ECOLOGY LETTERS

Ecology Letters, (2015) **18**: 1068–1077 doi: 10.1111/ele.12489

LETTER

Integrating macroecological metrics and community taxonomic structure

John Harte,[1]* Andrew Rominger[2] and Wenyu Zhang[3]

Abstract

We extend macroecological theory based on the maximum entropy principle from species level to higher taxonomic categories, thereby predicting distributions of species richness across genera or families and the dependence of abundance and metabolic rate distributions on taxonomic tree structure. Predictions agree with qualitative trends reported in studies on hyper-dominance in tropical tree species, mammalian body size distributions and patterns of rarity in worldwide plant communities. Predicted distributions of species richness over genera or families for birds, arthropods, plants and microorganisms are in excellent agreement with data. Data from an intertidal invertebrate community, but not from a dispersal-limited forest, are in excellent agreement with a predicted new relationship between body size and abundance. Successful predictions of the original species level theory are unmodified in the extended theory. By integrating macroecology and taxonomic tree structure, maximum entropy may point the way towards a unified framework for understanding phylogenetic community structure.

Keywords

Abundance distribution, energy equivalence, macroecology, maximum entropy, metabolic rate distribution, taxonomic tree.

Ecology Letters (2015) **18**: 1068–1077

INTRODUCTION

The maximum entropy theory of ecology (METE) is built upon a state variable description of ecosystems (Harte *et al.* 2008, 2009; Harte 2011; Harte & Newman 2014). The state variables used in the original formulation of METE are the area of an ecosystem, A_0, the total number of species, S_0, within some taxonomic group such as plants or birds that are found in that area, the total number of individuals, N_0, in those species and the total metabolic rate of all those individuals, E_0. Knowledge of numerical values of these state variables provides the constraints that are the input to the maximum information entropy (MaxEnt) inference calculation. The outputs of the theory are predictions of many of the metrics of macroecology: the functional forms describing patterns in spatial distribution, abundance and energetics within and across species.

We refer to this original version of METE, with state variables A_0, S_0, N_0 and E_0, as the ASNE model. Consistent with that designation, a model in which those state variables are augmented with G_0, the number of genera, will be referred to as the AGSNE model of METE. Most analyses of census data are carried out using species and individuals as the units of analysis. Census data that are resolved to species, however, can also be viewed at the level of genus or family. Hence, many other macroecological metrics can be defined and will be derived here, including the distribution of abundances over species within a genus with some specified number of species, the distribution of metabolic rates over individuals within a species that is in a genus with some specified number of species, and the relationship between the total metabolic rate of a species and the species richness of the genus to which it belongs. We also derive the form of one additional metric, the distribution of species richness over genera or families, that has received theoretical attention (Yule 1925; Scotland & Sanderson 2004; Etienne *et al.* 2012; Maruvka *et al.* 2013; Stadler *et al.* 2014) but not, as is achieved here, within a unified theoretical framework that also predicts the other metrics listed above.

The MaxEnt concept and its applications to ecology are thoroughly described elsewhere (Harte 2011; Harte & Newman 2014) but the essential idea is as follows. MaxEnt is a procedure for deriving least-biased probability distributions that are consistent with prior knowledge expressed as constraints on that distribution (Jaynes 1957, 1982). 'Constraints' refers, for example to values of some set of moments of the distribution and 'least-biased' here means that the resulting distributions do not embody information other than the known constraints. Graphically, the MaxEnt solutions are the smoothest, flattest possible distributions that satisfy the constraints. The method of Lagrange multipliers (Stewart 2007) is used to derive those least-structured distributions.

The middle column of Table 1 presents the functional forms of the non-spatially explicit macroecological metrics predicted by ASNE (Harte 2011). The only parameters appearing in these functions are the Lagrange multipliers, λ_1 and λ_2, the

[1]*Energy and Resources Group, University of California at Berkeley, Berkeley, CA 94720, USA*

[2]*Department of Environmental Science, Policy and Management, University of California at Berkeley, Berkeley, CA 94720, USA*

[3]*Department of Statistical Science, Cornell University, Ithaca, NY 14850, USA*

**Correspondence: E-mail: jharte@berkeley.edu*

numerical values of which are determined by the constraints derived from ratios of the state variables S_0, N_0, E_0.

Numerous empirical tests indicate support for most of the predictions of the ASNE model of METE. Tests of the predicted species abundance distribution (SAD, 2nd row, middle column, of Table 1) indicate that the majority of results from censuses are more consistent with the predicted log series form of the SAD than with a widely considered alternative, the log-normal distribution, although exceptions exist (Harte *et al*. 2008, 2009; Harte 2011; White *et al*. 2012; Newman *et al*. 2014; Xiao *et al*. 2015). Similarly, the distribution of metabolic rates across all individuals predicted by ASNE (5th row, second column in Table 1) is in good agreement with the tests that have been carried out (Harte 2011; Newman *et al*. 2014; Xiao *et al*.). Finally, the spatial metrics in the ASNE version of METE predict that all species-area relationships (SARs) collapse onto a universal curve if the local (i.e. at specified area) slope of the log-log SAR is plotted against the ratio of total abundance to species richness at that same area; this prediction has been validated with numerous empirical SARs (Harte *et al*. 2009; Harte 2011). All the above successful predictions of the ASNE model remain valid in the taxonomically extended version of METE.

On the other hand, tests of the other predictions of ASNE often fail. In particular, the ASNE prediction for the intraspecific distribution of metabolic rates across individuals within species of specified abundance (8th row in Table 1) has been shown in two studies (Newman *et al*. 2014; Xiao *et al*. 2015) to deviate from empirical data. Related to this, ASNE predicts a form of the energy equivalence principle (Nee *et al*. 1991; Enquist *et al*. 1998; White *et al*. 2007). In this form, energy equivalence asserts an inverse relationship between the

Table 1 The forms of some of the metrics predicted by ASNE (column 2) and AGSNE (column 3)

