ARTICLES

Patterns of relative species abundance in rainforests and coral reefs

Igor Volkov[1,2], Jayanth R. Banavar[1], Stephen P. Hubbell[3,4] & Amos Maritan[5]

A formidable many-body problem in ecology is to understand the complex of factors controlling patterns of relative species abundance (RSA) in communities of interacting species. Unlike many problems in physics, the nature of the interactions in ecological communities is not completely known. Although most contemporary theories in ecology start with the basic premise that species interact, here we show that a theory in which all interspecific interactions are turned off leads to analytical results that are in agreement with RSA data from tropical forests and coral reefs. The assumption of non-interacting species leads to a sampling theory for the RSA that yields a simple approximation at large scales to the exact theory. Our results show that one can make significant theoretical progress in ecology by assuming that the effective interactions among species are weak in the stationary states in species-rich communities such as tropical forests and coral reefs.

A variety of patterns have been observed in the RSA distributions, which are measures of the number of species having a given number of individuals, of ecological communities. In particular, tropical forests[1−5] and coral reefs[6,7] exhibit contrasting RSA patterns. In tropical-tree communities there are fewer rare species in the local community than in the metacommunity, whereas the opposite pattern is found in coral reefs. Reference 6 reported log-series-like RSA distributions in local communities, and log-normal-like RSA distributions when a geographically widespread set of coral-reef communities was pooled to estimate the RSA distribution for the metacommunity. In contrast, local tropical-tree communities exhibit log-normal-like RSA distributions, which become more log-series-like at large landscape scales[1−5]. The log-series RSA distribution has a larger proportion of rare species than the log-normal. Here we consider two distinct types of community structure: first, a relatively small semi-isolated local community surrounded by a very large metacommunity acting as a source of immigrants, as in Hubbell's theory[1], and, second, spatially isolated island communities whose assemblage acts as the metacommunity[8,9]. For the tropical forest, the timescale for species turnover in the metacommunity is very long compared to the characteristic timescale for immigration, leading to an effectively frozen metacommunity acting as a backdrop for immigration. In coral reefs, in contrast, each local community receives immigrants from all the surrounding semi-isolated local communities, within each of which the species abundances are not frozen in time. We present a simple unified theory for understanding the RSA patterns of tropical forests and coral reefs.

Coral reefs

Consider a metacommunity consisting of many small semi-isolated local communities, each of which receives immigrants from other local communities. Because of the isolation of the local communities from each other, changes in the RSA distribution of the aggregated metacommunity may be assumed to occur more rapidly than immigration. One may make a simplifying assumption that the immigration parameter Y is species-independent, corresponding to immigration occurring not from a frozen metacommunity, as commonly assumed for a tropical forest, but from a time-averaged metacommunity in a species-symmetric manner. Using equations (12)

and (14) (see Box 1), the mean number of species with n individuals $\langle \varphi_n \rangle$ is given by:

$$\langle \varphi_n \rangle = \theta \frac{x^n}{n!} \Gamma(n+Y) \tag{1}$$

where $\theta = S/[(1-x)^{-Y} - 1]\Gamma(Y)$ is the Hubbell biodiversity number[1], Γ is the gamma function, S is the number of observed species and x is the per capita birth-to-death-rate ratio.

Given the isolation of individual coral reefs under the island metacommunity model, the value of the immigration parameter Y is very small because the local communities are separated from each other by large distances. In such a situation, the RSA for the local communities resembles the Fisher log-series, and does not have an interior mode (at abundance $n > 1$). Now let us gradually assemble the metacommunity RSA distribution by considering the joint RSA distributions of multiple local communities. First consider the joint RSA of two local communities A and B comprising the metacommunity. Consider a species that has n_A individuals in community A with probability $P(n_A)$ and n_B individuals in community B with probability $P(n_A)$. The probability that the species has n individuals in A and B is (see Supplementary Materials):

$$P(n_A + n_B = n) = \sum_{n_A + n_B = n} P(n_A)P(n_B) \propto \frac{x^n}{n!} \Gamma(n+2Y) \tag{2}$$

