, Haren, Netherlands*
[4]*The Santa Fe Institute, Santa Fe, NM*
[5]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA*
[6]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA*
[7]*Center for Applied Biodiversity, Science Conservation International, Arlington, VA*
[8]*Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada*

[9]*Department of Biology, University of North Carolina, Chapel Hill, NC, USA*
[10]*School of Biology, University of St Andrews, St Andrews, UK*
[11]*Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA*
[12]*School of Natural Resources, University of Arizona, Tucson, AZ, USA*
[13]*Center for Theoretical Study, Charles University of Academy Sciences of the CR, Praha, Czech Republic*
[14]*Department of Ecology, Faculty of Science, Charles University, Praha, Czech Republic*
*Correspondence: E-mail: morlon.helene@gmail.com

this distribution reflects common underlying principles that structure ecological communities. The SAD is widely used to test alternative hypotheses regarding the determinants of community structure and biodiversity (Caswell 1976; Hubbell 2001; Etienne & Olff 2005; Chave *et al.* 2006; McGill *et al.* 2006a).

Most work on the SAD has focused on the frequency distribution of the number of individuals per species (numerical abundance). This is a natural way to characterize SADs and predictions for this type of abundance distribution are available for a variety of models. Numerical abundance is commonly used in community ecology as a measure of the local dominance or success of a species (Preston 1962). The use of numerical abundance also has a precedence in population biology as the most natural measure of demography (Volterra 1926). In evolutionary biology, it is an important determinant of the genetic variability in populations, influencing the relative strength of selection and genetic drift (Wright 1938). However, the number of individuals is only one way to characterize the relative abundance of species within communities, and there are several reasons why abundance has also been measured differently. Abundance has been measured as the biomass and/or resource use of species populations in systems ecology, in order to better understand compartmentalization and fluxes of carbon, nutrients and energy in ecosystems (Odum 1983). Combined measures of abundance and biomass have also been used by marine biologists as a diagnostic tool (Warwick 1986). Numerical abundance, biomass, resource use and their interrelationship have been the main focus of empirical and theoretical macroecology, to understand the division of food and space among species (Brown & Maurer 1989; Thibault *et al.* 2004; White *et al.* 2004; Connolly *et al.* 2005; Harte *et al.* 2008). Finally, for pragmatic reasons (e.g. modularity in plants) field ecologists have sometimes used percent cover, point quadrat frequency or biomass as proxies for numerical abundance (Wilson *et al.* 1996; Chiarucci *et al.* 1999; Tokeshi 1999; Latimer *et al.* 2005).

Patterns observed for different abundance measures are expected to be similar if numerical abundance, biomass and resource use are proportional to one another. This condition may be achieved if interspecific differences are negligible, or if all individuals can be considered to be ecologically equivalent (MacArthur 1960; Hubbell 2001). For example, ecologists have often viewed the 'classical' SAD (constructed from numerical abundance) as representing the outcome of rules governing how species partition resources (Motomura 1932; MacArthur 1957, 1960; Sugihara 1980; Tokeshi 1990, 1999; Marquet *et al.* 2003). For this characterization to be valid, it is necessary to assume that individuals across species are comparable in their body-size and resource use, as originally noted by MacArthur (1960):

"It is assumed that the species whose abundances are being compared are roughly of same size, so that an individual of one species is comparable to an individual of any of the others."

This assumption is likely violated in natural systems for several reasons. First, within and between communities, species' body-sizes typically vary by several orders of magnitude (Brown & Nicoletto 1991), and large-bodied species generally use more energy than small-bodied ones (Kleiber 1947; Peters 1983). Second, the most numerous species do not always contain the most biomass or use the most resources, as they sometimes contain many small individuals (Brown & Maurer 1986; Maurer & Brown 1988). Third, individuals of the same size can have a very different resource use, as a consequence of differences in traits such as internal nutrient concentrations (Sterner & Elser 2002) or life histories (Bielby *et al.* 2007). Finally, it has been increasingly realized that individuals within populations of the same species can strongly vary in size, due in part to age structure, with important consequences for ecological interactions (Werner & Gilliam 1984; Kooijman 2000). Hence, different abundance measures are not necessarily well correlated (Saint-Germain *et al.* 2007), and SADs constructed from different abundance measures may well exhibit fundamental differences. Partly for this reason, discussion on the use of SADs has drifted away from the perspective of resource partitioning and ecosystem functioning.

It has recently been suggested that investigating abundance measures other than the number of individuals may lead more directly to mechanistic theory in ecology, placing alternative currencies at the front line of future SAD research (McGill *et al.* 2007). Three fundamental questions follow naturally: (i) What are the similarities and differences in the way individuals, biomass and resource use are distributed among species? (ii) What are the constraints that link the distributions constructed from different abundance measures? (iii) What insights can be gained from exploring the distribution of biomass and resource use among species? Aspects of these questions generated a suite of papers with equivocal conclusions (Harvey 1987; Sugihara 1989; Pagel *et al.* 1991; Taper & Marquet 1996). In these papers, the authors focused on comparing the variance between the distributions of numerical abundance, biomass and resource use, and drew upon two classical allometries (relationships of the form cx^a): the allometry between population density and body-size (Damuth 1981), and between body-size and energy use (as a proxy for resource use, Kleiber 1947). Some of the results from this body of work suggested that the equitability with which species divide individuals, biomass or resource use is similar (Sugihara 1989; Taper & Marquet 1996) while others suggested important differences (Harvey 1987; Tokeshi 1990; Pagel *et al.* 1991). Since then, there has been continued focus on species numerical abundance

distributions, but few papers have examined SADs in parallel using alternative measures for abundance (but see Thibault *et al.* 2004; Connolly *et al.* 2005; Saint-Germain *et al.* 2007).

In this paper we generalize the 'classical' SAD based on numerical abundance into a set of Generalized Species Abundance Distributions (GSADs) that encompass alternative measures, specifically biomass and resource use. We start with an empirical analysis of GSADs for four major taxonomic groups at various spatial scales. This empirical analysis aims to test the equivalence between GSADs, and to search for general patterns in the way GSADs differ from one another. We then develop, under parsimonious assumptions, a conversion framework that links GSADs to one another. This theoretical analysis provides the first step towards a theoretical understanding of differences between GSADs. We illustrate the utility of our theoretical framework for interpreting the data, and discuss the implications of our findings for our understanding of community structure. Finally, we highlight future developments that are needed to fully understand the interrelationships between GSADs.

EMPIRICAL EVIDENCE FOR MAJOR DIFFERENCES BETWEEN GSADS

We compare GSADs within a wide variety of taxonomic groups and habitats, by comparing their variance, skewness and spatial scaling. We use the terminology 'species individuals distribution' (SID) to refer to the 'classical' SAD (the one constructed from numerical abundance), while SBD and SED refer to the species biomass and energy distributions respectively (Table 1).

Data and methods

We use data from a total of 1533 communities spanning four major taxonomic groups: birds, fish, small mammals and trees. The bird data is from the Breeding Bird Survey of North America (BBS; $n = 1400$; Robbins *et al.* 1986; Sauer *et al.* 2008), the fish data from an island scale survey of Trinidad ($n = 76$; Magurran & Phillip 2001), the small mammal data from the Sevilleta LTER (Ernest *et al.* 2000) and the Portal Project ($n = 7$; Brown 1998; Ernest *et al.* in press), and the tree data from a 50-ha tropical forest plot at Barro Colorado Island (BCI; $n = 50$ plots of 1 ha; Condit 1998). Each bird species is assigned an average mass from Dunning (1993). The population biomass of each fish species comes from a direct measurement in the field. The population biomass of each mammal species is obtained from summation of individual body-size measurements. Individual tree biomass is estimated from an empirical interspecific allometry based on diameter at breast height (Brown 1997), providing a first approximation of more specific allometries incorporating wood density (Chave *et al.* 2003).

Empirical data on the energy use (a proxy for resource use) of each species within sampled communities are typically unavailable. We estimate per-capita energy use for birds and mammals based on published allometries for field metabolic rates for the respective groups (Nagy *et al.* 1999). For fish and trees, we use published allometries for basal metabolic rates (Gillooly *et al.* 2001). A detailed description and treatment of the data sources can be found in Appendix S1. In evaluating the results, the reader should note that while numerical abundance and biomass are either

Table 1 Measurements, their corresponding distributions and notations

	Measurement	Symbol	Macroecological distribution	Notation
Generalized SADs (GSADs)	Number of individuals: measured in a fixed area (density)	N	Species individuals distribution	SID
	Biomass: sum of individual masses over a species' population	M	Species biomass distribution	SBD
	Energy use: sum of individual energy use over a species' population (reflects resource utilization)	E	Species energy distribution	SED
Other macrological distributions	Body-size: average body-mass over a species' population	$\bar{m}$	Species size distribution	SsD
	Individual energy use: average individual energy use over a species' population (reflects individual resource use)	$\bar{e}$	Species individual energy distribution	SeD

We denote abundance measures by upper-case symbols (e.g. N, M and E) and measures at the individual level (per-capita measurements) by lower-case symbols (e.g. m and e). Overlines stand for averages over individuals of a given species (e.g. $\bar{m}$ and $\bar{e}$). The 'classical' species abundance distribution (the one constructed from the number of individuals) is coined 'species individuals distribution' (SID), while the term 'generalized species abundance distributions' (GSADs) is used to refer to any of the distributions of numerical abundance (SID), biomass (SBD) or energy use (SED).

measured directly or estimated with confidence, energy use is necessarily calculated by making assumptions about the form of the relationship between metabolic rate and body-size. We used the best available estimations of energy distributions given the nature of our data.

We compare the SID, SBD and SED using three simple characterizations: (i) the variance of log-transformed abundance values, a measure directly related to many common evenness metrics (e.g. Evar; May 1975; Smith & Wilson 1996), and used in previous comparisons of GSADs (Harvey 1987; Sugihara 1989; Pagel *et al.* 1991; Taper & Marquet 1996), (ii) the skewness of log-transformed abundance values (e.g. lognormal distributions have no skew), a measure of asymmetry in a frequency distribution based on the third central moment, widely used to characterize SADs (e.g. Nee *et al.* 1991; Gregory 1994; Hubbell 2001; Magurran 2003; McGill 2003) and (iii) visual evaluation of the variation of the shape of the distributions with spatial scale (i.e. unveiling; Preston 1962; Nee *et al.* 1991; Connolly *et al.* 2005). Using the log-transformed data focuses on the shape of the distribution as usually characterized by ecologists (e.g. the famous 'Preston plots'; Preston 1962) and eliminates any dependence on measurement units. In combination these three characterizations provide a broad and general picture of the similarities and differences between species abundance distributions (SADs) based on different measures.

How GSADs differ from one another

We compare for each local community within each data set the variance of log-transformed abundance measures against one another (Fig. 1, left panels, see also details and statistics in Appendix S2). Under the null hypothesis that GSADs are equivalent, the resulting data points would cluster around the 1 : 1 line. Rather than this null result, the data points consistently lay above the 1 : 1 line, demonstrating that the partitioning of individuals, biomass and energy is not equivalent. Overall, SIDs are more evenly distributed than SBDs and SEDs, and SEDs more evenly distributed than SBDs. This general pattern occurs across all the data sets, though it is less consistent in the fish data.