Metric	Predicted in ASNE	Predicted in AGSNE
Distribution of species richness over genera: $\Gamma(m)$		$\Gamma(m) \approx \frac{e^{-\lambda_1 m}}{m\ln(1/\lambda_1)}$
Distribution of abundances across species (SAD)	$\varphi(n) \approx \frac{e^{-\beta n}}{n\ln(\frac{1}{\beta})}$	$\varphi(n) \approx \frac{\lambda_1 e^{-(\lambda_1+\beta n)}}{n\ln(1/\beta)(1-e^{-(\lambda_1+\beta n)})}$
Distribution of species abundances for all the species in a genus with m species		$\varphi(n\|m) \approx \frac{e^{-\beta mn}}{n\ln(1/\beta)}$
Mean and variance of the abundances of the species in a genus with m species		$<n\|m> \approx \frac{e^{-\beta m}}{\beta m\ln(1/\beta m)} \approx \frac{N_0}{G_0 m}$ $\sigma^2(n\|m) \approx \frac{N_0^2\ln(1/\beta)}{G_0^2 m^2}$
Distribution of metabolic rates across all individuals in community	$\Psi(\varepsilon) \approx \frac{\beta\lambda_2 e^{-\gamma(\varepsilon)}}{(1-e^{-\gamma(\varepsilon)})^2}$	$\Psi(\varepsilon) \approx \frac{\beta\lambda_3}{\ln(1/\lambda_1)}\sum_m \frac{m e^{-m(\lambda_1+\gamma(\varepsilon))}}{(1-e^{-\gamma(\epsilon)})^2} \approx \frac{\beta\lambda_3}{\gamma^2(\varepsilon)}$
Distribution of metabolic rates over the individuals in a species, selected at random from the pool of all species (ASNE), or in a genus with m species (AGSNE)	$\nu(\varepsilon) \approx \frac{\lambda_2}{\ln(1/\beta)}\frac{e^{-(\lambda_2(\varepsilon-1)+\beta)}}{(\lambda_2(\varepsilon-1)+\beta)}$	$\xi(\varepsilon\|m) \approx \frac{\lambda_3}{\ln(1/\beta m)}\frac{e^{-(\lambda_3 m(\varepsilon-1)+\beta m)}}{(\lambda_3(\varepsilon-1)+\beta)}$
Mean and variance of the $\nu(\varepsilon)$ and $\xi(\varepsilon \mid m)$ distributions	$<\varepsilon> \approx \frac{1}{\lambda_2\ln(1/\beta)}$ $\sigma^2(\varepsilon) \approx \frac{1}{\lambda_2^2\ln(1/\beta)}\left(1-\frac{1}{\ln(1/\beta)}\right)$	$<\varepsilon\|m> \approx \frac{1}{(m\lambda_3)\ln(1/\beta)}$ $\sigma^2(\varepsilon\|m) \approx \frac{1}{(m\lambda_3)^2\ln(1/\beta m)}\left(\frac{1}{2}-\frac{1}{\ln(1/\beta m)}\right)$
Distribution of metabolic rates across individuals in a species with n individuals (ASNE), or the individuals in the species with n individuals that are in a genus with m species (AGSNE)	$\Theta(\varepsilon\|n) = \lambda_2 n e^{\lambda_2 n(\varepsilon-1)}$	$\Theta(\varepsilon\|m,n) = \lambda_3 mn e^{-\lambda_3 mn(\varepsilon-1)}$
Dependence on species abundance of metabolic rates averaged over individuals within species (ASNE) or within species in a genus with m species (AGSNE)	$<\varepsilon\|n> = 1+\frac{1}{n\lambda_2}$	$<\varepsilon\|m,n> = 1+\frac{1}{mn\lambda_3}$

Note that the symbols β and γ, which play a parallel role in the two models, are each defined differently in columns 2 and 3; in column 2, $\beta = \lambda_1 + \lambda_2$ and $\gamma = \lambda_1 + \lambda_2\varepsilon$, whereas in column 3, $\beta = \lambda_2 + \lambda_3$ and $\gamma = \lambda_2 + \lambda_3\varepsilon$. The variable m is the number of species in a genus. Expressions determining the Lagrange multipliers, λ_i, in AGSNE are given in Table 2.

local population density of a species and the average metabolic rate of its individuals. Restating this, species within a community all have the same total (summed over individuals within the species) metabolic rate. Assuming the metabolic rate of an individual varies as mass to the ¾ power (Brown *et al*. 2004), the related Damuth (1981) rule follows: the population density of a species is inversely proportional to the ¾ power of the average mass of individuals in the species. Log-log plots of abundance vs. metabolic rate tend, however, to either exhibit considerable scatter around, or deviate entirely from, the prediction (Marquet *et al*. 1990; White *et al*. 2007; Newman *et al*. 2014; Xiao *et al*. 2015).

An additional rationale for extending METE beyond ASNE derives from several publications revealing systematic relationships between the species richness of higher taxonomic categories and the macroecological patterns exhibited by the species in those categories. For example Ter Steege *et al*. (2013) show that the most abundant Amazonian tree species belong to genera that contain relatively few species. Schwartz & Simberloff (2001) and Lozano & Schwartz (2005) show that rare vascular plant species are over-represented in species-rich families. Smith *et al*. (2004) show that species of mammals with the largest body sizes, and therefore largest metabolic rates of individuals, belong to genera with the fewest species. Moreover, the variance of body size across species is also greatest in mammalian genera with the fewest species (Smith *et al*. 2004). We show here that the taxonomically extended version of METE predicts all these general trends.

In Materials and Methods, we explain how the extended theory is defined and constructed, and list the data sources for testing predictions. In Theoretical Results, we extract and discuss testable ecological predictions from the derived metrics, and, in Comparisons with Data, we examine the validity of the predictions. In the Discussion, we speculate on interpretations and implications of our findings. We conclude with a summary of the current status of METE. In Supplementary Material we provide mathematical details of the derivations that are left out in the main text and discuss two possible extensions of the theory.

MATERIALS AND METHODS

To illustrate our approach to theory construction, we focus here on the specific case of adding genus as a category, and thus constructing a realisation of METE that we denote by AGSNE; the method readily generalises to other taxonomic categories.

The extended ecological structure function

An 'ecological structure function', denoted $R(n, \varepsilon \mid S_0, N_0, E_0)$, is the core of ASNE (Harte *et al*. 2008). R is a joint conditional distribution over abundance (n) and metabolic rate (ε) defined so that $R{\cdot}d\varepsilon$ is the probability that if a species is picked at random from the species pool, then it has abundance n, and if an individual is picked at random from that species, then its metabolic energy requirement is in the interval $(\varepsilon, \varepsilon + d\varepsilon)$.

To construct AGSNE, we augment the list of state variables, A_0, S_0, N_0, E_0, by adding G_0, the number of genera in area A_0. In analogy to R, a new joint, conditional probability distribution, $Q(m,n,\varepsilon \mid G_0, S_0, N_0, E_0)$, can be defined by:

> Pick a genus at random from the pool of genera; then $Qd\varepsilon$ is the probability it has m species and, if you pick one of those species, that it has n individuals, and that if you pick one of those individuals from that species, that it has metabolic rate in the interval $(\varepsilon, \varepsilon + d\varepsilon)$.