The corresponding RSA has the same form as for the single community but with the effective immigration parameter $2Y$. Extending the calculation of the joint RSA distribution to more and more local communities, one arrives at the RSA of the metacommunity characterized by an effective immigration parameter LY, where L is the total number of local communities comprising the metacommunity. When L is large, the RSA distribution exhibits a clear, interior mode at abundance $n > 1$, and the rare species constitute a smaller fraction of all the species than in the local community. This mean field analysis does not take into account the actual spatial locations of the local reef communities. Figure 1 shows the fits of the island model to the coral-reef RSA data[6,7]. We can see from the fits that, on local scales (local communities and reefs), immigration is almost absent so

[1]Department of Physics, 104 Davey Laboratory, [2]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [3]Department of Ecology and Evolutionary Biology, The University of California, Los Angeles, California 90095, USA. [4]The Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002, Panama. [5]Dipartimento di Fisica 'G. Galilei', Università di Padova CNISM and INFN, via Marzolo 8, 35131 Padova, Italy.

Box 1 | General framework

We neglect inter-species interactions after the community has reached a steady state, and consider the dynamics of the population of a single species. These dynamics are governed by generalized birth and death events (including speciation, immigration and emigration). $b_{n,k}$ and $d_{n,k}$ represent the probabilities of birth and death, respectively, in the kth species with n individuals with $b_{-1,k} = d_{0,k} = 0$. $P_{n,k}(t)$ denotes the probability that the kth species contains n individuals at time t. In the simplest scenario, the time evolution of $P_{n,k}(t)$ is regulated by the master equation[21]:

$$\frac{\partial P_{n,k}(t)}{\partial t} = P_{n-1,k}(t)b_{n-1,k} + P_{n+1,k}(t)d_{n+1,k} - P_{n,k}(t)(b_{n,k} + d_{n,k}) \quad (8)$$

which leads to the steady-state or equilibrium solution:

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (9)$$

for $n > 0$ and where $P_{0,k}$ can be deduced from the normalization condition $\Sigma_{n \geq 0} P_{n,k} = 1$. One can show that the system is guaranteed to reach the stationary solution (9) in the infinite time limit[22].

Let us consider a simple, ecologically meaningful form for the effective birth and death rates of the kth species:

$$b_{n,k} = b_k(n + \Upsilon_k) \quad (10)$$

and

$$d_{n,k} = d_k n \quad (11)$$

where b_k and d_k denote the per-capita density-independent birth and death rates and a non-zero Υ_k could arise from either immigration or owing to intraspecific interactions such as those giving rise to density dependence[3]. We do not incorporate speciation explicitly into the model because it does not affect the functional form of the results (it can be incorporated into the immigration term at $n = 0$ by adding a constant).

The steady-state solution of the master equation for $P_k(n)$, the probability that the kth species has n individuals, yields a negative binomial distribution[23]:

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} = \frac{(1 - x_k)^{\Upsilon_k}}{\Gamma(\Upsilon_k)} \frac{x_k^n}{n!} \Gamma(n + \Upsilon_k) \quad (12)$$

where $x_k = b_k/d_k$, the ratio of the per-capita birth rate to the per-capita death rate, controls the mean species abundance given by $x_k \Upsilon_k/(1 - x_k)$. Henceforth, we make the neutral ecological equivalence assumption that the per-capita birth and death rates are the same for all the species, that is, $b_k = b$, $d_k = d$ and $x_k = x$, so we can discard the subscript k for b, d and x.

The number of species containing n individuals is given by:

$$\varphi_n = \sum_{k=1}^{S} I_{n,k} \quad (13)$$

where S is the total number of species that may potentially be present in the community and the indicator $I_{n,k}$ is a random variable that takes the value 1 with probability $P_{n,k}$ and 0 with probability $(1 - P_{n,k})$. Thus, the average number of species containing n individuals is given by:

$$\langle \varphi_n \rangle = \sum_{k=1}^{S} I_{n,k} = \sum_{k=1}^{S} P_{n,k} \quad (14)$$

and the variance is given by:

$$\sigma_{\varphi_n}^2 = \left\langle (\varphi_n - \langle \varphi_n \rangle)^2 \right\rangle = \left[\sum_{k=1}^{S} (I_{n,k} - P_{n,k}) \right]^2 = \sum_{k=1}^{S} P_{n,k}(1 - P_{n,k}) \quad (15)$$

This result is based on the assumption that the fluctuations of $I_{n,k}$ do not depend on the species label k, a consequence of the absence of interactions between species. The RSA relationship we seek to derive is the dependence of $\langle \varphi_n \rangle$ on n. Another quantity of interest is the average number of species observed in the community:

$$\langle S_{\text{obs}} \rangle = S - \langle \varphi_0 \rangle = S - \sum_{k=1}^{S} (1 - x)^{\Upsilon_k} \quad (16)$$