We compare for each local community within each data set the skewness of log-transformed abundance measures against one another (Fig. 1, right panels and Appendix S2). Under the null hypothesis that GSADs are equivalent, the resulting data points would cluster around the 1 : 1 line. Deviations from this null assumption are less consistent among data sets than the results concerning variance. The mammal communities appear to have equivalent skewness regardless of the abundance measure. The other three data sets do tend to exhibit changes in the average sign and/or magnitude of the skewness among the different abundance
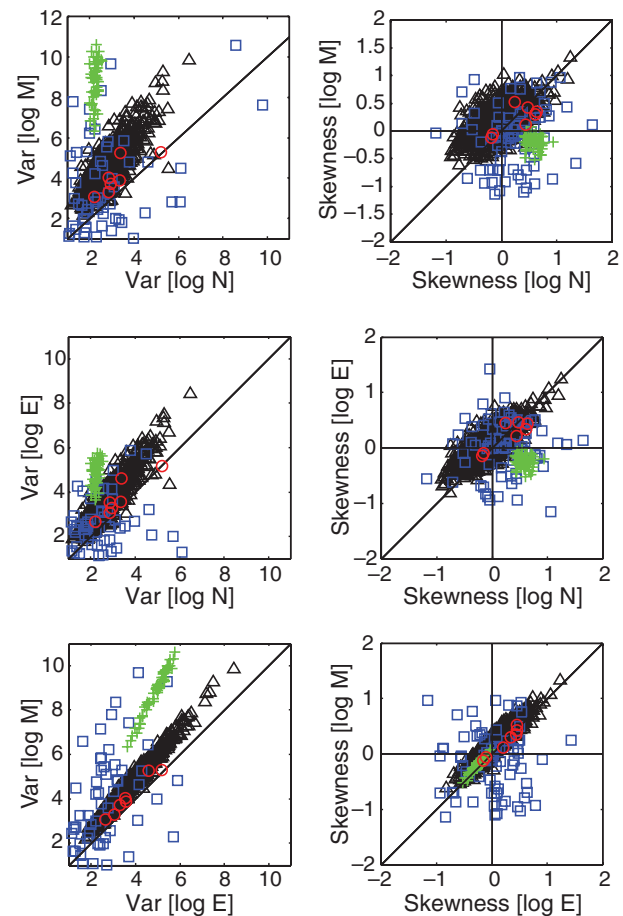

Figure 1 How diversity distributions constructed from different abundance measures differ from one another. Comparison of the variance (left panels) and skewness (right panels) between GSADs. Each data point represents a local community. Black triangles: birds; *n* = 1400. Blue squares: fish; *n* = 76. Red circles: mammals; *n* = 7. Green crosses: trees; *n* = 50. The data points would fall on the black 1 : 1 line if GSADs were equivalent with respect to variance and skewness. A Var–Var plot between two different abundance measures (left panels) illustrates which of the two is the most evenly distributed between species. Top left: individuals are overall more evenly distributed than biomass (statistically significant in all data sets); middle left: individuals are overall more evenly distributed than energy use (significant in all data sets except the fish); bottom left: energy use is overall more evenly distributed than biomass (significant in all data sets). The skewness–skewness plots (right panels) show that, except for the mammals, skewness is not equivalent among abundance measures. Detailed results and statistics are provided in Appendix S2. Two outliers from the fish data were excluded from the results in the left panels to allow for clearer presentation (Outlier 1: Var[log(*N*)] = 1.0, Var[log(*E*)] = 20.3, Var[log(*M*)] = 48.5; Outlier 2: Var[log(*N*)] = 1.7, Var[log(*E*)] = 4.0, Var[log(*M*)] = 15.3).

measures. The bird data do not show a clear tendency for the direction of skew of the SID (there are about as many sites with positively skewed distributions as with negatively

skewed distributions), while the SBD and SED are generally positively skewed. The fish data show the opposite pattern with more communities having positively skewed SIDs, and no clear tendency in the skewness of the other GSADs. The plants show the most extreme differences, cleary shifting from positively skewed SIDs to negatively skewed SBDs and SEDs. In sum, with the exception of the mammal data, all data sets exhibit differences in skewness among different GSADs, but with no clear general pattern.

We compare within each data set the global spatial scaling of GSADs. In general, the shape of sampled SADs differs from that of the underlying community SAD (Green & Plotkin 2007). In particular, SADs often appear truncated when the rarest species are not sampled (veil effect; Preston 1962). Recently, Connolly *et al.* (2005) showed that the SID unveils at much larger spatial scales than the SBD. In other words, non-truncated SBDs are obtained at spatial scales where SIDs are still mostly truncated. Our data suggest that what Connolly *et al.* refer to as a 'scale discrepancy', a phenomenon in which the SID and the SBD unveil at different spatial scales, is not specific to the coral systems they studied. Indeed, this phenomenon also occurs in each data set with multiple scales of observation (i.e. all the systems except the mammals; Fig. 2), and also holds for the SED
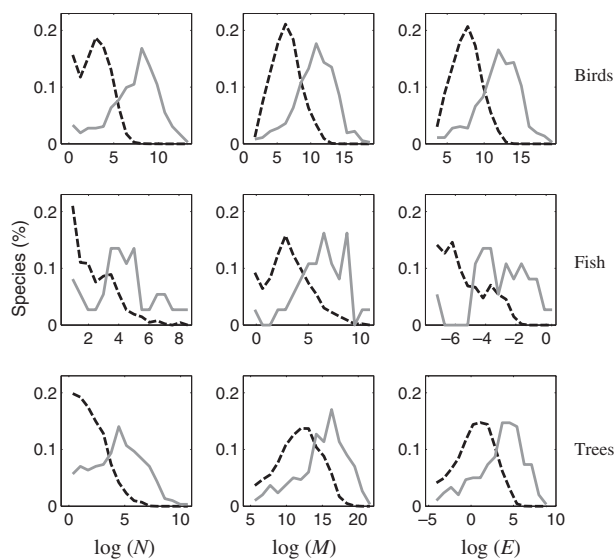

Figure 2 Diversity distributions constructed from biomass and energy use unveil faster than distributions constructed from numerical abundance. Plain lines represent GSADs at the largest available spatial scale. Birds: data at the continental scale; Fish: data at the island scale; Trees: data at the 50 ha scale. Dotted lines represent data averaged over the local communities. Variation around the mean is small and does not influence the general conclusion stemming from these analyses. The mammal data contain local communities that cannot legitimately be combined for this type of unveiling analysis.

(although in the case of the fish data the unveiling is less rapid for energy than for biomass). This phenomenon provides further support for fundamental differences between GSADs.

A THEORETICAL EXPLORATION OF THE LINK BETWEEN GSADS

The empirical results on variance, skewness and unveiling presented above demonstrate clear deviations from the null hypothesis that GSADs are equivalent. These results also show general patterns in the difference between GSADs. Understanding the origin of deviations between GSADs in natural communities will likely provide clues about the processes underlying their structure. Here, we provide the first steps of a conversion framework to analytically link GSADs to one another. Developing such a framework necessarily involves making assumptions about the way different abundance measures relate to one another. After presenting general conversions between the shape and moments of distributions under these assumptions, we evaluate their accuracy in predicting our data.

Conversion framework

We aim for generality and investigate the conversion between any two diversity distributions (probability density function describing the frequency of species with a given attribute) S_X and S_Y constructed from any log-transformed measurements X and Y. In the context of converting GSADs, X and Y may be replaced by $\log(N)$, $\log(M)$ or $\log(E)$ (see Empirical Evaluation below). The conversion between two diversity distributions S_X and S_Y is given by basic probability theory:

$$S_Y(Y) = \int P(Y|X)S_X(X)\mathrm{d}X \qquad (1)$$

$P(Y|X)$ is the probability that a species with measurement X also has measurement Y. For simplicity, we use the integral notation for both continuous and discrete X, but the integral should be interpreted as a sum when X is discrete. If X is independent of Y, or in other words if X has no explanatory power on Y, $P(Y|X) = P(Y)$ and no relationship between S_X and S_Y can be derived. However, abundance measures are not independent (Figs S1 and S2). We thus assume that X has an explanatory power on Y, and write:

$$Y = f(X) + Z \qquad (2)$$

where Z is a random variable capturing the deviation of Y from its functional dependency (f) on X. By writing this equation, we implicitly assume that ecological and evolutionary processes act primarily on X, and that Y is determined, with some variation, by X. We further assume that Z

is independent of X. These assumptions may be violated in natural communities, and should be considered as first parsimonious approximations.

In the particular case where X is the main explanatory variable for Y, Z can be thought of as a residual error ε independent of X, usually normally distributed. Denoting P_ε the probability density distribution of this error, it is straightforward to show (Appendix S3) that the distribution S_Y is related to the distribution S_X by:

$$S_Y(Y) = \int P_\varepsilon(Y - f(X))S_X(X)\mathrm{d}X \qquad (3)$$

This general equation may be used to derive relationships between S_X and S_Y for any specific form of the distributions of X and Y, the functional relationship between them, and the form of the residual error, provided that the error is independent of X. Table 2 shows specific results when the variation Z around the relationships between X and Y is either ignored or normal (i.e. log-normal on a linear scale), and when f is either unspecified or linear, which is the case for a power-law allometry between variables (see Appendix S3 for details).

There is evidence for the existence of power-laws in nature. In particular, empirical and theoretical work support that individual energy use scales with body-size as approximately $e \propto m^{\frac{3}{4}}$ (size–energy relationship; Kleiber 1947; West *et al.* 1997) and that in some cases population density scales with body-size as $N \propto \bar{m}^{-\frac{3}{4}}$ (size–density relationship; Damuth 1981; Peters 1983; White *et al.* 2007). When these relationships are assumed to be power-law, the relationship between any of N, M and E is also expected to be power-law (Appendix S4). Power-laws with log-normal error (corresponding to linear relationships with normal error for logarithmically transformed variables) are also the simplest and most natural approximations describing the relationships

between numerical abundance, biomass and the energy use of species. By substituting two of $\log(N)$, $\log(M)$ or $\log(E)$ in Y and X, the formulas in Table 2 show that GSADs are interrelated patterns rather than independent ones. These formulas also show that *a priori*, the shape of GSADs is not conserved: resource division rules predicting a log-normal SBD (Sugihara 1980) may well produce a non log-normal SID.

Converting the variance and skewness of GSADs

Many studies are more interested in simple characterizations of the distributions based on their higher moments, such as the variance and skewness, than in their detailed shape. Here, we investigate the relationship between the variance and skewness of GSADs. We consider the case of power-law allometries, as they constitute convenient approximations. On a log-log scale, a power-law allometry with exponent $a_{Y|X}$ where X is the explanatory variable reads:

$$Y = \log\left(c_{Y|X}\right) + a_{Y|X}X + Z_{Y|X} \qquad (4)$$

where $c_{Y|X}$ is a normalization constant. Below, we assume as above that $Z_{Y|X}$ is independent of X. The relationship between the variance of X and Y is straightforward:

$$\mathrm{Var}[Y] = a_{Y|X}^2\mathrm{Var}[X] + \mathrm{Var}\left[Z_{Y|X}\right] \qquad (5)$$

Note that here the variance of X and Y are not (measurement) errors of a single value of X and Y for a single species. Rather, $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$ pertain to the overall variance in observed X and Y values across species. When applied to the SID, SBD and SED, this relationship may be used to examine under which conditions biomass and energy use are expected to be more or less equitably distributed than individuals (Appendix S4).