Note that Q, the ecological structure function in the extended theory, is a function of the discrete variables, m and n, and a continuous variable ε. For notational simplicity we will use the term 'distribution' regardless of whether the independent variable is discrete or continuous, with the understanding that in the latter case a probability density function is intended.

Constraints, metrics and lagrange multipliers

Table 2 summarises the equations that determine the predictions of AGSNE. The first three rows in Table 2 describe the constraints on Q derived from ratios of state variables. The 4th row describes the form of the structure function derived from applying the maximum entropy criterion with those constraints. The next six rows show how the macroecological metrics in AGSNE are determined from Q. And the last three rows show how the Lagrange multipliers that Q depends upon are determined from the constraints. Further explanations of all the entries in the third column of Table 2 and of how the actual derived metrics in the third column of Table 1 are calculated from the entries in Table 2 are in Supplementary Material.

Data sources

To test the predicted distribution of species richness over higher taxonomic categories, $\Gamma(m)$, we use 30 data sets spanning arthropods, birds, plants and microorganisms. Arthropod data are from Basset *et al*. (2012) and Gruner (2007), bird data consisting of 10 transects chosen randomly from the Breeding Bird Survey (Sauer *et al*. 2014), plant data from census plots at Cape Point Preserve (J. Slingsby, pers. comm.) and the Smithsonian Tropical Forest Research Institute plots at BCI (Condit 1998; Condit *et al*. 2004; Condit *et al*. 2012), Luquillo (Thompson *et al*. 2002), Sherman and Cocoli (Pyke *et al*. 2001; Condit *et al*. 2004), and Yasuni (Valencia *et al*. 2003, 2004), and microbiome data from Wu *et al*. (2013) and L. Smarr (pers. comm.).

To test the predicted modification in the energy equivalence rule, we use census data from an intertidal invertebrate community (Marquet *et al*. 1990) and BCI data (Condit 1998; Condit *et al*. 2004; Condit *et al*. 2012). To test a predicted relationship between the species richness of a higher taxonomic category and the abundance distribution of the species in that category we use BCI data (Condit 1998; Condit *et al*. 2004; Condit *et al*. 2012). Finally, we compare qualitatively our predictions regarding the dependence on genera species richness of the distributions of both abundances and metabolic rates against observed trends reported by Smith *et al*. (2004), Schwartz & Simberloff (2001), Lozano & Schwartz (2005) and Ter Steege *et al*. (2013).

Table 2 The structure of AGSNE, including constraint equations, the extended structure function derived from those constraints, the relationship between the metrics and the structure function, and the equations determining the Lagrange multipliers

Category	Description	Determining equation
Constraints	Average number of species per genus	$<m> = \frac{S_0}{G_0} = \sum_{m,n,\varepsilon} mQ(m,n,\varepsilon)$
	Average number of individuals per genus	$<n_G> = \frac{N_0}{G_0} = \sum_{m,n,\varepsilon} mnQ(m,n,\varepsilon)$
	Average metabolic rate per genus	$<\varepsilon_G> = \frac{E_0}{G_0} = \sum_{m,n,\varepsilon} mn\varepsilon Q(m,n,\varepsilon)$
Structure function	Distribution over species richness of genera (m), abundances of species (n), and metabolic rates of individuals (ε)	$Q(m,n,\varepsilon) = \frac{1}{Z(\lambda_1,\lambda_2,\lambda_3)} e^{-\lambda_1 m} e^{-\lambda_2 mn} e^{-\lambda_3 mn\varepsilon}$
		$Z(\lambda_1,\lambda_2,\lambda_3) = \sum_{m,n,\varepsilon} e^{-\lambda_1 m} e^{-\lambda_2 mn} e^{-\lambda_3 mn\varepsilon}$
Metrics	Distribution of species richness over genera	$\Gamma(m) = \sum_{n,\varepsilon} Q(m,n,\varepsilon)$
	Species abundance distribution	$\varphi(n) = \frac{\sum_{m,\varepsilon} mQ(m,n,\varepsilon)}{\sum_{m,n,\varepsilon} mQ(m,n,\varepsilon)} = \frac{G_0}{S_0} \sum_{m,\varepsilon} mQ(m,n,\varepsilon)$
	Distribution of abundances over species in a genus with m species	$\varphi(n\|m) = \frac{\Sigma_\varepsilon Q(m,n,\varepsilon)}{\Sigma_{n,\varepsilon} Q(m,n,\varepsilon)} = \frac{\Sigma_\varepsilon Q(m,n,\varepsilon)}{\Gamma(m)}$
	Distribution of metabolic rates over all individuals	$\Psi(\varepsilon) = \frac{\sum_{m,n} mnQ(m,n,\varepsilon)}{\sum_{m,n,\varepsilon} mnQ(m,n,\varepsilon)} = \frac{G_0}{N_0} \sum_{m,n} mnQ(m,n,\varepsilon)$
	Distribution of metabolic rates over individuals in a species with n individuals and in a genus with m species	$\theta(\varepsilon\|m,n) = \frac{Q(m,n,\varepsilon)}{\sum_\varepsilon Q(m,n,\varepsilon)} = \frac{Q(m,n,\varepsilon)}{\Gamma(m)\varphi(n\|m)}$
	Distribution of metabolic rates over individuals in a species that is in a genus with m species	$\xi(\varepsilon\|m) = \sum_n \Theta(\varepsilon\|m,n)\varphi(n\|m)$
Lagrange multipliers	λ_1	$S_0/G_0 = 1/\lambda_1 \ln(\frac{1}{\lambda_1})$
	$\beta = \lambda_2 + \lambda_3$	$N_0/G_0 = 1/\beta \ln(\frac{1}{\beta})$
	λ_3	$G_0/(E_0 - N_0) = \lambda_3$

For notational convenience we denote integrals over ε with sums over ε in the table, but in the actual calculations we integrate over ε. As in ASNE, we define the unit of energy such that $\varepsilon = 1$ is the lowest metabolic rate among the N_0 individuals. And we leave the limits off of summations, which are understood to range from 1 to S_0, N_0, E_0 for m, n, ε respectively.

THEORETICAL RESULTS

The third column in Table 1 shows the results, for the AGSNE model of METE, of the calculations for each of the metrics defined in Materials and Methods. Some of the metrics in the third column of Table 1 yield predictions about phenomena that are not predicted in ASNE (second column of Table 1). These metrics include $\Gamma(m)$, $\varphi(n \mid m)$, $\xi(\varepsilon \mid m)$ and $\Theta(\varepsilon \mid n,m)$. In this section, we discuss the most salient properties of these new predictions. Other metrics derived here are also predicted in ASNE, but because the constraints imposed here include additional prior information not in ASNE, AGSNE yields different predictions for these same metrics. We also discuss these differences here.