For a sample of J individuals comprising $\bar{S}$ species, the sampling multivariate probability distribution is a compound multinomial Dirichlet distribution[19,24]:

$$P(n_1, n_2, \ldots, n_{\bar{S}}|J) = \frac{1}{P(J)} \prod_{k=1}^{\bar{S}} P_{n_k,k} \delta(J - n_1 - n_2 - \ldots - n_{\bar{S}}) =$$

$$\binom{J + \sum_k \Upsilon_k - 1}{J}^{-1} \prod_{k=1}^{\bar{S}} \binom{n_k + \Upsilon_k - 1}{n_k} \delta(J - n_1 - n_2 - \ldots - n_{\bar{S}}) \quad (17)$$

where the normalization factor $P(J)$ is the probability of observing a community of size J. Interestingly, the same formula also applies to community dynamics with a fixed total population obeying a zero-sum rule—that is, when the death of an individual is immediately followed by a birth or by the addition of an immigrant into the community. Refs 9 and 13 have derived the expression for the compound multinomial Dirichlet distribution[19,24] by using a master equation approach and have presented a sampling theory that considers both ecological and evolutionary times in a single sampling formula.

---

that the local RSA resembles the Fisher log-series distribution. Our theory explains how the RSA becomes log-normal-like on aggregating the local communities into one metacommunity.

We can estimate the species-similarity index between two local communities can be defined as:

$$\frac{2S_{A \cap B}}{S_A + S_B} 100\% \quad (3)$$

where S_A and S_B are the number of observed species in communities A and B, respectively, and $S_{A \cap B}$ is the number of species that are present in both communities. Because immigration is almost absent for local communities, one can assume that for each local community the species are randomly drawn from the metacommunity species pool. With this assumption, the average similarity index becomes:

$$\frac{2S_A S_B}{S(S_A + S_B)} 100\% \quad (4)$$

where S is the total number of species in the metacommunity. Figure 2 shows the frequency distribution of the similarity index taken for all pairs of local communities (within a given region). The size of the species pool was chosen to be the same as the number of species in the regional metacommunity. Our results are in qualitative accord with those in ref. 7, which reported that the distribution of the similarity

index among local coral reefs was quite different from the predictions of neutral theory valid for tropical forests. The average community similarity is low (about 35%) and has a large variance. A more refined model of the coral reefs would not only consider the biology of coral-reef reproduction but also the specific locations, environments and the degree of stationarity of local communities.

Tropical forests

We now consider the metacommunity model introduced in ref. 1, in which a local community is embedded within a surrounding metacommunity, which is a source of immigrants. The dynamics of the local community is governed by births, deaths and immigration, whereas the metacommunity is characterized by births, deaths and speciation. The characteristic rate of species turnover in the local community is much faster than in the metacommunity so that we may treat the metacommunity as a fixed backdrop for immigration processes. We consider the non-interacting species case; this means that we have to consider only one species at a time, greatly simplifying the theory (see Box 1).

For the metacommunity, we can introduce speciation with a rate $v \ll 1$, a very small probability for the creation of a new species. The v term does not contribute significantly when $n \neq 0$, but has a crucial role when $n = 0$. (Under the equivalence assumption of neutral theory, the species label of the new species is of no consequence.)
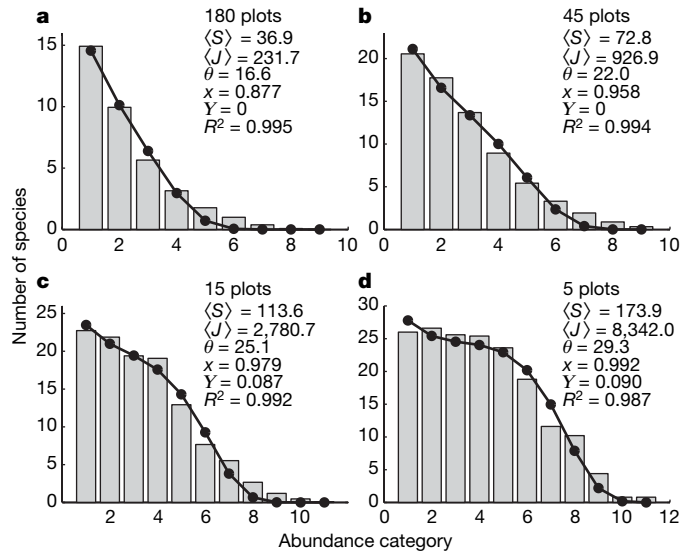
Figure 1 | Relative species abundance of coral-reef communities. We plot the fits of equation (1) (solid line) to the coral-reef species abundance data for the coral-reef local community (**a**), reef (**b**), metacommunity (**c**) and metacommunity (habitats pooled, **d**)[6,7]. The bars are observed numbers of species binned into $\log_2$ abundance categories. The first histogram bar represents $\langle\phi_1\rangle$, the second bar $\langle\phi_2\rangle + \langle\phi_3\rangle$, the third bar $\langle\phi_4\rangle + \langle\phi_5\rangle + \langle\phi_6\rangle + \langle\phi_7\rangle$, and so on. Graphs show the average RSA of: **a**, 180 local communities; **b**, 45 reef communities, each of which consists of 4 local communities; **c**, 15 metacommunities, each of which consists of 3 reef communities; **d**, metacommunities (habitats pooled), each of which consists of 5 metacommunities. Also shown are the values of the fitted parameters and the R^2 values.