Table 2 Conversion formulas

	No error	Log-normally distributed error
General allometry	$S_{(Y)}(Y) = \dfrac{S_{(x)}(f^{-1}(Y))}{\lvert f'(f^{-1}(Y))\rvert}$	$S_{(Y)}(Y) = \displaystyle\int \frac{1}{\sigma\sqrt{2\pi}}\exp{-\frac{(Y-f(x))^2}{2\sigma^2}}S_X(X)\mathrm{d}X$
Power-law allometry	$S_{(Y)}(Y) = \dfrac{1}{\lvert a\rvert}S_X\left(\dfrac{1}{a}(Y-\log(c))\right)$	$S_{(Y)}(Y) = \displaystyle\int \frac{1}{\sigma\sqrt{2\pi}}\exp{-\frac{(Y-\log(c)-ax)^2}{2\sigma^2}}S_X(X)\mathrm{d}X$

Formulas providing the conversion from any macroecological distribution S_X (constructed from measurement X) to any macroecological distribution S_Y (constructed from measurement Y). We assume that X is an explanatory variable for Y, that Y is functionally linked to X through the relationship f, and that the deviation Z from this functional relationship is independent of X. X and Y are typically log-transformed data. On a log-log scale, a power-law allometry between X and Y, with exponent $a_{Y|X}$ reads $Y = \log(c_{Y|X}) + a_{Y|X}X + Z_{Y|X}$ where $c_{Y|X}$ is a normalization constant, i.e. f is linear $(f(x) = \log(c) + ax)$. For example, $a_{M|N} = [1 + (1/a_{N|\bar{m}})]$ and $a_{E|N} = [1 + (a_{e|m}/a_{N|\bar{m}})]$ where $a_{N|\bar{m}}$ is the power-law exponent of the size–density relationship (Damuth 1981; White *et al.* 2007), and $a_{e|m}$ is the power-law exponent of the relationship between individual energy use and body-size (Kleiber 1947; Peters 1983; Appendix S3). In the absence of error about the relationships between measurements, these conversions correspond to a simple change of variable.

The skewness of a variable X is given by:

$$\text{Skew}[X] = \frac{E\left[(X - E[X])^3\right]}{\text{Var}[X]^{\frac{3}{2}}} \tag{6}$$

The relationship between the skewness of X and Y, for any centred and unskewed random variable $Z_{Y|X}$ (e.g. normally distributed on a log-scale; see Appendix S3 for the more general case of skewed error), is given by:

$$\text{Skew}[Y] = \frac{a_{Y|X}^3 E\left[(X - E[X])^3\right]}{\left(a_{Y|X}^2 \text{Var}[X] + \text{Var}\left[Z_{Y|X}\right]\right)^{\frac{3}{2}}} \tag{7}$$

This relationship shows that the direction of the skewness of distributions may change if the power-law exponent $a_{Y|X}$ is negative. The amplitude of the skewness is expected to be conserved for a perfect allometry between X and Y ($Z_{Y|X} = 0$), but to decrease from S_X to S_Y in the more realistic case when the allometry is not perfect.

Equations 5 and 7 provide the relationships between the second and third moments of S_X and S_Y, under the hypothesis that X is an explanatory variable for Y. Assuming instead that Y is an explanatory variable for X, the relationships will be different, as a result of error propagation (Taper & Marquet 1996; Appendix S3). These relationships may thus be used to evaluate, in a given system, if data are most consistent with a model where X or Y is considered to be an explanatory variable for the other. This possible application is illustrated below.

Illustration and empirical evaluation

We illustrate the developments above by focusing on the relationship between $\text{Var}[\log M]$ and $\text{Var}[\log N]$; we test the ability of Equation 5 to predict patterns in the empirical data, and discuss their potential application for deciphering processes driving community assembly. We note that the biomass of a species is directly given by $M = N\bar{m}$, where $\bar{m}$ is the average body-size of the species' population.

To derive the relationship between $\text{Var}[\log M]$ and $\text{Var}[\log N]$, we need to assume, as noted above, that either M or N is an explanatory variable for the other. In natural systems, complex ecological interactions act in concert to determine the biomass and numerical abundance of species. It is likely that neither M is a direct explanatory variable for N, nor N is a direct explanatory variable for M, but rather that processes act both on N and M in a correlative manner (Taper & Marquet 1996). However, we ask the following questions: are our data consistent with the hypothesis that community structure is primarily driven by processes such as resource division rules (e.g. Marquet *et al.* 2003), acting on the relative biomass of species more directly than on their relative

population size? In this case, M is the natural explanatory variable, constraining population sizes and the demographic processes from which they result. Alternatively, are our data consistent with the hypothesis that community structure is primarily driven by demographic processes (e.g. Maurer 1999), acting on the numerical abundance of species more directly than on their biomass? In this case, N is the natural explanatory variable, constraining the biomass of species.

To answer these questions, we consider two simplified scenarios used as benchmarks for more realistic models. In the first scenario, ecological processes determine the variance in the SBD, and we aim to predict the variance in the SID. We test two alternative hypotheses: either the body-size of species has no effect on their ability to accumulate biomass (e.g. large species have small population sizes and no advantage in resource monopolization), or the body-size of species influences their ability to accumulate biomass (e.g. large species have an advantage in resource monopolization). In the later case, we assume that the dependency of biomass on body-size is power-law (exponent $a_{M|\bar{m}}$). The resulting relationship between the variances, obtained with the approach outlined in the Conversion Framework, is given in Table 3 (formulas in the top table, Appendix S3). In the second scenario, we assume that ecological processes determine the variance in the SID, and we aim to predict the variance in the SBD. We test two alternative hypotheses: either the body-size of species has no effect on their numerical abundance (i.e. N is independent of $\bar{m}$), or the body-size of species influences demographic processes, and thus ultimately numerical abundance (e.g. small species have shorter generation times and reach higher densities). In the later case, we assume that the dependency of numerical abundance on body-size (i.e. the size–density relationship; Damuth 1981; White *et al.* 2007) is power-law (exponent $a_{N|\bar{m}}$). The predicted relationships between the variance corresponding to each scenario are given in Table 3 (formulas in the bottom table; Appendix S3).

We test the ability of each scenario to reproduce the data. When body-size is assumed to have an effect on either numerical abundance or biomass, the power-law exponent of the corresponding relationship is estimated by regression in each local community. We compare observed and predicted values for the variance (Fig. 3), and test the performance of the various predictions by evaluating their accuracy, as measured by the standardized root mean squared error (Table 3; predictions with best accuracy are the ones with lowest values). The results suggest that our data are best described when the numerical abundance of species is considered to be the primary explanatory variable in the system, and species' biomass results from an attribution of numerical abundance across species independent of their body-size. The second best prediction is obtained when the relative biomass of species is considered

Table 3 Four different predictions for the variance–variance relationship between GSADs, and their comparison to data

	Numerical abundance predicted from biomass	
	M does not depend on species' body-size	M depends on species' body-size
Predicted relationship	$\text{Var}[\log N] = \text{Var}[\log M] + \text{Var}[\log \overline{m}]$	$\text{Var}[\log N] = \left(1 - \frac{1}{a_{M\vert\overline{m}}}\right)^2 \text{Var}[\log M] + \text{Var}[Z_{N\vert M}]$
Birds	174	**17.3**
Fish	27.5	**12.3**
Mammals	**4.59**	6.95
Trees	729	**0.824**

	Biomass predicted from numerical abundance	
	N does not depend on species' body-size	N depends on species' body-size
Predicted relationship	$\text{Var}[\log M] = \text{Var}[\log N] + \text{Var}[\log \overline{m}]$	$\text{Var}[\log M] = \left(1 + \frac{1}{a_{N\vert\overline{m}}}\right)^2 \text{Var}[\log N] + \text{Var}[Z_{M\vert N}]$
Birds	**14.3**	1.51e+07
Fish	**10.9**	6.98e+03
Mammals	**0.108**	1.04e+03
Trees	**0.114**	1.07e+08

Upper table: M is considered to be an explanatory variable for N. Lower table: N is considered to be an explanatory variable for M. Left column: the attribution of biomass (top) or population size (bottom) across species is assumed to be independent of body-size. Right column: the body-size of species is assumed to have an effect on their capacity to accumulate biomass (top) or individuals (bottom). In this case, the dependence of M or N on $\overline{m}$ is assumed to be power-law (exponents $a_{M\vert\overline{m}}$ and $a_{N\vert\overline{m}}$, respectively). Reported values are standardized root mean square errors (RMSE/SD, where $\text{RMSE} = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2 / n}$ with x_i and y_i observed and predicted values, respectively, and SD is the standard deviation of observed values). $a_{M\vert\overline{m}}$ and $a_{N\vert\overline{m}}$ are estimated directly from the raw data, in each community separately. Standardized RMSE measure the accuracy of predictions to observations (low values correspond to a good accuracy, in bold).

to be the primary explanatory variable, and their numerical abundance results from an attribution of biomass across species dependent on their body-size (i.e. body-size influences the capacity of species to accumulate biomass). The third best prediction is obtained when the relative biomass of species is considered to be the primary explanatory variable, and their numerical abundance results from an attribution of biomass across species independent of their body-size. Finally, considering the numerical abundance of species as the primary explanatory variable, with a dependency of species' capacity to accumulate individuals on their body-size, yields particularly unrealistic predictions. This simple analysis suggests that in our systems, the main effect of the body-size of a species is to affect its ability to accumulate biomass rather than to influence population density via the production or recruitment of new offspring.

SYNTHESIS AND FUTURE DIRECTIONS

Differences between GSADs

Models of SADs based on resource division rules often assume that individuals, biomass or resource use are interchangeable:

"A niche translates ultimately into numbers of organisms (or biomass)" (Sugihara 1980).

The question of the interchangeability between abundance measures was raised over 20 years ago (Harvey 1987), generating a decade long debate (Sugihara 1989; Pagel *et al.* 1991; Taper & Marquet 1996) that ended in the following terms (Taper & Marquet 1996):

"The apportionment of individuals and resources, among species, are equivalent processes."

Our results contradict this statement. Below we discuss the cause and consequence of differences in the variance, skewness, and unveiling of GSADs.