The distribution of species richness across higher taxonomic categories: $\Gamma(m)$

Entry 1 in Table 1 shows that the AGSNE model predicts an approximate log series distribution of species richness values across genera. The same prediction would then hold for the distribution of species richness values across families in the AFSNE model. Numerical evaluations of the exact expression in Supplementary Material differ by at most a per cent or two from the log series distribution (eqns S-24 vs. S-25) for realistic combinations of state variables.

The species abundance distribution: $\varphi(n)$

The species abundance distribution, $\varphi(n)$, predicted in AGSNE is given in the 2nd entry in Table 1. For realistic combinations of state variables, $\varphi(n)$ is closely approximated by the log series SAD, predicted by ASNE, as discussed in the Supplementary Material (eqns S-26 vs. S-28).

Abundance distributions within genera: $\varphi(n \mid m)$

The 3rd entry in Table 1, $\varphi(n \mid m)$, is the distribution of abundances over the set of species in a genus with m species. This is a slightly modified log series, with its mean and variance (4th entry in Table 1) now dependent on m. Thus, the fraction of rare or very abundant species in a genus will depend on the species richness of the genus. For fixed values of the state variables, genera with fewer species will tend to have more abundant species and the variance of abundances of those species will be larger. The dependence on m of mean and variance vary approximately as $1/m$ and $1/m^2$ respectively.

The expected fraction of singleton species (species with 1 individual) in a genus with m species, given by $\varphi(1 \mid m)$, is as follows:

$$\varphi(1|m) \approx \frac{e^{-\beta m}}{\ln(1/\beta m)}. \quad (1)$$

Here, β is determined from the MaxEnt constraint (eqn S-23) and is a function of G_0 and N_0. Both the numerator and the denominator are decreasing functions of m, but the numerator decreases more slowly, and so the fraction of singleton species in a genus with m species is an increasing function of m. Our $\varphi(n \mid m)$ also predicts that the abundance of the most abundant species in a genus with m species varies approximately (see Supplementary Material) as:

$$n_{max} \approx \frac{0.56}{\beta m(1-\beta m)^{0.643}}. \quad (2)$$

Because βm is of order $S_0/N_0 << 1$, n_{max} varies to an excellent approximation as $1/m$.

The distribution of metabolic rates over all individuals: $\Psi(\varepsilon)$

Numerical evaluation of the exact AGSNE prediction (eqn S-40) shows that this metric can be well approximated by the summation in the 5th entry in Table 1. AGSNE predicts a slightly larger proportion of species with small values of ε than does ASNE (second column of Table 1). Paralleling the situation with the SAD, $\Phi(n)$, the differences between the ASNE and AGSNE predictions are too small to alter the results of previous testing of these metrics.

Metabolic rate distribution within genera: $\xi(\varepsilon \mid m)$

The 6th entry in Table 1 indicates an approximately exponentially decreasing distribution of metabolic rates across the individuals in species belonging to genera with m species, with the rate constant in the exponent proportional to m. ASNE, in contrast, predicts that the rate constant is a true constant independent of m. We note that this ASNE function, $\nu(\varepsilon)$, was incorrectly defined and derived in Harte (2011); the expression in Table 1 should replace the expression given in that book.

From $\xi(\varepsilon \mid m)$, the mean and the variance of the metabolic rates of individuals in species belonging to genera with m species can be derived and are shown in the 7th entry in Table 1; their m-dependence is $1/m$ and $1/m^2$ respectively.

At large ε, $\xi(\varepsilon \mid m)$ has the form of a log series distribution:

$$\xi(\varepsilon|m) \approx \frac{e^{-\lambda_3 m\varepsilon}}{\varepsilon \ln(\frac{1}{\beta m})} \quad (3)$$

and thus genera with few species are more likely to contain individuals with high metabolic rates. The fraction of individuals within a species in a genus with m species that metabolise at the lowest rate, $\varepsilon = 1$, is given by $\xi(1|m) \approx \frac{e^{-\lambda_3 m}}{\ln(\frac{1}{\beta m})}$. For realistic combinations of state variables this is an increasing function of m (see discussion following eqn S-45 in SM). Hence, just as the fraction of low-abundance species is predicted to be greatest in genera with many species (eqn 1), so is the fraction of very small individuals in the species within such genera.

Energy equivalence in *GSNE*

Consider next the function, $\Theta(\varepsilon \mid m,n)$, the distribution of metabolic rates within species that have abundance n and that belong to a genus with m species. The shape of the AGSNE prediction for this metric, the 8th entry in Table 1, differs from the shape of the ASNE prediction for $\Theta(\varepsilon \mid n)$ by the factor of m appearing in the exponent. From the distribution,

we can evaluate (the last entry in Table 1) the average metabolic rate of the individuals that are in a genus with m species and a species with n individuals:

$$<\varepsilon|m,n> = \sum_{\varepsilon} \varepsilon\Theta(\varepsilon|m,n) = 1 + \frac{1}{\lambda_3 mn}. \quad (4)$$

The expected total metabolic rate of all individuals in a species with n individuals that is in a genus with m species is

$$n<\varepsilon|m,n> = n + \frac{1}{\lambda_3 m} \quad (5)$$

and hence, for all combinations of n and m such that $nm << 1/\lambda_3 = (E_0 - N_0)/G_0$, the total metabolic rate of each species is inversely proportional to m and independent of n.

In ASNE, the metric $\Theta(\varepsilon \mid n)$ was defined analogously: the distribution of metabolic rates over the individuals in a species with abundance n, and its predicted form is given in Table 1. It results in the following expression for the average energy of the individuals in a species with abundance n:

$$n<\varepsilon|n> = n + \frac{1}{\lambda_2} \quad (6)$$

For those species for which $n << 1/\lambda_2 = (E_0 - N_0)/S_0$, eqn 6 is a statement of energy equivalence (White *et al*. 2007): the average metabolic rate of the individuals in a species is inversely proportional to abundance. Using the relationship that metabolic rate $\sim$ mass$^{3/4}$, energy equivalence is another statement of the Damuth rule (Damuth 1981) relating abundance to the inverse ¾ power of average body mass for individuals in a species.

The difference between eqns 5 and 6 is noteworthy. Equation 5 implies that in AGSNE, if the inequality is satisfied (which is nearly always the case, but see discussion in Harte *et al*. 2008), then on a graph of average energy or body mass of individuals against abundance, the species will not all fall on the universal curve that eqn 6 had predicted. Rather, the species in genera with differing numbers of species will lie on different curves, characterised by the species richness of each genus.