Operationally, this is mathematically equivalent to replacing Y_k (for the kth species) in equation (12) by v, which yields the logarithmic distribution $P_n = vx^n/n + O(v^2)$, where $O(v^2)$ is a small term of the order of v^2.

Following ref. 2, one obtains an expression for the steady-state RSA distribution in the local community by setting $Y_k = \tilde{m}p_k$, where $\tilde{m}$ is a measure of the immigration rate measured in units of the birth rate b, and p_k is the fraction of individuals in the surrounding metacommunity belonging to the kth species (see Box 2):

$$\langle\varphi_n\rangle = \theta\frac{x^n}{n!}\int_0^\infty \frac{\Gamma(y+n)}{\Gamma(y+1)}e^{-\omega y}dy \tag{5}$$

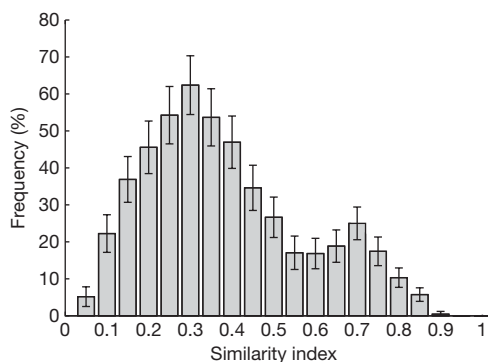where $\omega = \theta/\tilde{m} - \ln(1-x)$ and θ is the biodiversity parameter.

Figure 2 | Similarity index for coral-reef communities. We generated 10,000 random realizations of 60 samples with sizes equal to those of the coral-reef slopes[7] assuming that all the species have equal abundance in the metacommunity. The histogram shows the distribution of the similarity index predicted by the island model for the coral-reef data (the mean index is equal to 0.38 and the variance is 0.19) using equation (3). The distribution is qualitatively similar to that reported in ref. 7. Error bars represent standard deviation.

Box 2 | Relative species abundance of tropical forests

The mean number of species with n individuals in a community can be written as:

$$\langle\varphi_n\rangle = \sum_{k=1}^{S_M} P_{n,k} = S_M \int_0^\infty d\mu\,\hat{\rho}(\mu)P_{n,\mu} \tag{18}$$

where S_M is the number of species in the metacommunity. Here $\hat{\rho}(\mu)d\mu$ is the continuous probability distribution of the mean populations of the species in the metacommunity and has the form of the familiar Fisher log-series (in a singularity-free description[25,26]):

$$\hat{\rho}(\mu)d\mu = \frac{1}{\Gamma(\varepsilon)\delta^\varepsilon}\exp(-\mu/\rho)\mu^{\varepsilon-1}d\mu \tag{19}$$

where $\delta = x/(1-x)$ to match the first moment of the discrete and continuous distributions and:

$$P_{n,\mu} = \frac{(1-x)^{\tilde{m}\mu/J_M}}{\Gamma(\tilde{m}\mu/J_M)}\frac{x^n}{n!}\Gamma(n+\tilde{m}\mu/J_M) \tag{20}$$