The first difference between GSADs observed in our data is their variance (Fig. 1). This result differs from the result of Taper & Marquet (1996) due to the difference in our approach to analyzing empirical data. The authors estimated the empirical variance in energy among species using a theoretical relationship linking variances in energy and numerical abundance. As they note, their estimation likely suffers from error propagation. In our empirical analyses, we directly estimate the energy use of each species from fitted relationships between field or basal metabolic rate and body-size, and calculate the variance among species *a posteriori*. Because we have direct data on organismal body-size (except for the birds, although the error introduced by using a mean species mass is expected to be small), we eliminate the principal source of error in the theory-based variance prediction. One limitation of our approach is that
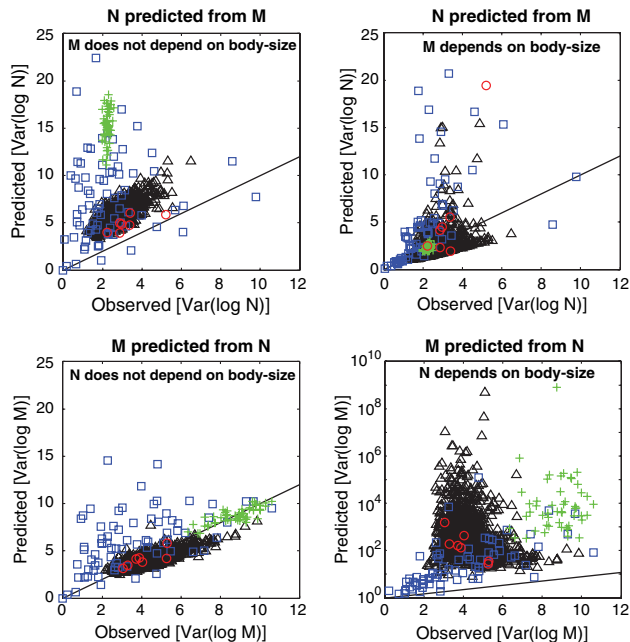
Figure 3 The relative ability of different variance–variance relationships to predict data can provide insights into processes. Comparison between variance values observed (*x*-axis) and predicted (*y*-axis) under four different scenarios. Each data point represents a local community. Black triangles: birds; *n* = 1400. Blue squares: fish; *n* = 76. Red circles: mammals; *n* = 7. Green crosses: trees; *n* = 50. Upper panels: *M* is assumed to be an explanatory variable for *N*. Lower panels: *N* is assumed to be an explanatory variable for *M*. Left panels: the attribution of biomass (top) or population size (bottom) across species is assumed to be independent of body-size. Right panels: the attribution of biomass (top) or population size (bottom) across species depends on their body-size. Note the log scale on the *y*-axis of the bottom right panel showing the particularly bad prediction obtained when assuming that *N* is an explanatory variable and that the attribution of population size across species depends on their body-size.

we ignore the error introduced by using the size–energy allometry. However, including error propagation would increase the variance in energy-use, given the biologically realistic assumption that metabolic rate is determined by body-size ('body-size causal' model). This suggests that the disparity in variances between the SED and the SID is likely to be even greater than we observed. Furthermore, as our analyses of biomass and numerical abundance are based on direct field measurements, the disparity in variances between the SBD and the SID is unequivocal.

In the four taxonomic groups analysed, biomass and resources are overall less equitably distributed among species than are individuals. Intuitively, small-bodied species live below the density for which they would receive an equal share of the total biomass or resources, while large-bodied species live above this density. In other words, large species

have an advantage in resource monopolization, an idea sometimes summarized as 'bigger is better' (Brown & Maurer 1986; Maurer & Brown 1988; Pagel *et al.* 1991; McGill 2008). These observations go against the idea embedded in the so-called 'energetic equivalence rule' (Damuth 1981). Nearly 30 years ago, Damuth (1981) gathered data showing that the numerical abundance of a species decreases as a $-\frac{3}{4}$ power of its body-size, which, combined with the $\frac{3}{4}$ power-law relationship between body-size and metabolic rate, suggests that the resource use of a species is independent of its population density and body-size. Under this rule, one would expect the variance in energy among species to be approximately constant with respect to the variance in numerical abundance. While our estimation of energy use is to be taken with caution, our analyses strongly suggest that the energy equivalence rule is not achieved in our data sets, where variance in energy and numerical abundance are highly correlated. These results are consistent with the observation that the energetic equivalence rule is rarely supported within local communities (Blackburn & Gaston 1997; White *et al.* 2007).

The second difference between GSADs observed in our data is their skewness (Fig. 1). Early statistical explanations for the origin of the shape of the SAD (May 1975) were largely based upon the normal (unskewed) nature of the SAD (plotted on a log-scale). More recent developments have been motivated by the repeated observation that SADs tend to exhibit a significant left skew on a logarithmic scale (Nee *et al.* 1991; Gregory 1994; Hubbell 2001; Magurran 2003; McGill 2003). Our results suggest that future theories could profitably focus on why different abundance measures have different degrees of skew.

The third difference between GSADs observed in our data is the scale at which they unveil (Fig. 2). Biomass and energy distributions unveil at a much smaller spatial scale than SIDs. These results are consistent with patterns observed by Connolly *et al.* (2005) in coral and fish communities. Connolly *et al.* (2005), basing their justification on Taper & Marquet (1996), suggested that the differences in unveiling between the SID and SBD provide support against niche-based origins of the SAD. Our results suggest that SID and SBD are not expected to be similar, irrespective of which ecological processes generate them. The difference in the spatial scale at which SID and SBD unveil appears to be a general phenomenon, and also applies to the energy distribution.

The rapid unveiling of biomass and resource use suggests that GSADs based on these measures may avoid a number of problems associated with the distribution of numerical abundance. The number of individuals or areas sampled in typical ecological surveys is often too small to yield completely unveiled SIDs, particularly in species-rich systems, such as tropical or microbial communities. Recon-

structing the shape of the 'true', 'untruncated' SID is not an easy task (Green & Plotkin 2007), and estimating the underlying parameters of truncated distributions is also problematic (McGill 2003). Focusing on biomass or energy-based GSADs may thus provide a more consistent and more direct characterization of community structure than individual-based distributions.

Linkages between GSADs

Linking the SID to other important macroecological distributions is fundamental to the quest towards 'unified' theories that can explain many macroecological patterns in concert (Hubbell 2001; McGill *et al.* 2007; Harte *et al.* 2008; Storch *et al.* 2008). While it is well established that the SID is linked to spatial biodiversity patterns (e.g. the species area and distance–decay relationships) through aggregated sampling (He & Legendre 2002; Plotkin & Muller-Landau 2002; Morlon *et al.* 2008), we show here a series of linkages of a different nature. Linkages between the SID and (non-spatial) macroecological distributions (in particular the SBD and SED) arise from relationships between body-size, energy use and numerical abundance, and may be derived using the formulas presented in Table 2. These results reinforce the vision of the SID as 'the master pattern', from which many others may be derived (McGill *et al.* 2007).

Our results suggest that the shape of GSADs is not generally conserved when converting from one abundance measure to an other, so that a theoretical prediction for one abundance measure is not directly applicable to another. In particular, theory relying on resource partitioning assumptions (e.g. MacArthur 1957, 1960; Sugihara 1980; Tokeshi 1990, 1999) will only directly predict the shape of SEDs. To the extent to which abundance is related to body-size or energy use, the conversion from one GSAD to another may be approached using the general formulas presented in Table 2. Such conversions can potentially extend the breadth of current ecological theory from one to multiple GSADs. From a practical standpoint, such conversions may be used to assess the shape of GSADs that are difficult to measure empirically. For example, our conversion framework can be used for predicting SEDs, which are relevant to resource-based models but often empirically intractable. This conversion framework could also provide a tool for estimating the classical SAD in systems where this is currently challenging, for example in microbial systems.

The relationships between the variance and skewness of GSADs depend on two major factors: which abundance measure is considered to be an explanatory variable, and the nature of the dependency between species' traits (e.g. body-size) and abundance (Table 3 and Appendix S3). These findings support and extend the pioneer work by Taper & Marquet (1996) showing that the expected relationship between the variance of GSADs depends on what the authors called 'models of causality'. We further suggest to leverage this dependency to identify which of the models is most consistent with empirical data. When data are available simultaneously for different abundance measures in a given system, we can indeed test the relative accuracy of predictions derived under different scenarios (Fig. 3 and Table 3).

When applying these analyses to our data, we find that the predictions with better accuracy are given by a 'N causal' model, from which biomass is deduced from independent pairing between species population size and body-size (i.e. body-size has no net effect on the demography of species), or a 'M causal' model, from which numerical abundance is deduced from directional pairing between species biomass and body-size (i.e. body-size influences the ability of species to accumulate biomass). Interestingly, while the 'N causal' model coupled with a dependency of population size on species body-size is the first model one would think about given Damuth's rule (Damuth 1981), this model yields particularly poor predictions (Fig. 3 bottom right panel; see also Taper & Marquet 1996; their fig. 2a). This result is not surprising given the weak dependence of population size on body-size in our data (results not shown), and would probably be different in communities showing a clear size–density relationship (e.g. Schmid *et al.* 2000). Drawing inference about causality from relationships between the variance and skewness of distributions offers promising avenues; however, further empirical and theoretical research is needed to extend the approach, and to fully understand its promise and limitations.

Future directions

We have described empirical patterns in the comparison between GSADs, and theoretical approaches for understanding these patterns. These analyses constitute a preliminary investigation in an area of research that clearly needs further development.

Collecting data encompassing various abundance measures
Comparing distributions constructed from various abundance measures can provide insights into pattern and process in ecological communities. However, data sets encompassing several measures are scarce, and those containing direct information on resource use are almost nonexistent. Future empirical work on SADs will greatly beneficiate from simultaneously collecting measures of numerical abundance, biomass and resource use.

Relaxing assumptions in the conversion framework
We based our conversion framework on the simplest reasonable set of assumptions. We assumed that X is an

explanatory variable for Y (or inversely), and that there exists a functional relationship between X and Y. However, two variables depending on one another do not necessarily do so directly, as the dependence can also be driven by a third variable influencing both of them simultaneously (Taper & Marquet 1996). It is not clear how distinguishing between direct causal pathways vs. joint dependence will change the results, a question that will require future work.

We assumed that the deviation Z about the functional relationship between two variables can be considered as independent of the explanatory variable. In the case of converting between numerical abundance and biomass, this hypothesis is equivalent to assuming that the deviation about the size–density relationship is independent of body-size. This hypothesis is well supported in some data (e.g. Schmid *et al.* 2000), but violated in others (e.g.Marquet *et al.* 1995).

We did not consider intraspecific variation in our framework. Intraspecific variation in body-size is not expected to have any effect on the conversion between the SID and the SBD, as the relationship between the biomass and numerical abundance of a species only involves its average body-size ($M = N\bar{m}$ is exact, by definition of $\bar{m}$). However, an exact conversion between the SID (or SBD) and the SED will involve writing the average individual energy use of a species as a function of its body-size distribution (frequency distribution of individual body-sizes within species).

Integrating sampling theory within the conversion framework

Recent theoretical developments based on sampling theory have advanced our understanding of the spatial scaling of the SID (Green & Plotkin 2007). It appears that species biomass and energy distributions unveil at different spatial scales than SID do (our Fig. 2; see also Connolly *et al.* 2005). While this phenomenon has potentially important theoretical and practical implications (Connolly *et al.* 2005), it is not well understood. Coupling a sampling framework with the conversion formulas presented in this paper will likely allow for a better understanding of the different spatial scaling perceived by various GSADs.

Towards a mechanistic modelling of GSADs

Given the complexity of ecological processes, statistically based approaches for modelling SADs receive a renewed interest (e.g. Green & Plotkin 2007; Pueyo *et al.* 2007; Harte *et al.* 2008). The theoretical part of this paper is purely statistical. For example, it does not depend on whether species are considered to be ecologically equivalent or idiosyncratic, or whether they occupy similar or different trophic levels or niches. We have shown, extending the work of others (Taper & Marquet 1996), that even a purely

statistical conversion between GSADs does depend on assumed 'models of causality'. While this dependence may be used to provide insights into ecological processes, more robust tests of theories will necessitate a mechanistic modelling of the SBD and SED along with that of the SID (e.g. O'Dwyer *et al.* 2009).

Linking GSADs to other macroecological patterns

We focused in this paper on the relationships between GSADs. However, the generality of the probability theory we use offers broader perspectives for linking frequency distributions in ecology. The allometries between numerical abundance, body-size and metabolic rate allow linking GSADs to two other macroecological patterns: the well-known species size distribution (SsD; Hutchinson & MacArthur 1959; Brown & Maurer 1989; Etienne & Olff 2004; Allen *et al.* 2006; Clauset & Erwin 2008), and the newly introduced species individual energy distribution (SeD; Appendix S3, Figs S1 and S2). Loehle (2006) noted previously that the SID and SsD are related to one another through the size–density relationship. He showed that the geometric SID can result from a uniform species-size distribution and an exact power-law size–density relationship. This result is easily reproduced using our framework (with $Y = N$ and $X = \bar{m}$; Appendix S3), and the sensitivity of this result to deviations from the exact power-law size–density relationship could further be investigated using our framework.