COMPARISONS WITH DATA

We compare predicted to observed properties of the distribution of species richness across genera or families, $\Gamma(m)$, using 30 census data sets spanning a wide variety of types of organisms and habitats. We also conduct two tests, one with invertebrate data and one with BCI data, of the modified energy equivalence rule. Then we again use BCI data to test the predicted dependence of species abundance on the species richness of the genus the species is in. And finally we compare qualitatively our theoretical results with available published summaries of trends in the dependence of species abundance or body size on the species richness of genera.

The distribution of species richness across higher taxonomic categories

To test the predicted form of $\Gamma(m)$, we examined a broad spectrum of data sets that included taxonomic data for various communities of plants, birds, arthropods or microorganisms. The results are shown in Fig. 1. In each of the data sets, the species list used to compile values of m describes a local community that is either obtained from censusing a plot, a transect, a canopy, or in the case of microorganisms, the human gut.

For each of the data sets, we have compared prediction vs. observation for three features of the distribution: (1) the number of higher level taxa with exactly one species, denoted $m = 1$ in the figure legend; (2) the number with fewer than 10 species, denoted $m < 10$; and (3) the number of species in the most species-rich family (or genus in the case of birds), denoted 'maximum m'. For each data set, each prediction is uniquely determined from knowledge of the total number of species and families (or genera) in the set. We focus in detail on the tails of the distribution because that is often where theoretical predictions fail, and, if alternative distributions are proposed, where they are likely to differ most. Excellent agreement between observation and prediction is evident in Fig. 1 and characterised in the figure caption.

The modified energy equivalence rule

AGSNE predicts that the relationship between abundance and body size is not a universal curve but rather a family of curves, each characterised by a different genus- or family level

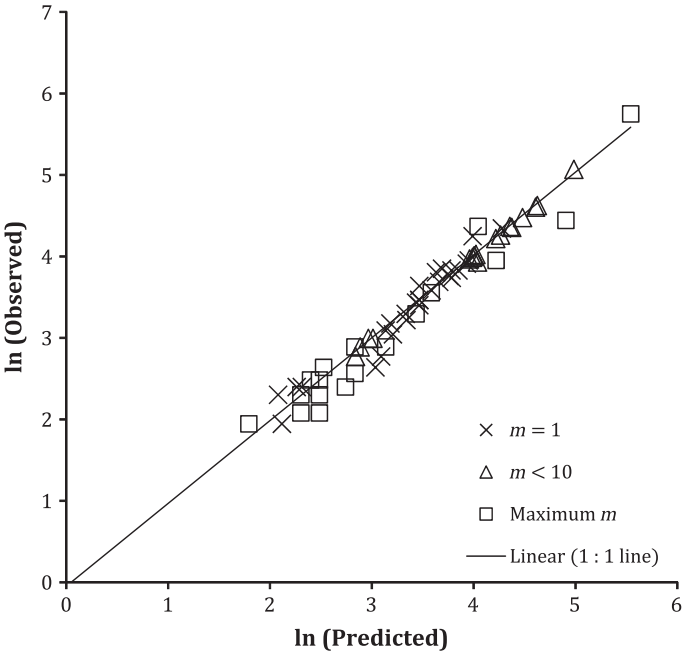

Figure 1 Comparison of observed and predicted distributions of species, $\Gamma(m)$ (first entry in Table 1), across families for arthropods, plants and microorganisms and across genera for birds. For each of the 30 data sets graphed, three types of data are plotted: numbers of families (or genera) with one species, numbers of families (or genera) with fewer than 10 species and the number of species in the most species-rich family (or genus). For those three types of data, theory predicts 94, 99 and 95% of the variance respectively; regression slopes of observed against predicted are equal to 1 and intercepts are equal to 0 within 95% confidence intervals.

value of species richness (eqn 5). Equivalently, it predicts that the total metabolic rate of a species, given by the product of its abundance and its average individual metabolic rate, should vary inversely with the species richness of the genus or family containing the species.

We carried out two tests of this prediction. First, we turned to a classic study examining the size-abundance relationship in an invertebrate intertidal community (Marquet *et al.* 1990). Data from that study showed considerable scatter in a log (abundance) vs. log(body size) plot, although the central tendency lay along a line with slope of $\sim -3/4$, as predicted by the Damuth rule. Figure 2 shows that our modified energy equivalence rule (eqn 5) resolves at least some of this scatter in the sense that species within the community that are in genera with only one species ($m = 1$) tend to lie above the trend line and those in more species-rich genera tend to lie below as predicted. Letting y = log(abundance), x = log(body size) and w = log(number of species in genus), we evaluated two regression models: i. $y = ax + bw + c$ (this is the AGSNE prediction, if $a = -3/4$, $b = -1$), and ii. $y = ax + c$ (this is the unmodified Damuth rule, as predicted by ASNE, if $a = -3/4$). In model i, $a = -0.70 \pm 0.10$ (SE) and $b = -1.18 \pm 0.43$, indicating consistency with the AGSNE prediction, and the adjusted R^2 is 0.53. In model ii, the adjusted R^2 is 0.45; model comparison supports the AGSNE model (i) over the ASNE model (ii), with $F(1, 3) = 10.3$, $P < 0.05$, $n = 47$.

An additional model, $y = ax + bw + cx \cdot e^w + d$, with the cross-term indicating m-dependence of the mass-abundance scaling exponent, has an adjusted R^2 of 0.51 and the coefficient multiplying the cross-term is -0.046 ± 0.056.

For an additional test, we used BCI tropical tree data (Condit 1998; Condit *et al.* 2004). Here, we assume that the metabolic rates of trees are proportional to their basal areas. In Fig. 3a, b, we plotted log(total metabolic rate of species) against log(species richness of the family the species is in) for unbinned and binned data respectively. In Fig. 3a, the fitted slope is -0.46 ± 0.17, with $P = 0.0067$ for zero dependence on family species richness. Thus for this data set, the fitted slope for dependence of total species level metabolic rate on family species richness is significantly greater than the predicted slope of 0 under ASNE, but less than the predicted slope of -1 under AFSNE.

