

Neutral theory and relative species abundance in ecology

Igor Volkov¹, Jayanth R. Banavar¹, Stephen P. Hubbell^{2,3} & Amos Maritan^{4,5}

¹Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

²Department of Plant Biology, The University of Georgia, Athens, Georgia 30602, USA

³The Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama

⁴International School for Advanced Studies (SISSA), Via Beirut 2/4, 34014 Trieste, Italy

⁵INFN and The Abdus Salam International Center for Theoretical Physics, Trieste, Italy

The theory of island biogeography¹ asserts that an island or a local community approaches an equilibrium species richness as a result of the interplay between the immigration of species from the much larger metacommunity source area and local extinction of species on the island (local community). Hubbell² generalized this neutral theory to explore the expected steady-state distribution of relative species abundance (RSA) in the local community under restricted immigration. Here we present a theoretical framework for the unified neutral theory of biodiversity² and an analytical solution for the distribution of the RSA both in the metacommunity (Fisher's log series) and in the local community, where there are fewer rare species. Rare species are more extinction-prone, and once they go locally extinct, they take longer to re-immigrate than do common species. Contrary to recent assertions³, we show that the analytical solution provides a better fit, with fewer free parameters, to the RSA distribution of tree species on Barro Colorado Island, Panama⁴, than the lognormal distribution^{5,6}.

The neutral theory in ecology^{2,7} seeks to capture the influence of speciation, extinction, dispersal and ecological drift on the RSA under the assumption that all species are demographically alike on a per capita basis. This assumption, while only an approximation⁸⁻¹⁰, appears to provide a useful description of an ecological community on some spatial and temporal scales^{2,7}. More significantly, it allows the development of a tractable null theory for testing hypotheses about community assembly rules. However, until now, there has been no analytical derivation of the expected equilibrium distribution of RSA in the local community, and fits to the theory have required simulations² with associated problems of convergence times, unspecified stopping rules, and precision³.

The dynamics of the population of a given species is governed by generalized birth and death events (including speciation, immigration and emigration). Let $b_{n,k}$ and $d_{n,k}$ represent the probabilities of birth and death, respectively, in the k th species with n individuals with $b_{-1,k} = d_{0,k} = 0$. Let $p_{n,k}(t)$ denote the probability that the k th species contains n individuals at time t . In the simplest scenario, the time evolution of $p_{n,k}(t)$ is regulated by the master equation¹¹⁻¹³

$$\frac{dp_{n,k}(t)}{dt} = p_{n+1,k}(t)d_{n+1,k} + p_{n-1,k}(t)b_{n-1,k} - p_{n,k}(t)(b_{n,k} + d_{n,k}) \quad (1)$$

which leads to the steady-state or equilibrium solution, denoted by P :

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (2)$$

for $n > 0$ and where $P_{0,k}$ can be deduced from the normalization condition $\sum_n P_{n,k} = 1$. Note that there is no requirement of

Box 1

Derivation of the RSA of the local community

We study the dynamics within a local community following the mathematical framework of McKane *et al.*²⁷, who studied a mean-field stochastic model for species-rich communities. In our context, the dynamical rules² governing the stochastic processes in the community are:

(1) With probability $1-m$, pick two individuals at random from the local community. If they belong to the same species, no action is taken. Otherwise, with equal probability, replace one of the individuals with the offspring of the other. In other words, the two individuals serve as candidates for death and parenthood.

(2) With probability m , pick one individual at random from the local community. Replace it by a new individual chosen with a probability proportional to the abundance of its species in the metacommunity. This corresponds to the death of the chosen individual in the local community followed by the arrival of an immigrant from the metacommunity. Note that the sole mechanism for replenishing species in the local community is immigration from the metacommunity, which for the purposes of local community dynamics is treated as a permanent source pool of species, as in the theory of island biogeography¹.

These rules are encapsulated in the following expressions for effective birth and death rates for the k th species:

$$b_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \frac{\mu_k}{J_M} \left(1 - \frac{n}{J}\right) \quad (8)$$

$$d_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \left(1 - \frac{\mu_k}{J_M}\right) \frac{n}{J} \quad (9)$$

where μ_k is the abundance of the k th species in the metacommunity and J_M is the total population of the metacommunity.