Realizing that the SID and SsD are linked raises the question of which of the two patterns we should try to find mechanisms or models for. Both the SID and SsD have been the subject of much attention in the literature, and many models have been developed to explain their shape (Hutchinson & MacArthur 1959; Brown & Maurer 1989; Etienne & Olff 2004; Allen *et al.* 2006; Clauset & Erwin 2008). However, the SID and SsD have been studied in isolation. Applying the general formulas provided in this paper with specific shapes for the species individuals or size distributions will help bridge this gap.

In an even broader context, frequency distributions are not limited to the GSADs, SsD and SeD. Frequency distributions of traits other than body-size and individual-energy use are commonly used. Ecologists look at species trait distributions (Reich *et al.* 2003), trait-trait allometries (Wright *et al.* 2004) and abundance-traits relationships McGill *et al.* (2006a,b). The mathematical framework we use is relevant to linking any two GSADs, any GSAD to any species trait distribution (given a mathematical link between traits and abundance), or any two mathematically related traits (Table 2; Stegen & White 2008). These relationships add to the growing literature on integrating patterns in ecology (Reuman *et al.* 2008).

REFERENCES

Allen, C., Garmestani, A., Havlicek, T., Marquet, P., Peterson, G., Restrepo, C. *et al.* (2006). Patterns in body mass distributions: sifting among alternative hypotheses. *Ecol. Lett.*, 9, 630–643.

Alonso, D., Ostling, A. & Etienne, R.S. (2008). The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.*, 11, 93–105.

Bielby, J., Mace, G., Bininda-Emonds, O., Cardillo, M., Gittleman, J., Jones, H. *et al.* (2007). The fast-slow continuum in mammalian life history: an empirical reevaluation. *Am. Nat.*, 169, 748–757.

Blackburn, T.M. & Gaston, K.J. (1997). A critical assessment of the form of the interspecific relationship between abundance and body size in animals. *J. Anim. Ecol.*, 66, 233–249.

Brown, S. (1997). *Estimating Biomass and Biomass Change of Tropical Forests: a Primer*. UNFAO Forestry Paper 134. Food and Agricultural Organization, Rome.

Brown, J.H. (1998). *The Desert Granivory Experiments at Portal*. Oxford University Press, New York.

Brown, J.H. & Maurer, B.A. (1986). Body size, ecological dominance, and cope's rule. *Nature*, 324, 248–250.

Brown, J.H. & Maurer, B.A. (1989). Macroecology: the division of food and space among species on continents. *Science*, 243, 1145–1150.

Brown, J. & Nicoletto, P. (1991). Spatial scaling of species composition: body masses of North American land mammals. *Am. Nat.*, 138, 1478.

Caswell, H. (1976). Community structure: a neutral model analysis. *Ecol. Monogr.*, 46, 327–354.

Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B. & Hubbell, S.P. (2003). Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J. Ecol.*, 91, 240–252.

Chave, J., Alonso, D. & Etienne, R.S. (2006). Theoretical biology – comparing models of species abundance. *Nature*, 441, E1.

Chiarucci, A., Wilson, J., Anderson, B. & De Dominicis, V. (1999). Cover versus biomass as an estimate of species abundance: does it make a difference to the conclusions? *J. Veg. Sci.*, 10, 35–42.

Clauset, A. & Erwin, D.H. (2008). The evolution and distribution of species body size. *Science*, 321, 399–401.

Condit, R. (1998). *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes Company, Berlin and Georgetown.

Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005). Community structure of corals and reef fishes at multiple scales. *Science*, 309, 1363–1365.

Damuth, J. (1981). Population density and body size in mammals. *Nature*, 290, 699–700.

Dornelas, M. & Connolly, S.R. (2008). Multiple modes in a coral species abundance distribution. *Ecol. Lett.*, 11, 1008–1016.

Dunning, J. (1993). *Handbook of Avian Body Masses*. CRC Press, Boca Raton.

Ernest, S.K.M., Brown, J.H. & Parmenter, R.R. (2000). Rodents, plants, and precipitation: Spatial and temporal dynamics of consumers and resources. *Oikos*, 88, 470–482.

Ernest, S.K.M., Valone, T.J. & Brown, J.H. (2009). Long-term monitoring and experimental manipulation of a Chihuahuan desert ecosystem near Portal AZ. *Ecology*, in press.

Etienne, R.S. & Olff, H. (2004). How dispersal limitation shapes species body size distributions in local communities. *Am. Nat.*, 163, 69–83.

Etienne, R.S. & Olff, H. (2005). Confronting different models of community structure to species abundance data: a Bayesian model comparison. *Ecol. Lett.*, 8, 493–504.

Fisher, R., Corbet, A. & Williams, C. (1943). The relation between the number of individuals and the number of species in a random sample from an animal population. *J. Anim. Ecol.*, 12, 42–58.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M. & Charnov, E.L. (2001). Effects of size and temperature on metabolic rate. *Science*, 293, 2248–2251.

Green, J.L. & Plotkin, J.B. (2007). A statistical theory for sampling species abundances. *Ecol. Lett.*, 10, 1037–1047.

Gregory, R.D. (1994). Species abundance patterns of British birds. *Proc. R. Soc. Lond. B*, 257, 299–301.

Harte, J., Zillio, T., Conlisk, T. & Smith, A.B. (2008). Maximum entropy and the state-variable approach to macroecology. *Ecology*, 89, 2700–2711.

Harvey, P.H. (1987). How species divide resource. *Am. Nat.*, 129, 318–320.

He, F. & Legendre, P. (2002). Species diversity patterns derived from species-area models. *Ecology*, 83, 1185–1198.

Hubbell, S. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Hutchinson, G.E. & MacArthur, R.H. (1959). A theoretical ecological model of size distributions among species of animals. *Am. Nat.*, 93, 117–125.

Kleiber, M. (1947). Body size and metabolic rate. *Physiol. Rev.*, 27, 511–541.

Kooijman, S. (2000). *Dynamic Energy and Mass Budgets in Biological Systems*. Cambridge University Press, Cambridge.

Latimer, A.M., Silander, J.A. & Cowling, R.M. (2005). Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, 309, 1722–1725.

Loehle, C. (2006). Species abundance distributions result from body size-energetics relationships. *Ecology*, 87, 2221–2226.

MacArthur, R.H. (1957). On the relative abundance of bird species. *Proc. Natl. Acad. Sci. U.S.A.*, 43, 293–295.

MacArthur, R.H. (1960). On the relative abundance of species. *Am. Nat.*, XCIV, 25–35.

Magurran, A.E. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 309, 1363–1365.

Magurran, A.E. & Phillip, D.A.T. (2001). Implications of species loss in freshwater fish assemblages. *Ecography*, 24, 645–650.

Marquet, P.A., Navarrete, S.A. & Castilla, J.C. (1995). Body size, population density, and the energetic equivalence rule. *J. Anim. Ecol.*, 64, 325–332.

Marquet, P.A., Keymer, J. & Cofre, H. (2003). *Breaking the Stick in Space: Of Niche Models, Metacommunities, and Patterns in the Relative Abundance of Species*. Blackwell, Oxford.

Maurer, B.A. (1999). *Untangling Ecological Complexity: The Macroscopic Perspective*. University of Chicago Press, Chicago.

Maurer, B.A. & Brown, J.H. (1988). Distribution of energy use and biomass among species of north-american terrestrial birds. *Ecology*, 69, 1923–1932.

May, R. (1975). *Patterns of Species Abundance and Diversity*. Harvard University Press, Cambridge.

McGill, B.J. (2003). Does mother nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecol. Lett.*, 6, 766–773.

McGill, B.J. (2008). Exploring predictions of abundance from body mass using hierarchical comparative approaches. *Am. Nat.*, 172, 88–101.

McGill, B.J., Maurer, B.A. & Weiser, M.D. (2006a). Empirical evaluation of neutral theory. *Ecology*, 87, 1411–1423.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006b). Rebuilding community ecology from functional traits. *Trends Ecol. Evol.*, 21, 178–185.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995–1015.

Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D. *et al.* (2008). A general framework for the distance-

decay of similarity in ecological communities. *Ecol. Lett.*, 11, 904–917.

Motomura, I. (1932). On the statistical treatment of communities. *Zool. Mag.*, 44, 379–383.

Nagy, K.A., Girard, I.A. & Brown, T.K. (1999). Energetics of free-ranging mammals, reptiles, and birds. *Annu. Rev. Nutr.*, 19, 247–277.

Nee, S., Harvey, P.H. & May, R.M. (1991). Lifting the veil on abundance patterns. *Proc. R. Soc. Lond. B*, 243, 161–163.

O'Dwyer, J., Lake, J., Ostling, A., Savage, V. & Green, J.L. (2009). An integrative framework for stochastic, size-structured community assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 6170–6175.

Odum, H. (1983). *Systems Ecology: An Introduction*. University Press of Colorado, Niwot.

Pagel, M.D., Harvey, P. & Godfray, H.C.J. (1991). Species-abundance, biomass and resource use distribution. *Am. Nat.*, 138, 836–850.

Peters, R. (1983). *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge.

Plotkin, J.B. & Muller-Landau, H.C. (2002). Sampling the species composition of a landscape. *Ecology*, 83, 3344–3356.

Preston, F.W. (1948). The commonness, and rarity of species. *Ecology*, 29, 254–283.

Preston, F.W. (1962). The canonical distribution of commonness and rarity. *Ecology*, 43, 185–215.

Pueyo, S., He, F. & Zillio, T. (2007). The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol. Lett.*, 10, 1017–1028.

Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M. *et al.* (2003). The evolution of plant functional variation: traits, spectra, and strategies. *Int. J. Plant Sci.*, 164, S143–S164.

Reuman, D., Mulder, C., Raffaelli, D. & Cohen, J.E. (2008). Three allometric relations of population density to body mass: theoretical integration and empirical tests in 149 food webs. *Ecol. Lett.*, 11, 1216–1228.

Robbins, C.S., Bystrak, D. & Geissler, P.H. (1986). *The Breeding Bird Survey: Its First Fifteen Years, 1965–1979*. Resource Publication 157. U.S. Department of the Interior, Fish and Wildlife Service, Washington D.C.

Saint-Germain, M., Buddle, C.M., Larrivee, M., Mercado, A., Motchula, T., Reichert, E. *et al.* (2007). Should biomass be considered more frequently as a currency in terrestrial arthropod community analyses? *J. Appl. Ecol.*, 44, 330–339.

Sauer, J.R., Hines, J.E. & Fallon, J. (2008). *The North American Breeding Bird Survey, Results and Analysis 1966–2007*. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD.

Schmid, P.E., Tokeshi, M. & Schmid-Araya, J.M. (2000). Relation between population density and body size in stream communities. *Science*, 289, 1557–1560.

Smith, B. & Wilson, J.B. (1996). A consumer's guide to evenness indices. *Oikos*, 76, 70–82.

Stegen, J.C. & White, E.P. (2008). On the relationship between mass and diameter distributions in tree communities. *Ecol. Lett.*, 11, 1287–1293.