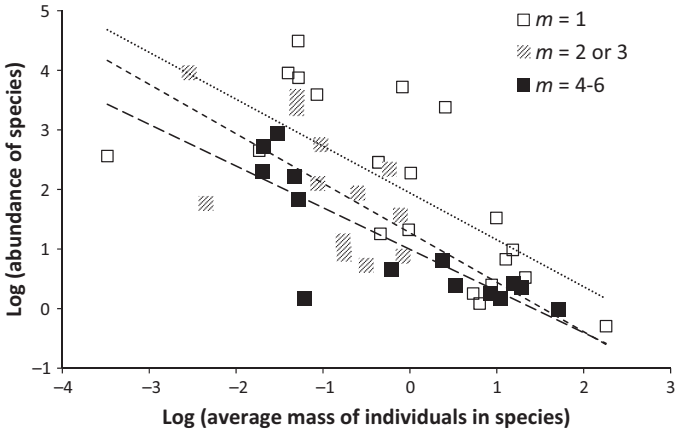

Figure 2 Log–log plot of density of individuals vs. average individual body mass for each species in an intertidal invertebrate community in Chile. The straight lines are power-law fits of density vs. body mass. The dotted line corresponds to $m = 1$, the short-dashed line to $m = 2$ or 3, and the long-dashed line to $m = 4$–6. The fitted slopes of the power-law fits are -0.79, -0.83 and -0.70 for those three cases respectively; see text for a comparison of ASNE and AGSNE regression models. Data are from Marquet *et al.* (1990).

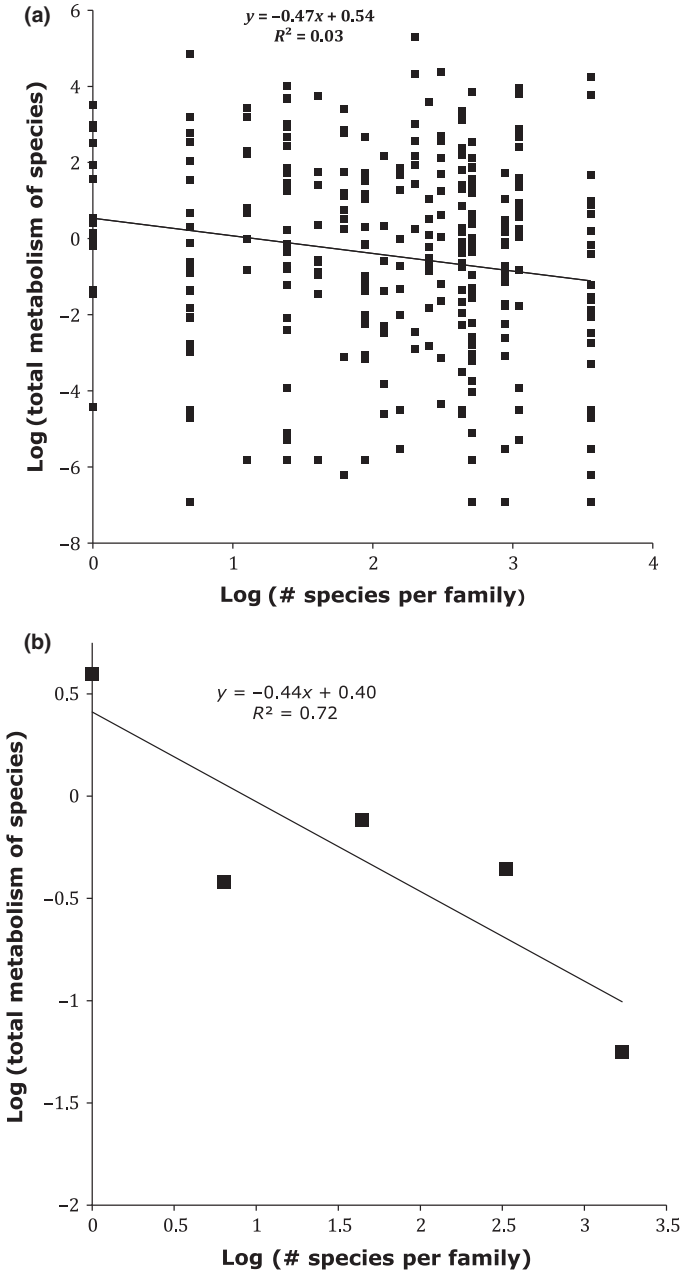

Figure 3 Test of the modified energy equivalence rule for tree species at BCI (Condit 1998; Condit *et al.* 2004; Condit *et al.* 2012), using all species from the 1982 census. (a) Total metabolic rate of each species is plotted against m, the species richness of the family the species belongs to. (b) Same as 3a except data are binned into logarithmic intervals of m. The metabolic rate of each censused tree is taken to be proportional to its basal area and the total metabolic rate of a species is then proportional to the summed basal area of all individuals in the species. Perfect agreement with the AFSNE version of METE would result in a slope of -1. The original ASNE form of the energy equivalence rule predicts a slope of 0.

Abundance distributions within genera

From the function $\varphi(n \mid m)$ describing the abundance distribution of species in genera with m species, we predict that the mean and standard deviation of this distribution should decrease inversely with m. Tree abundance data from the BCI 50 ha plot (Fig. 4) indicate that both mean and standard deviation are indeed decreasing function of species richness, although the observed dependences on m are weaker than predicted (see Figure Caption).

The predicted form of $\varphi(n \mid m)$ also implies that rare species are proportionally more likely to occur in genera or families with more species (eqn 1). This is qualitatively consistent with the findings of Schwartz & Simberloff (2001) and Lozano & Schwartz (2005), who showed that families with few species of vascular plants worldwide tend to have fewer than expected numbers of low-abundance species. At the other tail of the abundance distribution, we predict that the most abundant species are most likely to be found in the least speciose genera or families, with $n_{\max}$ decreasing inversely with m. This is qualitatively consistent with the findings of Ter Steege *et al.* (2013), who examined hyperdominance in Amazonian trees and found that the most abundant species were more likely than expected by chance to be found in the least speciose genera.

The distribution of metabolic rates over the individuals in species found in genera with *m* species

From the function $\xi(\varepsilon \mid m)$ (6th entry in Table 1) we predict that species with the largest metabolic rates (or body sizes) should tend to be found in genera with the fewest species. Moreover, the mean and standard deviation of the distribution of metabolic rates of species in a genus with m species should vary inversely with m (7th entry in Table 1. These predictions are qualitatively consistent with the findings of Smith *et al.* (2004), who examined such patterns in body sizes of mammals; species of mammals with the largest body sizes, and therefore largest metabolic rates of individuals, belong to genera with the fewest species. Moreover, the variance of the distribution of body sizes across species in genera with m species was found to be greatest in genera with the fewest species.