The right hand side of equation (8) consists of two terms. The first corresponds to rule (1) with a birth in the k th species accompanied by a death elsewhere in the local community. The second term accounts for an increase of the population of the k th species due to immigration from the metacommunity. The immigration is, of course, proportional to the relative abundance μ_k/J_M of the k th species in the metacommunity. Equation (9) follows in a similar manner. Note that $b_{n,k}$ and $d_{n,k}$ not only depend on the species label k but also are no longer simply proportional to n .

Substituting equation (8) and (9) into equation (2), one obtains the expression²⁷:

$$P_{n,k} = \frac{J!}{n!(J-n)!} \frac{\Gamma(n+\lambda_k)}{\Gamma(\lambda_k)} \frac{\Gamma(\vartheta_k-n)}{\Gamma(\vartheta_k-J)} \frac{\Gamma(\lambda_k+\vartheta_k-J)}{\Gamma(\lambda_k+\vartheta_k)} \equiv F(\mu_k) \quad (10)$$

where

$$\lambda_k = \frac{m}{(1-m)} (J-1) \frac{\mu_k}{J_M} \quad (11)$$

and

$$\vartheta_k = J + \frac{m}{(1-m)} (J-1) \left(1 - \frac{\mu_k}{J_M}\right) \quad (12)$$

Note that the k dependence in equation (10) enters only through μ_k . On substituting equation (10) into equation (4), one obtains:

$$\langle \phi_n \rangle = \sum_{k=1}^{S_M} F(\mu_k) = S_M \langle F(\mu_k) \rangle = S_M \int d\mu \hat{\rho}(\mu) F(\mu) \quad (13)$$

Here $\hat{\rho}(\mu)d\mu$ is the 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^{15,28}):

$$\hat{\rho}(\mu)d\mu = \frac{1}{\Gamma(\varepsilon)\delta^\varepsilon} \exp(-\mu/\delta) \mu^{\varepsilon-1} d\mu \quad (14)$$

where $\delta = x/(1-x)$. Substituting equation (14) into the integral in equation (13), taking the limits $S_M \rightarrow \infty$ and $\varepsilon \rightarrow 0$ with $\theta = S_M \varepsilon$ approaching a finite value^{15,28} and on defining $y = \mu \frac{\theta}{\delta \theta}$, one can obtain our central result, equation (7).

conservation of community size. One can show that the system is guaranteed to reach the stationary solution (2) in the infinite time limit¹⁴.

The frequency of species containing n individuals is given by:

$$\phi_n = \sum_{k=1}^S I_k \quad (3)$$

where S is the total number of species and the indicator I_k is a random variable which 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:

$$\langle \phi_n \rangle = \sum_{k=1}^S P_{n,k} \quad (4)$$

The RSA relationship we seek to derive is the dependence of $\langle \phi_n \rangle$ on n .

Let a community consist of species with $b_{n,k} \equiv b_n$ and $d_{n,k} \equiv d_n$ being independent of k (the species are assumed to be demographically identical).

From equation (4), it follows that $\langle \phi_n \rangle$ is simply proportional to P_n , leading to:

$$\langle \phi_n \rangle = S P_0 \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}} \quad (5)$$

We consider a metacommunity in which the probability d that an individual dies and the probability b that an individual gives birth to an offspring are independent of the population of the species to which it belongs (density-independent case), that is, $b_n = bn$ and $d_n = dn$ ($n > 0$). Speciation may be introduced by ascribing a non-zero probability of the appearance of an individual of a new species, that is, $b_0 = v \neq 0$. Substituting the expressions into equation (5),

one obtains the celebrated Fisher log series¹⁵:

$$\langle \phi_n^M \rangle = S_M P_0 \frac{b_0 b_1 \dots b_{n-1}}{d_1 d_2 \dots d_n} = \theta \frac{x^n}{n} \quad (6)$$

where M refers to the metacommunity, $x = b/d$ and $\theta = S_M P_0 v/b$ is the biodiversity parameter (also called Fisher's α). We follow the notation of Hubbell² in this paper. Note that x represents the ratio of effective per capita birth rate to the death rate arising from a variety of causes such as birth, death, immigration and emigration. Note that in the absence of speciation, $b_0 = v = \theta = 0$, and, in equilibrium, there are no individuals in the metacommunity. When one introduces speciation, x has to be less than 1 to maintain a finite metacommunity size $J_M = \sum_n n \langle \phi_n \rangle = \theta x / (1 - x)$.