Sterner, R. & Elser, J. (2002). *Ecological Stoichiometry. The Biology of Elements from Molecules to the Biosphere*. Princeton University Press, Princeton, NJ.

Storch, D., Sizling, A., Reif, J., Polechova, J., Sizlingova, E. & Gaston, K. (2008). The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. *Ecol. Lett.*, 11, 771–784.

Sugihara, G. (1980). Minimal community structure: an explanation of species abundance patterns. *Am. Nat.*, 116, 770–787.

Sugihara, G. (1989). How do species divide resources? *Am. Nat.*, 133, 458–463.

Taper, M.L. & Marquet, P.A. (1996). How do species really divide resources? *Am. Nat.*, 147, 1072–1086.

Thibault, K.M., White, E.P. & Ernest, M. (2004). Temporal dynamics in the structure and composition of a desert rodent community. *Ecology*, 85, 2649–2655.

Tokeshi, M. (1990). Niche apportionment or random assortment: species abundance patterns revisited. *J. Anim. Ecol.*, 59, 1129–1146.

Tokeshi, M. (1999). *Species Coexistence. Ecological and Evolutionary Perspectives*. Blackwell Science, Oxford.

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, 118, 558–560.

Warwick, R. (1986). A new method for detecting pollution effects on marine macrobenthic communities. *Mar. Biol.*, 92, 557–562.

Werner, E.E. & Gilliam, J.F. (1984). The ontogenetic niche and species interactions in size structured populations. *Annu. Rev. Ecol. Syst.*, 15, 393–425.

West, G.B., Brown, J.H. & Enquist, B.J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276, 122–126.

White, E.P., Ernest, S. & Thibault, K.M. (2004). Trade-offs in community properties through time in a desert rodent community. *Am. Nat.*, 164, 670–676.

White, E.P., Ernest, S., Kerkhoff, A.J. & Enquist, B. (2007). Relationships between body size and abundance in ecology. *Trends Ecol. Evol.*, 22, 323–330.

Wilson, J.B., Wells, T.C.E., Trueman, I.C., Jones, G., Atkinson, A.D., Crawley, N.J. *et al.* (1996). Are there assembly rules for plant species abundance? An investigation in relation to soil resources and successional trends. *J. Ecol.*, 84, 527–538.

Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science*, 87, 430–2264.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F. *et al.* (2004). The worldwide leaf economics spectrum. *Nature*, 428, 821–828.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 General link between diversity distributions.
Figure S2 Macroecological space.
Figure S3 Equitability in the distribution of individuals, biomass and energy.
Appendix S1 Detailed data sources and methods.
Appendix S2 Detailed empirical results.
Appendix S3 Detailed derivations.
Appendix S4 Equitability in the distribution of individuals, biomass and energy.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Taking species abundance distributions beyond individuals: Appendix

Appendix S1: Detailed data sources and methods

Birds

We use data from the North American Breeding Bird Survey (BBS; Robbins et al., 1986; Sauer et al., 2008), which consists of several thousand survey routes scattered across the continental United States and southern Canada. Data are gathered by volunteer observers who identify and count individuals of every bird species seen or heard at each of 50 stops along a 40 km route. The BBS dataset thus allows for the observation of patterns at both the local scale (individual survey routes) and continental scale (aggregating data across routes). We only use data from the 1400 routes for which surveys were conducted every year over the 5-year period 2002-2006 in order to minimize the chances of failing to detect rare species (McGill, 2003; Hurlbert and White, 2005, 2007). We exclude species not well-covered by BBS survey methodology (i.e., waterbirds, raptors, nocturnal species) and focus on 349 species of terrestrial land birds. Since sampling effort is constant across years and routes, we calculate density for each species in each route using the sum of the counts over the 5 year period and at the continental scale using the sum of densities per route. Mean species body-mass measurements are taken from the literature (Dunning, 1993), and per-capita energy use is calculated using empirically derived field metabolic rates (FMR; Nagy et al., 1999) for all bird species: $\overline{e} \approx FMR \approx 10.5\,\overline{m}^{0.681}$.

Fish

Fish data were obtained from a stratified survey of all major drainages in Trinidad, which took place between 1996 and 1998. Data come from seventy-six sites. The section of stream (average length 50 m) sampled at each site is short enough to be fished thoroughly, yet long enough for all species present to be represented in the catch. The sampling protocol includes major habitat types present in the river at that point (e.g. pool and riffle). Electrofishing is employed where possible, but is replaced by seining (mesh size 1.25 cm) when rivers are turbid. Large deep rivers are sampled with gillnets and a trammel net. Guppies *Poecilia reticulata* and other small fish are collected with dip nets. It is necessary to use a variety of methods to sample the range of habitats found in Trinidad, and in all cases fishing continued until no further individuals were caught. This type of removal sampling (Southwood and Henderson, 2000) is an effective way of sampling stream habitats, and sampling effort can be considered consistent across sites. The total number of individuals is recorded for each species at each site. Biomass is measured in the field at the time of fishing for each species and represents the total wet weight of all individuals caught. Per-capita energy use is calculated based on re-fitting data on resting metabolic rates Gillooly et al. (2001) using a multiple regression on appropriately transformed data. This relationship accounts for variation in local temperature as well as body size: $\bar{e} \approx MR \approx 31382\,\overline{m}^{0.75262} \exp\left(-\frac{0.4319}{kT}\right)$ where k is Boltzman's constant ($8.6\ 10^{-5}\ eV\ K^{-1}$) and T is temperature measured in degrees Kelvin. Temperature is calculated at each site from the average of three temperatures measurements recorded at the beginning, middle, and end of sampling. The value chosen for temperature has no meaningful effect on the local scale results because it is the same for every species within each local community.

Mammals

We use data from several small mammal communities from the Sevilleta LTER in New Mexico (Ernest et al., 2000) and the Portal Project in Arizona (Brown, 1998; Ernest, in press) . These studies include individual measurements of body mass and thus biomass can be directly calculated by summation without relying on mean species values (contrary to the bird data). The Sevilleta data comes from six sets of mark-recapture webs sampled continuously from 1994 to 1998 (Five

Points Grass, Five Points Larrea, Goat Draw, Rio Salado Grass, Rio Salado Larrea and Two 22). Data is summed over the three days within each census, the two annual censuses, and over the five year period. Recaptures within a single census are excluded. We use data from the control plots of the Portal Project (see Brown 1998 for details of the study) and sum the values of numerical abundance and biomass over the 12 monthly censuses and the five years from 1994-1998. As all sites are located in the desert southwest, per-capita energy use for both studies is calculated from the allometry reported in Nagy et al. (1999) relating field metabolic rate (FMR) and mass for desert mammals: $\overline{e} \approx FMR \approx 3.18 \, \overline{m}^{0.785}$.

Trees

We use data on trees from the Center for Tropical Forest Studies network (http: //www.ctfs.si.edu). Within a 50-ha plot in Barro Colorado Island in central Panama, spatial location, species identity and diameter at breast height (d) are reported for every stem $> 1cm$ (Condit, 1998; Hubbell et al., 1999, 2005). To estimate individual aboveground mass we use the empirical interspecific allometry $m = 0.124 \, d^{2.53}$ relating individual mass (m) to diameter at breast height (Brown, 1997); we then sum the individual tree masses to obtain an estimate of biomass. More accurate species-specific allometries incorporating wood density are available for BCI (Chave et al., 2003), but the allometry we use provides a first good approximation close to a theoretical prediction $m \propto d^{\frac{8}{3}}$ (West et al., 1997). The energetics of trees *sensu stricto* is not well characterized, but it has been proposed that surrogates such as biomass production, water consumption or respiration rates scale as $m^{\frac{3}{4}}$ West et al. (1997) (thus d^2) and foresters have traditionally assumed that total basal area is a decent measure of total production or resource use. To estimate per-capita energy use, we use the relationship derived by refitting the data from Gillooly et al. (2001) using a multiple regression on appropriately transformed data: $e \approx respiration \approx \exp{(16.949)} \, m^{0.692} \exp{\left(-\frac{0.642}{kT}\right)}$, where k is Boltzman's constant $(8.6 * 10^{-5} \; eV \; K^{-1}$) and T is temperature measured in degrees Kelvin. Temperature is calculated as the mean annual temperature (average within days, then within months and then the whole year) averaged over the last 5 years. The value chosen for temperature has in practice no effect on our results since it is the same for every species. The data available

for estimating the allometric relationship of energy use by plants is based on seedlings and plant parts Gillooly et al. (2001), and is thus necessarily a rough estimation. In addition, photosynthesis is strongly dependent on light availability, and light availability is highly dependent on size in tropical forest: small trees in the shaded understory are likely not at their maximum metabolic rate (Muller-landau 2006a,b). As this is the best data available, we use it to get a reasonable though coarse characterization of the energy use of tree species.

Appendix S2: Detailed empirical results

We focus here in testing the equivalence of GSADs when communities are considered as a whole (see Zar, 1999; Sokal and Rohlf, 2000; Cojbasic and Tomovic, 2007 for parametric and non-parametric approaches to comparing individual-level communities).

Variance

The bird data consist of 1400 local communities. Individuals are distributed with more equitability than biomass in all local communities but 5 (99.6% of the communities, $p < 0.05$ -run test-), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(M)]}{Var[\log(N)]}$) of $[1.64; 1.67]$. Individuals are distributed with more equitability than energy use in all local communities but 15 (98.9%, $p < 0.05$), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(E)]}{Var[\log(N)]}$) of $[1.32; 1.34]$. Energy use is more equitably distributed than biomass in all local communities ($p < 0.05$), with a 95% confidence interval for the factor of deviation ($\frac{Var[\log(M)]}{Var[\log(E)]}$) of $[1.23; 1.24]$. The fish data consist of 76 local communities. Individuals are distributed with more equitability than biomass in 55 of these 76 local communities (72.3%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.9; 5.0]$), and with more equitability than energy use in 43 of them (56.6%, n.s; 95% confidence interval for the factor of deviation: $[0.90; 1.01]$). Energy use is more equitably distributed than biomass in 59 of the 76 local communities (77.6%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[0.95; 6.25]$). The mammal data consist of 7 local communities. Individuals are distributed with more equitability than biomass in all of them (100%, $p < 0.05$; 95% confidence

interval for the factor of deviation: $[1.10; 1.44]$) and with more equitability than energy use in all local communities but 1 (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.03; 1.26]$). Energy use is more equitably distributed than biomass in all local communities (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.06; 1.15]$). The tree data consist of 50 local communities. Individuals are distributed with more equitability than biomass and energy use in all of them (100%, $p < 0.05$; 95% confidence intervals for the factors of deviation: $[3.77; 4.03]$ and $[2.10; 2.23]$ respectively). Energy use is more equitably distributed than biomass in all local communities (100%, $p < 0.05$; 95% confidence interval for the factor of deviation: $[1.78; 1.81]$). Taken together, these results demonstrate the non-equivalence of the variance of GSADs when communities are considered as a whole.

Skewness

Local bird communities are more often characterized by a left-skewed SID (814 out of 1400, 58%, $p < 0.05$), but right-skewed SBDs and SEDs (75% and 59% of the local communities respectively, $p < 0.05$). Local fish communities are characterized by mostly right-skewed SIDs (48 out of 76, 63%, $p < 0.05$) and slightly more right-skewed SEDs (54%, n.s.), but slightly more left-skewed SBDs (55%, n.s.). Five out of the seven local mammal communities show a positive skew for the SID, the SBD and the SED, while the two other local communities show a negative skew for the three distributions (n.s.). All the local 1 ha local tree plots show a positively-skewed SID, and all of them a negatively-skewed SBD (in the exception of 2 plots) and SED ($p < 0.05$).