where J_M is the total population of the metacommunity ($J_M = \sum_{n=1}^\infty \theta x^n = \theta x/(1-x)$). Then, on substituting equations (19) and (20) into equation (18) and defining $y = \tilde{m}\mu/J_M$, $\varepsilon\to0$ and $\varepsilon S_M\to\theta$, one obtains an analytical expression for the RSA of the local community:

$$\langle\varphi_n\rangle = \theta\frac{x^n}{n!}\int_0^\infty dy\,\frac{\Gamma(n+y)}{\Gamma(1+y)}\exp\{-y[J_M/(\tilde{m}\delta) - \ln(1-x)]\} =$$
$$\theta\frac{x^n}{n!}\int_0^\infty \frac{\Gamma(n+y)}{\Gamma(1+y)}e^{-\omega y}dy \equiv \theta\frac{x^n}{n!}f(n,\omega) \tag{21}$$

where $\omega = \theta/\tilde{m} - \ln(1-x)$ and the integral $f(n, \omega)$ in the above equation can be calculated analytically using the following recurrence equations: $f(1, \omega) = 1/\omega$, $f(n+1, \omega) = nf(n, \omega) - \partial f(n, \omega)/\partial\omega$. Equation (21) represents the average number of species with abundance n as a negative-binomial sampling from a metacommunity characterized by a log-series RSA[10,13,27–29].

For the limiting case of no immigration ($\tilde{m}\to0$), the RSA of the local community approaches the Fisher log-series. For large immigration, for tropical forests with the Fisher log-series metacommunity ($\tilde{m}\to\infty$), the probability distribution for each species is Poisson distributed with averages equal to those in the metacommunity so that one again obtains the Fisher log-series.

The expression equation (21) is much simpler than that presented in ref. 3:

$$\langle\varphi_n\rangle = \theta\frac{J!}{n!(J-n)!}\frac{\Gamma(\gamma)}{\Gamma(J+\gamma)} \times$$
$$\int_0^\gamma \frac{\Gamma(n+y)}{\Gamma(1+y)}\frac{\Gamma(J-n+\gamma-y)}{\Gamma(\gamma-y)}\exp(-y\theta/\gamma)dy \tag{22}$$

where $\gamma = m(J-1)/(1-m)$. The immigration parameter $\tilde{m}$ in equation (21) is equal to $Jm/(1-m)$ in ref. 3.

Using equation (5) we can readily obtain expressions for the average local community size $\langle J\rangle$ and the average number of species S_{obs} in the local community:

$$\langle J\rangle = \sum_{n=1}^\infty n\langle\varphi_n\rangle = \frac{\tilde{m}x}{1-x} \tag{6}$$

and

$$\langle S_{obs}\rangle = \sum_{n=1}^\infty \langle\varphi_n\rangle = \theta\ln\left[1 - \frac{\tilde{m}}{\theta}\ln(1-x)\right] \tag{7}$$

Table 1 summarizes the results of fits of data from six large species-rich plant communities using equation (5) (see Fig. 3). These data have been analysed previously using different varieties of neutral

Table 1 | Relative species abundance of tropical forests

Plot	S	J	θ	$\tilde{m}$	m	x	r^2
BCI, Panama	225	21,457	48	2,122	0.09	0.91	0.97
Yasuni, Ecuador	821	17,546	212	9,448	0.35	0.65	0.98
Pasoh, Malaysia	678	26,554	206	1,999	0.07	0.93	0.98
Korup, Cameroon	308	24,591	54	18,551	0.43	0.57	0.94
Lambir, Malaysia	1,004	33,175	305	3,281	0.09	0.91	0.99
Sinharaja, Sri Lanka	167	16,936	28	15,633	0.48	0.52	0.94

Estimates of the ecological parameters obtained from fitting equation (5) to the RSA data of six forests. S, the number of observed species; J, the local community size; θ, the biodiversity parameter[2]; m, the per capita immigration rate[2] ($\tilde{m} = Jm/(1 - m)$); and x, the per capita birth-to-death-rate ratio. Substituting the above expression for $\tilde{m}$ into equation (6), one obtains $m + x = 1$ (for the metacommunity $x + v = 1$).

theory[1−3,10]. Our simple analytical approximations of the exact sampling theory[8−14] yield virtually indistinguishable fits to the data, and we obtain very similar values of the biological parameters to the ones derived previously. The exact theory applies both under the assumption of species independence in a fluctuating community and for species undergoing zero-sum dynamics[9].