We turn now to the case of a local community of size J undergoing births and deaths accompanied by a steady immigration of individuals from the surrounding metacommunity. When the local community is semi-isolated from the metacommunity, one may introduce an immigration rate m , which is the probability of immigration from the metacommunity to the local community. For constant m (independent of species), immigrants belonging to the more abundant species in the metacommunity will arrive in the local community more frequently than those of rarer species.

Our central result (see Box 1 for a derivation) is an analytic expression for the RSA of the local community:

$$\langle \phi_n \rangle = \theta \frac{J!}{n!(J-n)!} \frac{\Gamma(\gamma)}{\Gamma(J+\gamma)} \int_0^\gamma \frac{\Gamma(n+y)}{\Gamma(1+y)} \frac{\Gamma(J-n+\gamma-y)}{\Gamma(\gamma-y)} \exp(-y\theta/\gamma) dy \quad (7)$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ which is equal to $(z-1)!$ for integer z and $\gamma = \frac{m(J-1)}{1-m}$. As expected, $\langle \phi_n \rangle$ is zero when n exceeds J . The computer calculations in Hubbell's book² as well as those more recently carried out by McGill³ were aimed at estimating $\langle \phi_n \rangle$ by simulating the processes of birth, death and immigration.

One can evaluate the integral in equation (7) numerically for a given set of parameters: J , θ and m . For large values of n , the integral can be evaluated very accurately and efficiently using the method of steepest descent¹⁶. Any given RSA data set contains information about the local community size, J , and the total number of species in the local community, $S_L = \sum_{k=1}^J \langle \phi_k \rangle$. Thus there is just one free fitting parameter at one's disposal.

McGill asserted³ that the lognormal distribution is a more parsimonious null hypothesis than the neutral theory, a suggestion which is not borne out by our reanalysis of the Barro Colorado Island (BCI) data. We focus only on the BCI data set because, as pointed out by McGill³, the North American Breeding Bird Survey data are not as exhaustively sampled as the BCI data set, resulting in fewer individuals and species in any given year in a given location. Furthermore, the McGill analysis seems to rely on adding the bird counts over five years at the same sampling locations even though these data sets are not independent.

Figure 1 shows a Preston-like binning⁵ of the BCI data⁴ and the fit of our analytic expression with one free parameter (11 degrees of freedom) along with a lognormal having three free parameters (9 degrees of freedom). Standard chi-square analysis¹⁷ yields values of $\chi^2 = 3.20$ for the neutral theory and 3.89 for the lognormal. The probabilities of such good agreement arising by chance are 1.23% and 8.14% for the neutral theory and lognormal fits, respectively. Thus one obtains a better fit of the data with the analytical solution to the neutral theory to BCI than with the lognormal, even though there are two fewer free parameters. McGill's analysis³ on the BCI data set was based on computer simulations in which there were difficulties in knowing when to stop the simulations, that is, when equilibrium had been reached. It is unclear whether McGill averaged over an ensemble of runs, which is essential to obtain repeatable and reliable results from simulations of stochastic