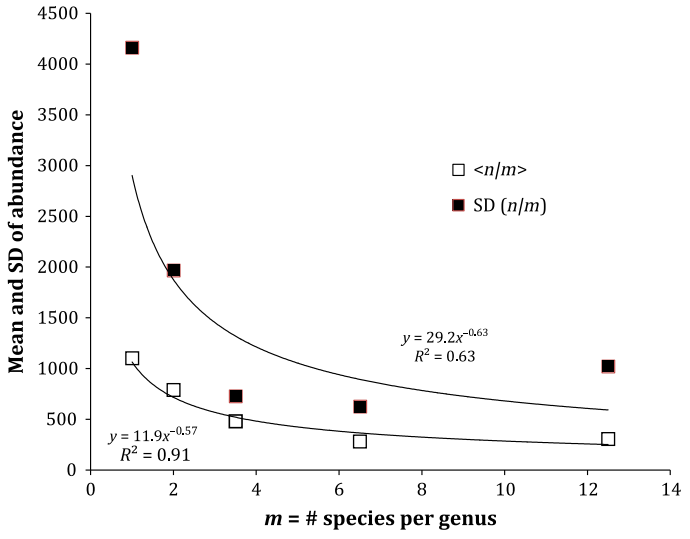

Figure 4 Dependence of the mean and standard deviation of the abundance of BCI trees in genera with m species (Condit 1998; Condit *et al.* 2004; Condit *et al.* 2012). Values of m are grouped in logarithmic intervals but plotted on a linear scale to better reveal discrepancies. The lines are power-law regression lines. The estimated mean and standard error of the exponents are -0.63 ± 0.28 and -0.57 ± 0.10.

DISCUSSION

The taxonomically extended theory connects local community macroecological metrics with the structure of the taxonomic tree describing the taxa in the community. Predictions of ASNE that have previously been confirmed empirically, in particular the SAD and the distribution of metabolic rates over all individuals in the community (Harte *et al.* 2009; Harte 2011; White *et al.* 2012; Newman *et al.* 2014; Xiao *et al.* 2015), as well as the species-area relationship and other spatially explicit metrics, are unmodified in the extended theory.

The extended theory predicts that genera or families with relatively few species are disproportionately likely to contain species with individuals having disproportionately larger body sizes and contain species having disproportionately more individuals. Referring to Fig. 2, these species are further from the origin in the upper right quadrant of the graph. Species-rich genera or families are predicted to have the converse: smaller body sizes and more rare species.

These predictions can be understood within the context of competition and ecological dominance. In particular, our predictions are consistent with the perspective that competition is keenest among taxonomically related species, and ecologically dominant species (with either large body sizes or large abundances) belong to lineages that have failed to diversify. If further testing reveals that the shape of the taxonomic tree influences the macroecology of the tips of the tree in conformity with the predictions of the extended MaxEnt-based theory, as we have suggested here, then this extended theory may shed light on the processes that help stabilise communities under strong competitive forces acting between species within species-rich genera.

The extended theory also predicts a log series distribution of species richness over genera or families, implying many genera or families with few species and few with many species. This pattern has been noted for many decades and models have been advanced to explain it (Yule 1925; Scotland & Sanderson 2004; Etienne *et al.* 2012; Maruvka *et al.* 2013; Stadler *et al.* 2014), but to our knowledge this work is the first to derive it from theory that also predicts the dependence of many other metrics of macroecology (Tables 1 and 2) on the structure of the taxonomic tree. The predicted distribution of species richness values over families or genera is in excellent agreement with observations (Fig. 1).

The predicted dependence of each species' total metabolic rate on the species richness of the genus the species belongs to is also in good agreement with intertidal invertebrate data (Fig. 2). And qualitatively, the dependence of abundance and

body size means and variances on the species richness of higher taxonomic categories are consistent with multiple observations (Schwartz & Simberloff 2001; Smith *et al*. 2004; Lozano & Schwartz 2005; Ter Steege *et al*. 2013).

The BCI data (Figs 3 and 4), however, reveal a weaker than predicted dependence of total species metabolic rates, and of average abundance of species, on the species richness of higher taxonomic categories. Interestingly, comparison of BCI abundance data with the predicted log series abundance distribution from either ASNE or AGSNE also reveals a relatively sizeable deviation from theory (Harte 2011). As discussed elsewhere (Harte 2011; Harte & Newman 2014), the predictions of METE often fail in ecosystems that are changing relatively rapidly as, for example in response to disturbance. Thus, the discrepancy between the predicted size-abundance relationship and the BCI census data may be a consequence of dispersal limitation at that site after the formation of Gatun Lake and resulting isolation of the plot from a large mainland source pool.

While taxonomic trees are not precise representations of the evolutionary history leading up to the community, the extended theory represents an attempt at bridging ecological and evolutionary metrics in the maximum entropy framework. Connections between evolution and ecology (Webb *et al*. 2002; Kraft *et al*. 2007; Cavender-Bares *et al*. 2009; Graham *et al*. 2009) have been insightful but the proper null model against which to test alternate hypotheses remains unclear (Swenson *et al*. 2006; Cavender-Bares *et al*. 2009). The maximum entropy framework used here may be able to provide powerful null models of phylogenetic community structure.

Our results could potentially also be of use in palaeoecology, where often individuals are only identified to genus or family (Webb 2013). In such a situation, if body size data from the fossil record are available, then from a fit of the predicted distribution of metabolic rates across individuals in any genus, the number of species in that genus might be calculable from $\xi(\varepsilon \mid m)$, for in that function the number of species in the genus, m, explicitly determines the shape of the function.

CONCLUSION

We have extended a state variable theory of macroecology constructed from the MaxEnt inference procedure to include an additional state variable characterising a taxonomic category above the species level. By preserving the successful predictions of METE, generally improving the unsuccessful ones, and making new successful predictions, our theory provides a glimpse of unified and potentially far-reaching connections between evolution and macroecology.

ACKNOWLEDGEMENTS

We are grateful to Pablo Marquet, Justin Kitzes, Jade Zhang, Erica Newman and Mark Wilber for useful discussions, to Larry Smarr and Jasper Slingsby for providing access to data, and four anonymous reviewers for constructive comments. Funding was provided by the Gordon and Betty Moore Foundation and by the US NSF through the Graduate Research Fellowship Program and grant NSF-EF-1137685.

AUTHOR CONTRIBUTION

JH conceived the theory. JH, AR and WZ worked out the implications of the theory. WZ and AR carried out analytical and numerical calculations. JH and AR analysed data and wrote the manuscript.

REFERENCES

Basset, Y., Cizek, L., Cuénoud, P., Didham, R.K., Guilhaumon, F., Missa, O., et al. (2012). Arthropod diversity in a tropical forest. *Science*, 338, 1481–1484.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85, 1771–1789.