Appendix S3: Detailed derivations

General conversion formula

Macroecological distributions are all interrelated by conditional probabilities (or probability densities; e.g. Figure S1). The conversion between the frequency distribution S_X and the frequency

distribution S_Y is given by the general formula:

$$S_Y(Y) = \int P(Y|X) S_X(X) dX \tag{1}$$

With

$$Y = f(X) + \varepsilon$$

where f is a general allometry and the error ε is independent of X and centered ($E(\varepsilon) = 0$), $P(Y|X)$ reads:

$$P(Y|X) = P(f(X) + \varepsilon|X) = P(\varepsilon|X) = P_\varepsilon(\varepsilon) = P_\varepsilon(Y - f(X))$$

Combining with 1, the conversion from X to Y is given by the general formula:

$$S_Y(Y) = \int P_\varepsilon(Y - f(X)) S_X(X) dX \tag{2}$$

Specific conversion formulas (Table 2)

- *General allometry, no error*

When there is no error around an allometric relationship f, P_ε may be written as a Dirac delta function:

$$P_\varepsilon(Y - f(X)) = \delta(Y - f(X))$$

For any monotic function g with root x_i:

$$\delta(g(x)) = \frac{\delta(x - x_i)}{|g'(x_i)|}$$

where the prime denotes the derivative. We thus have (for $f'(X) \neq 0$):

$$\delta\left(Y - f\left(X\right)\right) = \frac{\delta\left(X - f^{-1}\left(Y\right)\right)}{|f'\left(X\right)|}$$

Therefore, (2) becomes

$$S_Y\left(Y\right) = \frac{S_X\left(f^{-1}\left(Y\right)\right)}{|f'\left(f^{-1}\left(Y\right)\right)|} \tag{3}$$

- *Power-law allometry, no error*

Substituting in 3, for a power-law allometry $f : x \to \log\left(c\right) + ax$ without error:

$$S_Y\left(Y\right) = \frac{1}{|a|} S_X\left(\frac{1}{a}\left(Y - \log\left(c\right)\right)\right) \tag{4}$$

- *General allometry, normally distributed error*

Substituting in 2, for a normally distributed error $P_\varepsilon(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp^{-\frac{x^2}{2\sigma^2}}$ and a general allometric relationship f:

$$S_Y\left(Y\right) = \int \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(Y-f(X))^2}{2\sigma^2}} S_X\left(X\right) dX \tag{5}$$

- *Power-law allometry, normally distributed error*

Substituting in 5, for a normally distributed error and power-law allometry:

$$S_Y\left(Y\right) = \int \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(Y-\log(c)-aX)^2}{2\sigma^2}} S_X\left(X\right) dX \tag{6}$$

The geometric SID results from the uniform SsD (case power-law allometry with no error)

It has been show Loehle (2006) that the geometric SID results from the uniform SsD when an exact power-law allometry is assumed. This result is easily reproducibe using our framework. Assume

that the size distribution is uniform on a log scale, i.e:

$$S_s \left(\log \left(\overline{m} \right) \right) = \frac{1}{\log(\overline{m}_{max}) - \log(\overline{m}_{min})} \qquad if \quad \overline{m}_{min} \leq \overline{m} \leq \overline{m}_{max}$$
$$= 0 \qquad otherwise$$

substituting in 4 leads (with $a_{N|\overline{m}} \leq 0$)

$$S_N \left(\log \left(N \right) \right) = \frac{1}{|a|(\log(\overline{m}_{max}) - \log(\overline{m}_{min}))} \qquad if \quad c\overline{m}_{max}^{a_{N|\overline{m}}} \leq N \leq c\overline{m}_{min}^{a_{N|\overline{m}}}$$
$$= 0 \qquad otherwise$$

i.e. the species abundance distribution is also uniform on a log-scale. Note that the geometric rank abundance curve is equivalent to the uniform distribution on a log scale, and we therefore reproduce the results by Loehle (2006). The advantage of our framework is that it offers the possibility to extend such predictions to biologically more realistic cases, in particular incorporating the effect of variation around the size-density relationship.

Variance-variance relationships depend on the choice of "causal model"

There are two ways to describe the functional relationships between X and Y: 1) Y is functionally determined by X. For example, the biomass of a population may be largely a consequence of how numerous the population is if processes act primarily to determine the number of individuals in a population (e.g. processes of birth, death, speciation and migration) 2) X is functionally determined by Y. For example the density of a population may be largely a consequence of its' biomass if processes act primarily to determine population biomass (e.g. growth *versus* reproduction response in indeterminate growers such as trees). These two different ways to write the functional relationship between X and Y lead to two different relationships between the variance of X and Y (Taper and Marquet, 1996).

In the first case (Y functionally determined by X),

$$Y = \log \left(c_{Y|X} \right) + a_{Y|X} X + Z_{Y|X} \tag{7}$$

with $Z_{Y|X}$ independent of X. In this case,

$$Var\left[Y\right] = a_{Y|X}^2 Var\left[X\right] + Var\left[Z_{Y|X}\right] \tag{8}$$

In the second case (X functionally determined by Y)

$$X = \log\left(c_{X|Y}\right) + \frac{1}{a_{Y|X}}Y + Z_{X|Y} \tag{9}$$

with $Z_{X|Y}$ independent of Y. In this case, $Var\left[X\right] = \frac{1}{a_{Y|X}^2}Var\left[Y\right] + Var\left[Z_{X|Y}\right]$ and

$$Var\left[Y\right] = a_{Y|X}^2 Var\left[X\right] - a_{Y|X}^2 Var\left[Z_{X|Y}\right] \tag{10}$$

Equations 8 and 10 show that the relationship between the variance of two macroecological distributions S_X and S_Y depends on the steepness of the allometric relationship between X and Y and on the model of functional relationship between X and Y. Equations 8 and 10 may be used to understand most relevant "causal models" in natural communities. Indeed, variance values and allometric exponents can be measured from field data, which can then be used to infer the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$. Given that $Var\left[Z_{Y|X}\right]$ is positive, Equation 8 is satisfied if the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$ is positive, indicating a "X causal" model. If instead the sign of $Var\left[Y\right] - a_{Y|X}^2 Var\left[X\right]$ is negative, Equation 10 is satisfied, indicating a "Y causal" model.

Converting the skewness of macroecological distributions

We write

$$Y = \log\left(c_{Y|X}\right) + a_{Y|X}X + Z_{Y|X} \tag{11}$$

with $Z_{Y|X}$ independent of X. Assume that $Z_{Y|X}$ is centered ($E\left[Z_{Y|X}\right] = 0$). The expected value of Y reads:

$$E[Y] = E[\log\left(c_{Y|X}\right) + a_{Y|X}X + Z_{Y|X}] = \log\left(c_{Y|X}\right) + a_{Y|X}E\left[X\right] + E\left[Z_{Y|X}\right] = \log\left(c_{Y|X}\right) + a_{Y|X}E\left[X\right]$$

The variance of Y reads:

$$Var\left[Y\right] = a_{Y|X}^2 Var\left[X\right] + Var\left[Z_{Y|X}\right]$$

The skewness of Y is defined by:

$$\gamma_1 = \frac{E[(Y - E[Y])^3]}{(E[(Y - E[Y])^2])^{\frac{3}{2}}} = \frac{E[(Y - E[Y])^3]}{V\left[Y\right]^{\frac{3}{2}}}$$

$$E[(Y-E[Y])^3] = E\left[\left(a_{Y|X}\left(X - E\left[X\right]\right) + Z_{Y|X}\right)^3\right] = E\left[a_{Y|X}^3\left(X - E\left[X\right]\right)^3 + Z_{Y|X}^3 + 3a_{Y|X}^2\left(X - E\left[X\right]\right)^2 Z_{Y|X} + 3Z_{Y|X}^2 a_{Y|X}\left(X - E\left[X\right]\right)\right]$$

$Z_{Y|X}$ and X are independent, so that:

$$E\left[3a_{Y|X}^2\left(X - E\left[X\right]\right)^2 Z_{Y|X}\right] = 3a_{Y|X}^2 E\left[Z_{Y|X}\right] E\left[\left(X - E\left[X\right]\right)^2\right] = 0$$

$$E\left[3Z_{Y|X}^2 a_{Y|X}\left(X - E\left[X\right]\right)\right] = 3a_{Y|X} E\left[Z_{Y|X}^2\right] E\left[\left(X - E\left[X\right]\right)\right] = 0$$

Finally

$$E[(Y - E[Y])^3] = E\left[a_{Y|X}^3\left(X - E\left(X\right)\right)^3\right] = a_{Y|X}^3 E\left[\left(X - E\left(X\right)\right)^3\right] + E\left[Z_{Y|X}\right]^3$$

Thus the skewness becomes

$$\gamma_1 = \frac{a_{Y|X}^3 E[(X - E[X])^3] + E\left[Z_{Y|X}\right]^3}{(a_{Y|X}^2 Var[X] + Var\left[Z_{Y|X}\right])^{\frac{3}{2}}}$$

This expression shows that Y can be skewed even if X is not, if $Z_{Y|X}$ is skewed. In the case of unskewed $Z_{Y|X}$:

$$\gamma_1 = \frac{a^3 E[(X - E[X])^3]}{(a^2 Var[X] + Var\left[Z_{Y|X}\right])^{\frac{3}{2}}}$$

This formula shows that if X explains Y with a given unexplained variance, the absolute value

of the skewness of Y is always smaller than the absolute value of the skewness of X. This formula also shows that the sign of the skewness of Y is opposite to that of X for $a \leq 0$. In particular, with $a_{M|N} = 1 + \frac{1}{a_{N|\overline{m}}}$, N and M are expected to have opposite skew for $1 + \frac{1}{a_{N|\overline{m}}} < 0$, or equivalently $0 > a_{N|\overline{m}} > -1$. With $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|\overline{m}}}$, N and E are expected to have opposite skew for $1 + \frac{a_{e|m}}{a_{N|\overline{m}}} < 0$, or equivalently $0 > a_{N|\overline{m}} > -a_{e|m}$. With typical values reported for the exponent of the size-density relationship, the direction of the skewness is expected to switch between abundance measures. Under the assumption of unskewed error around allometric scaling laws, species individuals and biomass distributions are expected to have opposite skew as soon as the slope of the size-density relationship is shallower than -1, as is usually the case in natural communities. Also, species individuals and energy distributions are expected to have opposite skew as soon as the slope of the size-density relationship is shallower than the slope of the size-energy relationship, which is the case in our data and is also supported by others (Brown and Maurer, 1986).

Specific variance-variance relationships (Table 3)

We derive the variance-variance relationship between numerical abundance (N) and biomass (M) under the four simplified scenarios presented in section *Illustration and Empirical Evaluation* (Table 3). The biomass of a species is directly given by $M = N\overline{m}$, where $\overline{m}$ is the average body-size of the species' population. We thus have: $\log(M) = \log(N) + \log(\overline{m})$.

First scenario: M is the explanatory variable; we write $\log(N) = \log(M) - \log(\overline{m})$.