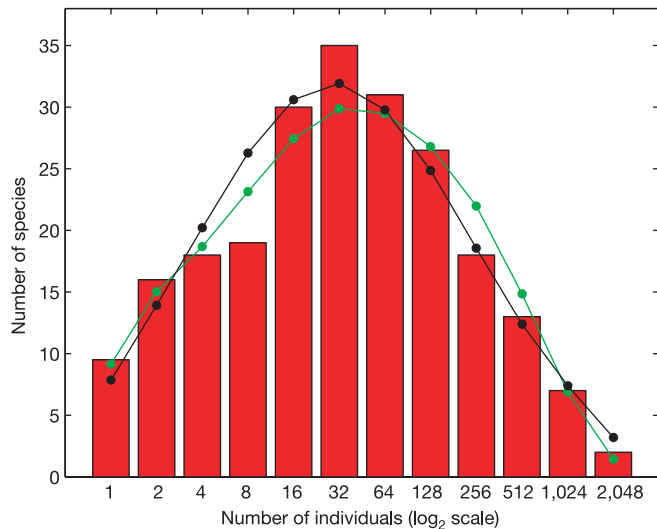


Figure 1 Data on tree species abundances in 50-hectare plot of tropical forest in Barro Colorado Island, Panama⁴. The total number of trees >10 cm DBH in the data set is 21,457 and the number of distinct species is 225. The red bars are observed numbers of species binned into log₂ abundance categories, following Preston's method⁵. The first histogram bar represents $\langle \phi_1 \rangle/2$, the second bar $\langle \phi_1 \rangle/2 + \langle \phi_2 \rangle/2$, the third bar $\langle \phi_2 \rangle/2 + \langle \phi_3 \rangle + \langle \phi_4 \rangle/2$ the fourth bar $\langle \phi_4 \rangle/2 + \langle \phi_5 \rangle + \langle \phi_6 \rangle + \langle \phi_7 \rangle + \langle \phi_8 \rangle/2$ and so on. The black curve shows the best fit to a lognormal distribution $\langle \phi_n \rangle = \frac{N}{n} \exp(-(\log_2 n - \log_2 n_0)^2 / 2\sigma^2)$ ($N = 46.29$, $n_0 = 20.82$ and $\sigma = 2.98$), while the green curve is the best fit to our analytic expression equation (7) ($m = 0.1$ from which one obtains $\theta = 47.226$ compared to the Hubbell² estimates of 0.1 and 50 respectively and McGill's best fits³ of 0.079 and 48.5 respectively).

processes because of their inherent noisiness. However, simulations of the neutral theory are no longer necessary, and all problems with simulations are moot, because an analytical solution is now available.

The lognormal distribution is biologically less informative and mathematically less acceptable as a dynamical null hypothesis for the distribution of RSA than the neutral theory. The parameters of the neutral theory or RSA are directly interpretable in terms of birth and death rates, immigration rates, size of the metacommunity, and speciation rates. A dynamical model of a community cannot yield a lognormal distribution with finite variance because in its time evolution, the variance increases through time without bound. However, as shown in ref. 18, the lognormal distribution can arise in static models, such as those based on niche hierarchy.

The steady-state deficit in the number of rare species compared to that expected under the log series can also occur because rare species grow differentially faster than common species and therefore move up and out of the rarest abundance categories owing to their rare-species advantage¹⁹. Indeed, it is likely that several different models (such as an empirical lognormal distribution, niche hierarchy models¹⁸ or the theory presented here) might provide comparable fits to the RSA data (we have found that the lognormal does slightly better than the neutral theory for the Pasoh data set²⁰, obtained in a tropical tree community in Malaysia). Such fitting exercises in and of themselves, however, do not constitute an adequate test of the underlying theory. Neutral theory predicts that the degree of skewing of the RSA distribution ought to increase as the rate of immigration into the local community decreases. Dynamic data on rates of birth, death, dispersal and immigration are needed to evaluate the assumptions of neutral theory and determine the role played by niche differentiation in the assembly of ecological communities.

Our analysis should also apply to the field of population genetics in which the mutation-extinction equilibrium of neutral allele frequencies at a given locus has been studied for several decades^{21–26}. □

Received 7 May; accepted 9 June 2003; doi:10.1038/nature01883.

1. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, NJ, 1967).
2. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
3. McGill, B. J. A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885 (2003).
4. Condit, R. et al. Beta-diversity in tropical forest trees. *Science* **295**, 666–669 (2002).
5. Preston, F. W. The commonness and rarity of species. *Ecology* **29**, 254–283 (1948).
6. May, R. M. *Ecology and Evolution of Communities* 81–120 (Harvard Univ. Press, Cambridge, MA, 1975).
7. Bell, G. Neutral macroecology. *Science* **293**, 2413–2418 (2001).
8. Diamond, J. & Case, T. J. (eds) *Community Ecology* (Harper and Row, New York, NY, 1986).
9. Tilman, D. *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, NJ, 1988).
10. Weiher, E. & Keddy, P. *Ecological Assembly Rules: Perspectives, Advances, and Retreats* (Cambridge Univ. Press, Cambridge, UK, 1999).
11. Boswell, M. T., Ord, J. K. & Patil, G. P. *Statistical Distributions in Ecological Work* 3–157 (International Co-operative Publishing, Fairland, MD, 1979).
12. Caraco, T. *Statistical Distributions in Ecological Work* 371–387 (International Co-operative Publishing, Fairland, MD, 1979).
13. Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 1 (John Wiley & Sons, Hoboken, NJ, 1968).
14. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (Amsterdam, North-Holland, 2001).
15. 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).
16. Morse, P. M. & Feshbach, H. *Methods of Theoretical Physics* Part 1 (McGraw-Hill, New York, NY, 1953).
17. 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, UK, 1993).
18. Sugihara, G., Bersier, L., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100**, 5246–5251 (2003).
19. Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: Theoretical consequences for patterns of diversity. *Am. Nat.* **159**, 1–23 (2002).
20. Manokaran, N. et al. *Stand Tables and Species Distributions in the Fifty Hectare Plot at Pasoh Forest Reserve* (Forest Research Institute Malaysia, Kuala Lumpur, 1992).
21. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
22. Ewens, W. J. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112 (1972).
23. Karlin, J. & McGregor, J. Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* **3**, 113–116 (1972).

24. Watterson, G. A. Models for the logarithmic species abundance distributions. *Theor. Pop. Biol.* **6**, 217–250 (1975).
25. Kimura, M. & Ohta, T. *Theoretical Aspects of Population Genetics* (Princeton Univ. Press, Princeton, NJ, 1971).
26. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, UK, 1983).
27. McKane, A., Alonso, D. & Solé, R. V. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E* **62**, 8466–8484 (2000).
28. Rao, C. R. *Statistical Ecology* Vol. 1 *Spatial Patterns and Statistical Distributions* 131–142 (The Penn. State Univ. Press, University Park, PA, 1971).

Acknowledgements We are grateful to O. Kargaltsev for a careful reading of the manuscript. This work was supported by NASA, by grants from the NSF, and by the Department of Plant Biology, University of Georgia.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.B. (banavar@psu.edu) or S.P.H. (shubbell@dogwood.botany.uga.edu).

The genome of a motile marine *Synechococcus*

B. Palenik¹, B. Brahamsha¹, F. W. Larimer^{2,3}, M. Land^{2,3}, L. Hauser^{2,3}, P. Chain^{3,4}, J. Lamerdin^{3,4}, W. Regala^{3,4}, E. E. Allen^{1*}, J. McCarren¹, I. Paulsen⁵, A. Dufresne⁶, F. Partensky⁶, E. A. Webb⁷ & J. Waterbury⁷

¹Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0202, USA
²Computational Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6480, USA
³Joint Genome Institute, Walnut Creek, California 94598, USA
⁴Lawrence Livermore National Laboratory, Livermore, California 94550-9234, USA
⁵TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA
⁶UMR 7127 CNRS Station Biologique de Roscoff, 29682 Roscoff, France
⁷Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

* Present address: Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

Marine unicellular cyanobacteria are responsible for an estimated 20–40% of chlorophyll biomass and carbon fixation in the oceans¹. Here we have sequenced and analysed the 2.4-megabase genome of *Synechococcus* sp. strain WH8102, revealing some of the ways that these organisms have adapted to their largely oligotrophic environment. WH8102 uses organic nitrogen and phosphorus sources and more sodium-dependent transporters than a model freshwater cyanobacterium. Furthermore, it seems to have adopted strategies for conserving limited iron stores by using nickel and cobalt in some enzymes, has reduced its regulatory machinery (consistent with the fact that the open ocean constitutes a far more constant and buffered environment than fresh water), and has evolved a unique type of swimming motility. The genome of WH8102 seems to have been greatly influenced by horizontal gene transfer, partially through phages. The genetic material contributed by horizontal gene transfer includes genes involved in the modification of the cell surface and in swimming motility. On the basis of its genome, WH8102 is more of a generalist than two related marine cyanobacteria².

Most species of picoplanktonic marine cyanobacteria currently known belong to two genera: *Synechococcus* and *Prochlorococcus*. Members must have the ability to acquire major nutrients and trace metals at the submicromolar concentrations found in the oligotrophic open seas. Their light-harvesting apparatus is uniquely