Comparison between coral reefs and tropical forests

To understand the qualitative difference between species composition in the coral-reef system and in tropical forests, we used the available relative tree species abundance data from the forest dynamics plot on Barro Colorado Island (BCI), Panama to generate random samples of the same size as in the coral-reef studies. Figure 4 shows the RSA for samples with 232, 927, 2,781 and 8,342 individuals (each histogram represents an RSA averaged over 100 samples). The abundance histograms are qualitatively similar to those for the coral-reef data. On fitting them with equation (5), the immigration parameter is very large for the small samples (for example, the per capita immigration rate $m \approx 1$ for the sample of 232 individuals) and decreases with an increase in sample size. This can be explained as follows: for a small population, the number of immigrants exceeds the number of internal birth events. As the population size increases, immigration has a less important role and becomes negligible as the population goes to infinity (and thus forms a metacommunity). Thus, one would observe a Fisher log-series in two limiting cases: in the metacommunity in which there are no immigration events, and in the very small local community that has a high immigration rate from the metacommunity characterized by a Fisher log-series RSA. The sampled community of the coral reef represents a very small portion of the total reef population. However, the distinct similarity indices for the tropical forest and the coral-reef communities underscore a key difference in the two
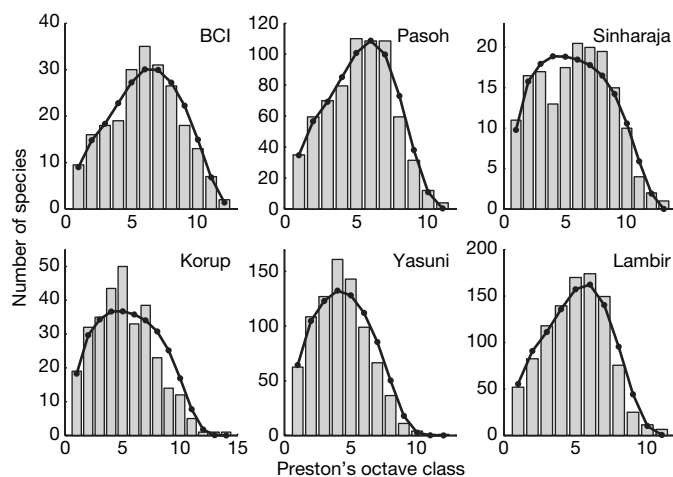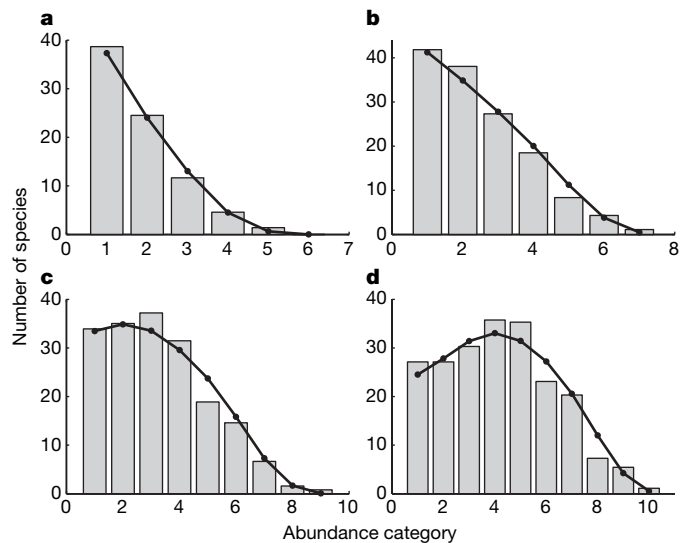
Figure 4 | Sampling of BCI tropical-forest data. We plot the fits of equation (5) (solid line) to the RSA of samples chosen randomly from the BCI data. Each plot shows RSA for the same number of samples with the same abundances as in the coral-reef data. The RSAs are qualitatively similar to those in Fig. 1 (the values of m are equal to 1, 0.84, 0.46 and 0.19 for **a**, **b**, **c** and **d**, respectively).

cases. For tropical forests, the assumption of equation (4)—that every species is equally likely to occur in a given local community—leads to a similarity index distribution that is significantly different from the actual distribution given by equation (3) (see Fig. 5). In contrast, for the coral-reef community, the observed distribution is consistent with the prediction of equation (4) (see Fig. 2). These results suggest a high degree of isolation between local communities in the coral-reef ecosystem. Unlike in the tropical forest, the set of local communities in the coral-reef data are effectively uncorrelated samples from the metacommunity[8,9].