In the first case, the body-size of species has no effect on their ability to accumulate biomass. In other words, M is independent of $\overline{m}$. We find:

$$Var\left[\log N\right] = Var\left[logM\right] + Var\left[log\overline{m}\right]$$

In the second case, we assume a power-law dependency of M on $\overline{m}$ (exponent $a_{M|\overline{m}}$), so that $\log(N) = (1 - \frac{1}{a_{M|\overline{m}}})\log(M) + Z_{N|M}$ with $Z_{N|M}$ independent of M. We find:

$$Var\left[\log N\right] = \left(1 - \frac{1}{a_{M|\overline{m}}}\right)^2 Var\left[logM\right] + Var\left[Z_{N|M}\right]$$

Second scenario: N is the explanatory variable; we write $\log\left(M\right) = \log\left(N\right) + \log\left(\overline{m}\right)$.

In the first case, the body-size of species has no effect on numerical abundance. In other words, N is independent of $\overline{m}$. We find:

$$Var\left[\log M\right] = Var\left[logN\right] + Var\left[log\overline{m}\right]$$

In the second case, we assume a power-law dependency of N on $\overline{m}$ (exponent $a_{N|\overline{m}}$), so that $\log\left(M\right) = (1 + \frac{1}{a_{N|\overline{m}}})\log\left(N\right) + Z_{M|N}$ with $Z_{M|N}$ independent of N. We find:

$$Var\left[\log M\right] = \left(1 + \frac{1}{a_{N|\overline{m}}}\right)^2 Var\left[logN\right] + Var\left[Z_{M|N}\right]$$

Appendix S4: Equitability in the distribution of individuals, biomass and energy

Assuming that the relationships between both N and $\overline{m}$ and $\overline{e}$ and $\overline{m}$ are power-law, we denote $c_{N|\overline{m}}$ and $a_{N|\overline{m}}$ the normalization constant and power-law exponent of the allometry between $\overline{m}$ and N. Writing $E = N\overline{e}$ and $M = N\overline{m}$, it is obvious that the relationship between any 2 of the 5 variables N, $\overline{m}$, $\overline{e}$, M and E is also power-law. Notations for the allometries between any two other variables are denoted accordingly. For example if $N \sim \overline{m}^{a_{N|\overline{m}}}$, then $\overline{m} \sim N^{\frac{1}{a_{N|\overline{m}}}}$, so that $M \sim N^{\left(1 + \frac{1}{a_{N|\overline{m}}}\right)}$ and $a_{M|N} = 1 + \frac{1}{a_{N|\overline{m}}}$. If furthermore we neglect intraspecific body-size variation (reasonable assumption in the case for determinant growers such as birds and mammals), we can write $\overline{e} \sim \overline{m}^{a_{e|m}}$, so that $E \sim N\left(N^{\frac{1}{a_{N|\overline{m}}}}\right)^{a_{e|m}} \sim N^{1 + \frac{a_{e|m}}{a_{N|\overline{m}}}}$ and $a_{E|N} = 1 + \frac{a_{e|m}}{a_{N|\overline{m}}}$. For example with Damuth ($a_{N|\overline{m}} = -0.75$) and Kleiber's ($a_{e|m} = 0.75$) coefficients: $a_{M|N} = -0.33$ and $a_{E|N} = 0$ (the energy equivalence rule).

Using the formulas presented in the manuscript (section 2), it is straightforward to derive

the relationships between the variance in number of individuals, biomass and energy use. These formulas depend on assumptions regarding which variable is explanatory vs. explained. Figure S3a illustrates the conditions under which biomass is expected to be more equitably distributed than individuals, and Figure S3b the conditions under which energy is expected to be more equitably distributed than individuals.

Captions

- Figure S1: *General link between diversity distributions*

We denote probabilities (or probability densities) associated with species level conversions (blue shade, conversion between species individual, biomass and energy distributions) by upper case symbols (P), and probabilities associated with per capita level conversions (yellow shade, conversion between species individual, size and per capita energy distributions) by lower case symbols (p). Conversions between population level distributions naturally stem from per capita level probabilities (in particular $p\left(N|\overline{m}\right)$ describing the relationship between density and body-size, and $p(\overline{e}|\overline{m})$ describing the relationship between metabolic rate and body-size). Equations in the figure provide the relationships between probabilities at the population and per capita level.

- Figure S2: *Macroecological space for a hypothetical ecological community*

This figure illustrates how the "classical" species abundance distribution (constructed from individual counts, SID) relates to species biomass (SBD) and energy (SED) distributions (a), but also to the species size (SsD) and individual energy use (SeD) distributions (b). Each point represents a species in a hypothetical community. The data are generated assuming: *1)* a log-series species individual distribution with parameter 0.9 *2)* a power-law allometry between the number of individuals and size with exponent $-\frac{3}{4}$ (Damuth's exponent) and normally distributed error with variance 0.5 *2)* a power-law allometry between individual mass and individual metabolic rate with exponent $\frac{3}{4}$ (Kleiber's exponent) and normally distributed error (variance 0.5); all individuals within a species are assumed of the same size, so that $m = \overline{m}$ and $e = \overline{e}$. The gray shadows in the figure are

projections of the species on the 2-dimensional surfaces, showing allometric relationships between any two variables. For example, the projection on the $\log(N) - \log(\overline{m})$ plane (panel b) is the size-density relationship, and the projection on the $\log(\overline{e}) - \log(\overline{m})$ plane is the size-energy relationship. The red dots are projections of the dataset on the 1-dimensional axes. The density of these dots constitutes the macroecological distributions (SID, SBD, SED, SsD and SeD), shown in inserts in their familiar histogram form. The energetic equivalence rule resulting from the choice of the exponents (same steepness for the size-density and size-energy relationships) is illustrated by the projection on the $\log(E) - \log(N)$ surface (panel a): E does not depend on N. Note that the right-skewed log-series SID corresponds to a left-skewed SsD (panel b).

- Figure S3: *Conceptual figure illustrating the effect of allometric slopes, error and causality on the relationships between: a) the equitability of individuals and biomass division b) the equitability of individuals and resource division*

If the error around allometries is ignored (plain lines), whether biomass (or energy) is more equitably distributed than the number of individuals depends on the slope of the size-density relationship (or the ratio of the slopes of the size-density and size-energy relationships, respectively). With Damuth exponent for the size-density relationship (-0.75), biomass is expected to be more equitably distributed than the number of individuals. With Kleiber's exponent for the size-energy relationship (0.75), energy is expected to be more equitably distributed than the number of individuals for a size-density relationship steeper than -0.375. Incorporating the effect of scatter around the allometries significantly change the results. The relationships between evenness depend on the direction of causality (short-dashed lines versus dashed-point lines).

References

Brown, J. H. (1998). *The desert granivory experiments at Portal.* Oxford University Press, New York.

Brown, J. H. and Maurer, B. A. (1986). Body size, ecological dominance, and cope's rule. *Nature*, 324:248–250.

Brown, S. (1997). *Estimating biomass and biomass change of tropical forests*. Food and Agricultural Organization of the United Nations Forestry Paper 134.

Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., and Hubbell, S. P. (2003). Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J. Ecol.*, 91:240–252.

Cojbasic, V. and Tomovic, A. (2007). Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples. *Computational Statistics and Data Analysis*, 51(12):5562–5578.

Condit, R. (1998). *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes Company, Berlin and Georgetown.

Dunning, J. (1993). *Handbook of Avian Body Masses*. CRC Press, Boca Raton.

Ernest, S. K. M., Brown, J. H., and Parmenter, R. R. (2000). Rodents, plants, and precipitation: Spatial and temporal dynamics of consumers and resources. *Oikos*, 88:470–482.

Ernest, S. K. M., Valone, T. J., and Brown, J. H. (in press). Long-term monitoring and experimental manipulation of a Chihuahuan desert ecosystem near Portal AZ. *Ecology*.

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., and Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science*, 293.

Hubbell, S., Condit, R., and Foster, R. (2005). Barro Colorado Forest Census Plot Data.

Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., Wright, S. J., and Loo de Lao, S. (1999). Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, 283:554–557.

Hurlbert, A. H. and White, E. P. (2005). Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecol. Lett.*, 8(3):319–327.

Hurlbert, A. H. and White, E. P. (2007). Ecological correlates of range occupancy in North American breeding birds. *Global Ecol. Biogeogr.*, 16:764–773.

Loehle, C. (2006). Species abundance distributions result from body size-energetics relationships. *Ecology*, 87(9):2221–2226.

McGill, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature*, 422:881–885.

Muller-landau, H. (2006a). Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecol. Lett.*, 9:575–588.

Muller-landau, H. (2006b). Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol. Lett.*, 9:589–602.

Nagy, K. A., Girard, I. A., and Brown, T. K. (1999). Energetics of free-ranging mammals, reptiles, and birds. *Annu. Rev. Nutr.*, 19(1):p247 – 277.

Robbins, C. S., Bystrak, D., and Geissler, P. H. (1986). *The breeding bird survey: its first fifteen years, 1965-1979.* Resource Publication 157. U.S. Department of the Interior, Fish and Wildlife Service.

Sauer, J. R., Hines, J. E., and Fallon, J. (2008). *The North American Breeding Bird Survey, Results and Analysis 1966 - 2007.* Version 5.15.2008. USGS Patuxent Wildlife Research Center.

Sokal, R. and Rohlf, F. (2000). *Biometry: the principles and practice of statistics in biological research.* Freeman New York.

Southwood, R. and Henderson, P. (2000). *Ecological Methods.* Blackwell Publishing.

Taper, M. L. and Marquet, P. A. (1996). How do species really divide resources? *Am. Nat.*, 147(6):1072–1086.

West, G. B., Brown, J. H., and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276:122–126.

Zar, J. (1999). *Biostatistical analysis.* Prentice Hall Upper Saddle River, NJ.

Figure S1

Per-capita level measurements $(\overline{m}, \overline{e})$

Population level measurements (N, M, E)

$S_{\overline{e}}$ (SeD)

$S_{\overline{m}}$ (SsD)

$p(\overline{e} \mid \overline{m})$

$p(\overline{m} \mid \overline{e})$

$p(\overline{e} \mid N)$ $p(N \mid \overline{e})$ $p(N \mid \overline{m})$ $p(\overline{m} \mid N)$

$P(M \mid N) = p\left(\overline{m} = \dfrac{M}{N} \mid N\right)$ $P(N \mid M) = p\left(N \mid \overline{m} = \dfrac{M}{N}\right)$

$P(E \mid N) = \tilde{p}\left(\overline{e} = \dfrac{E}{N} \mid N\right)$ $P(N \mid E) = p\left(N \mid \overline{e} = \dfrac{E}{N}\right)$

$P(M \mid E) = \int p\left(\overline{m} = \dfrac{M}{N} \mid N\right) p\left(N \mid \overline{e} = \dfrac{E}{N}\right) dN$

$P(E \mid M) = \int p\left(\overline{e} = \dfrac{E}{N} \mid N\right) p\left(N \mid \overline{m} = \dfrac{M}{N}\right) dN$

$E = N\overline{e}$ $S_N(N)$ (SAD) $M = N\overline{m}$

$P(E \mid N)$ $P(N \mid E)$ $P(N \mid M)$ $P(M \mid N)$

(SED) S_E

(SBD) S_M

$P(M \mid E)$

$P(E \mid M)$

Figure S2

(a)

Species Individuals Distribution (SID)

Species Biomass Distribution (SBD)

Species Energy Distribution (SED)

$\log(N)$

$\log(E)$

$\log(M)$

(b)

Species Individuals Distribution (SID)

Species-size Distribution (SsD)

Species individual-energy Distribution (SeD)

$\log(N)$

$\log(\bar{e})$

$\log(\bar{m})$

Figure S3a

Figure S3b