Cavender-Bares, J., Kozak, K.H., Fine, P.V. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett*., 12, 693–715.

Condit, R. (1998). Ecological implications of changes n drought patterns: shifts in forest composition in Panama. *Clim. Change*, 39, 413–427.

Condit, R., Aguilar, S., Hernandez, A., Perez, R., Lao, S., Angehr, G. *et al*. (2004). Tropical forest dynamics across a rainfall gradient and the impact of an El Nino dry season. *J. Trop. Ecol*., 20, 51–72.

Condit, R., Lao, S., Pérez, R., Dolins, S.B., Foster, R.B. & Hubbell, S.P. (2012). *Barro Colorado Forest Census Plot Data, 2012 Version*. DOI, Center for Tropical Forest Science Databases. doi:10.5479/data.bci.20130603.

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

Enquist, B.J., Brown, J.H. & West, G.B. (1998). Allometric scaling of plant energetics and population density. *Nature*, 395, 163–165.

Etienne, R., d Visser, S., Janzen, T., Olsen, J., Olff, H. & Rosindell, J. (2012). Can clade age alone explain the relationship between body size and diversity? *Interface Focus*, 2, 170–179.

Graham, C.H., Parra, J.L., Rahbek, C. & McGuire, J.A. (2009). Phylogenetic structure in tropical hummingbird communities. *Proc. Natl Acad. Sci*., 106, 19673–19678.

Gruner, D.S. (2007). Geological age, ecosystem development, and local resource constraints on arthropod community structure in the Hawaiian Islands. *Biol. J. Linn. Soc*., 90, 551–570.

Harte, J. (2011). *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford University Press, Oxford, UK.

Harte, J. & Newman, E.A. (2014). Maximum information entropy: a foundation for ecological theory. *Trends Ecol. Evol*., 29, 384–389.

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

Harte, J., Smith, A. & Storch, D. (2009). Biodiversity scales from plots to biomes with a universal species area curve. *Ecol. Lett*., 12, 789–797.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Phys. Rev*., 106, 620–630.

Jaynes, E.T. (1982). On the rationale of maximum entropy methods. *Proc. Instit. Elec. Electron. Eng*., 70, 939–952.

Kraft, N.J., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *Am. Nat*., 170, 271–283.

Lozano, F. & Schwartz, M. (2005). Patterns of rarity and taxonomic group size in plants. *Biol. Conserv*., 126, 146–154.

Marquet, P.A., Navarrete, S.N. & Castilla, J.C. (1990). Scaling population density to body size in rocky intertidal communities. *Science*, 250, 1125–1127.

Maruvka, Y., Schnerb, N., Kessler, D. & Ricklefs, R. (2013). Model for macroevolutionary dynamics. *Proc. Natl Acad. Sci*., 110, E2460–E2469.

Nee, S., Read, A.F., Greenwood, J.J.D. & Harvey, P.H. (1991). The relationship between abundance and body size in British birds. *Nature*, 351, 312–313.

Newman, E.A., Harte, M., Lowell, N., Wilber, M. & Harte, J. (2014). Empirical tests of within- and across-species energetics in a diverse plant community. *Ecology*, 95(10), 2815–2825.

Pyke, C.R., Condit, R., Aguilar, S. & Lao, S. (2001). Floristic composition across a climatic gradient in a neotropical lowland forest. *J. Veg. Sci.*, 12, 553–566.

Sauer, J., Hines, J.Fallon., J. Pardieck, K., Ziolkowski, .D.Jr & Link, W. (2014). *The North American Breeding Bird Survey, Results and Analysis 1966–2012. Version 02.19.2014*. USGS Patuxent Wildlife Research Center, Laurel, MD.

Schwartz, M. & Simberloff, D. (2001). Taxon size predicts rates of rarity in vascular plants. *Ecol. Lett.*, 4, 464–469.

Scotland, R. & Sanderson, M. (2004). The significance of few versus many in the tree of life. *Science*, 303, 643.

Smith, F., Brown, J., Haskell, J., Lyons, K., Alroy, J., Charnov, E. *et al.* (2004). Similarity of mammalian body size across the taxonomic hierarchy and across space and time. *Am. Nat.*, 163(5), 672–691.

Stadler, T., Rabosky, D., Ricklefs, R. & Bokma, F. (2014). On age and species richness of higher taxa. *Am. Nat.*, 184(4), 447–455.

Stewart, J. (2007). *Calculus*. Brooks Cole, Florence.

Swenson, N.G., Enquist, B.J., Pither, J., Thompson, J. & Zimmerman, J.K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology*, 87, 2418–2424.

Ter Steege, H., Pitman, N., Sabatier, D., Baraloto, C., Salomão, R., Guevara, J. *et al.* (2013). Hyperdominance in the Amazonian tree flora. *Science*, 342, 325–333.

Thompson, J., Brokaw, N., Zimmerman, K., Waide, R., Everham, E., Lodge, D. *et al.* (2002). Land use history, environment, and tree composition in a tropical forest. *Ecol. Appl.*, 12, 1344–1363.

Valencia, R., Hernández, C., Villa, G. & Condit, R. (2003). Demographic tree data from the 25-ha Yasuní Forest Dynamics Plot, 1992–2003. CTFS Forest Dynamics Plot Data Series. Quito, Ecuador.

Valencia, R., Foster, R., Villa, G., Condit, R., Svenning, J.-C., Hernandez, C. *et al.* (2004). *J. Ecol.*, 92, 214–229.

Webb, T.I.I.I. (2013). Paleoecology. In: *Encyclopedia of Biodiversity*, Vol. 5. (ed Levine, S.). Academic Press, New York, pp. 645–655.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.

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

White, E.P. Thibault, K. & Xiao, X. (2012). Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, 93, 1772–1778.

Wu, S., Li, W., Smarr, L., Nelson, K., Yooseph, S. & Torralba, M. (2013). Large memory high performance computing enables comparison across human gut microbiome of patients with autoimmune diseases and healthy subjects. In: *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* (XSEDE '13) (Ed. N., Wilkins-Deahr.). ACM, New York, NY, pp. 25–6. doi:10.1145/2484762.2484828.

Xiao, X., McGlinn, D. & White, E. (2015). A strong test of the maximum entropy theory of ecology. *Am. Nat.*, 75, E70–E80. doi:http://www.jstor.org/stable/10.1086/679576.

Yule, G.U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philos. Trans. R. Soc. B*, 213, 21–87.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Jonathan Chase
Manuscript received 14 May 2015
First decision made 17 June 2015
Manuscript accepted 10 July 2015