Intraspecific interactions can be taken into account in a simplified manner by attributing density-dependent effective birth and death rates, which depend on the population of a species[3]. On retaining the equivalence of all species and within a parsimonious model, it was shown that the RSA behaves as $\langle \varphi_n \rangle \propto x^n/(n + c)$, where a positive coefficient c describes a rare-species advantage[3]. The data presented in Figs 1 and 4 can also be fitted approximately with this model with effective values of c equal to 2.29, 1.26, 2.63, 0.87, 2.91 and 0.41 for tropical forests (BCI, Yasuni, Pasoh, Korup, Lambir and Sinharaja, respectively) and −0.24, 0.05, 0.30 and 0.42 for coral-reef communities (local, reef, metacommunity and metacommunity (habitats
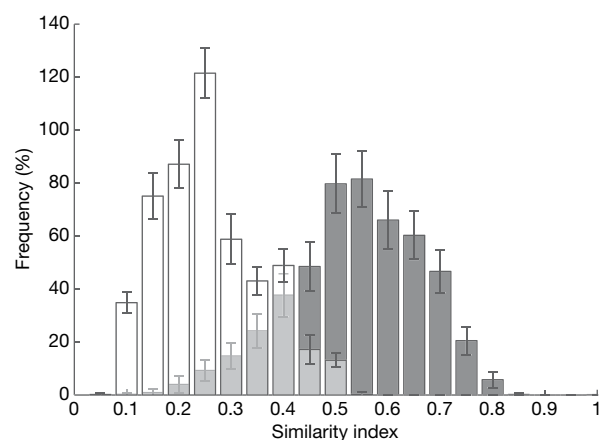
Figure 3 | Relative species abundance of tropical forests. We plot the fits of equation (5) (solid lines) to the tree species abundance data from the BCI, Pasoh, Sinharaja, Korup, Yasuni and Lambir plots (see Table 1). The frequency distributions are plotted using Preston's binning method[1]. The numbers on the x-axis represent Preston's octave classes.

Figure 5 | Similarity index for the BCI plot. We generated 10,000 random realizations of 60 samples with abundances equal to those of the coral reef slopes[7]. The dark and light histograms were calculated using equations (3) and (4), respectively. Error bars represent standard deviation.

pooled), respectively). Interestingly, the coral-reef local community is characterized by a negative *c* coefficient, which would correspond to a rare species disadvantage, analogous to the Allee effect[15], in the effective birth-to-death-rates ratio.

Discussion

We have presented a simple unified theoretical framework in which interspecific interactions are turned off, which yields analytical expressions for two distinct and very different types of metacommunities. One type[1], a continuous source area from which immigrants to local communities are drawn, yields log-series RSA distributions on large scales and log-normal-like distributions with interior modes on local scales, as one consistently observes in tropical-forest communities. The other type[8,9], a metacommunity consisting of an island archipelago of isolated communities, yields the opposite pattern: log-series-like RSA distributions in local communities, but interior-mode RSA distributions on large scales, as is observed in coral-reef communities.

In adopting this approach, we are not denying that species interact. There is a huge ecological literature on interspecific interactions, including competitive and predator–prey interactions, among others. What we are saying, however, is that we can make major theoretical progress in ecology by not considering species interactions at the outset. It is plausible that there is an underlying simplicity associated with an ecological community in the vicinity of its steady state. Although the effects of interactions could very well be a factor under non-equilibrium conditions, overt competition between a pair of species would be toned down in the steady state in favour of one or the other over local spatial and temporal scales. Indeed, what our theory shows is that a large fraction of the quantitative variation in patterns of RSA in tropical forests and coral reefs can be accounted for by our non-interacting species approach. Additional variation in patterns of RSA will undoubtedly be explained when theoretical ecologists incorporate species interactions, especially when we require the theory to be able to explain the abundances of named and identified species in ecological communities. This progression in the development of ecological theory has an analogy to the development of the theory of gases in physics[16]. First came the assumption that gases were 'ideal' and non-interacting, which then led to the ideal gas equation of state[17]. This does remarkably well, but deviations in the behaviour of gases from ideal behaviour led to refinements to the theory[18]. We look forward to analogous developments in theoretical ecology.

METHODS SUMMARY

When confronting the theory with data, it is necessary to relax the zero-sum rule. For example, in the case of tropical forests, the sampling process entails counting all the trees within a fixed area, but there is always some variation in the total abundance of trees in a fixed sample area. This is not a problem at the stationary state, at which one can prove that the zero-sum rule can be relaxed[9,19] and the species can be treated as being independent of each other (see also Supplementary Material in ref. 3). Indeed, we have found that a maximum likelihood estimator (MLE) method[19] based on equation (17) on a coral-reef metacommunity system leads to a similar value of the parameter Υ compared to the estimate based on the fitting of the average RSA with equation (5), $\Upsilon_{MLE} = 0.058$ and $\Upsilon_{RSA} = 0.087$. We have checked the validity of the relaxation of the zero-sum rule by randomly generating 1,000 samples and calculating the exact likelihood function for them. As expected, the coral-reef data are indistinguishable from the random pseudo-samples. For the fitting procedure, we assumed that the census data on species composition represents the average RSA. In all figures, we estimated the model parameters by a least-squares fit, minimizing the sum of squared residuals[20].

1. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
2. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037 (2003).
3. Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. Density and frequency dependence explains tree species abundance and diversity in tropical forests. *Nature* 438, 658–661 (2005).
4. Latimer, A. M., Silander, J. A. Jr & Cowling, R. M. Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science* 309, 1722–1725 (2005).
5. Etienne, R. S., Latimer, A. M., Silander, J. A. Jr & Cowling, R. M. Comment on ''Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot''. *Science* 311, 610 (2006).
6. Connolly, S. R., Hughes, T. P., Bellwood, D. R. & Karlson, R. H. Community structure of corals and reef fishes at multiple scales. *Science* 309, 1363–1365 (2005).
7. Dornelas, M., Connolly, S. R. & Hughes, T. P. Coral reef diversity refutes the neutral theory of biodiversity. *Nature* 440, 80–82 (2006).
8. Alonso, D. & Pascual, M. Comment on ''A keystone mutualism drives pattern in a power function''. *Science* 313, 1739 (2006).
9. Etienne, R. S., Alonso, D. & McKane, A. J. The zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* 248, 522–536 (2007).
10. Alonso, D. & McKane, A. J. Sampling Hubbell's neutral theory of biodiversity. *Ecol. Lett.* 7, 901–910 (2004).
11. Etienne, R. S. & Olff, H. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7, 170–175 (2004).
12. Etienne, R. S. A new sampling formula for neutral biodiversity. *Ecol. Lett.* 8, 253–260 (2005).
13. Etienne, R. S. & Alonso, D. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* 8, 1147–1156 (2005); erratum 9, 500 (2006).
14. Etienne, R. S. & Alonso, D. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J. Stat. Phys.* 128, 485–510 (2006).
15. Courchamp, F., Clutton-Brock, T. & Grenfell, B. Inverse density dependence and the Allee effect. *Trends Ecol. Evol.* 14, 405–410 (1999).
16. Harte, J. Tail of death and resurrection. *Nature* 424, 1006–1007 (2003).
17. Maxwell, J. C. *The Scientific Papers of James Clerk Maxwell* Vol. 1 (Dover, New York, 2003).
18. van der Waals, J. D. *On the Continuity of the Gaseous and Liquid States* (Dover, New York, 2004).
19. Rannala, B. The sampling theory of neutral alleles in an island population of fluctuating size. *Theor. Popul. Biol.* 50, 91–104 (1996).
20. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1993).
21. Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 1 (Wiley & Sons, Hoboken, 1968).
22. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2001).
23. Kendall, D. G. Stochastic processes and population growth. *J. Roy. Statist. Soc. B* 11, 230–282 (1949).
24. Mosimann, J. E. On the compound multinomial distribution, the multivariate distribution, and correlations among proportions. *Biometrika* 49, 65–82 (1962).
25. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12, 42–58 (1943).
26. Rao, C. R. *Statistical Ecology* Vol. 1, *Spatial Patterns and Statistical Distributions* 131–142 (Penn. State Univ. Press, University Park, Pennsylvania, 1971).
27. Pielou, E. C. *An Introduction to Mathematical Ecology* (Wiley, New York, 1969).
28. Bulmer, M. G. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* 30, 101–110 (1974).
29. Dewdney, A. K. A general theory of the sampling process with applications to the veil line. *Theor. Popul. Biol.* 54, 294–302 (1998).