THEORETICAL BIOLOGY

Comparing models of species abundance

Arising from: I. Volkov, J. R. Banavar, F. He, S. P. Hubbell & A. Maritan *Nature* **438,** 658–661 (2005)

Ecologists are struggling to explain how so many tropical tree species can coexist in tropical forests, and several empirical studies have demonstrated that negative density dependence is an important mechanism of tree-species coexistence[1,2]. Volkov *et al.*[3] compare a model incorporating negative density dependence with a dispersal-limited neutral model[4] and claim that each predicts six empirical species-abundance distributions of tropical-tree communities equally well. However, we show here that their main conclusion is premature: when the two models are compared in an improved analysis, we find that the dispersal-limited model outcompetes the density-dependent model in all six cases. Hence, although density dependence is certainly an important diversity-maintaining mechanism, our improved approach indicates that the dispersal-limited model provides a more parsimonious explanation of empirical species-abundance distributions.

Volkov *et al.* compare model performances by applying a maximum-likelihood method to empirical species-abundance distributions. They use an approximate-likelihood function[5], but an exact likelihood is now available for the dispersal-limited model that is based on a proper sampling theory[6,7]. We applied the new maximum-likelihood method to the data sets used by Volkov *et al.* and based our model comparison on the difference in the models' AIC, defined as twice the number of parameters in the model minus twice its maximum log-likelihood[8].

We found that the dispersal-limited model performed significantly better than the density-dependent model for all six forests (Table 1). The probability that the density-dependent model performed better than the dispersal-limited model (mathematically defined as the Akaike weight[8]) was always less than 1%. Even a dispersal-unlimited neutral model performed better than the density-dependent model in two cases out of six. Thus the claim by Volkov *et al.* that density dependence gives an equally sufficient mechanistic explanation for species-abundance distributions, in addition to and independent of dispersal limitation, is unsubstantiated.

How can one test the claim of Volkov *et al.* that alternative models cannot be distinguished, in principle or in practice, given just the species-abundance distributions? Volkov *et al.* support their assertion by arguing that the ratio $\hat{r}_n$, computed from the expected number of species with a given abundance[3], looks the same for all models when plotted against *n*. However, they do not convincingly show that these graphs are

Table 1 | Comparison of species-abundance models

Site	S	J	Θ	m	L_1	L_2	L_3	Model ranking	w_2	w_3
BCI	225	21,457	47.7	0.093	−308.7	−315.0	−318.8	1 > 2 > 3	0.0018	0.0001
Yasuni	821	17,546	204.2	0.429	−297.2	−303.6	−307.6	1 > 2 > 3	0.0017	0.0001
Pasoh	678	26,554	190.9	0.093	−359.4	−365.3	−392.5	1 > 2 > 3	0.0027	0.0000
Korup	308	24,591	52.7	0.547	−317.0	−323.1	−318.7	1 = 3 > 2	0.0022	0.3318
Lambir	1,004	33,175	285.6	0.115	−386.4	−391.2	−437.9	1 > 2 > 3	0.0090	0.0000
Sinharaja	167	16,936	436.8	0.0019	−252.9	−258.5	−253.8	1 = 3 > 2	0.0037	0.2928

The performance of three species-abundance models are compared in six large tropical forest dynamic plots containing *J* sampled individuals and *S* species. Parameters Θ and *m* are the maximum-likelihood estimates in model 1. L_1, L_2 and L_3 represent the maximum log-likelihood values for the respective models. Model ranking was based on the Akaike weights, w_i, the probability that model *i* is more likely than the model with the lowest AIC[8].
Methods. Model 1 is the neutral model with dispersal limitation of ref. 4, model 2 is the density-dependence model of ref. 3, model 3 is like model 1 but with no dispersal limitation (ref. 4). In models 1 and 3, likelihoods are calculated based on ref. 6. In model 2, likelihoods are calculated based on ref. 5 (see ref. 3). The AIC is defined as 2($p − L_i$) where *p* is the number of parameters ($p = 2$ in models 1 and 2, and $p = 1$ in model 3). The Akaike weights were computed by $w_i = \exp(-\Delta_i/2)/[1 + \exp(-\Delta_i/2)]$, where Δ_i is the difference of AIC between model *i* and the model with the lowest AIC.

statistically indistinguishable. Even if they had, this would still be a weak test of model equivalence. Proving that two species-abundance models are mathematically equivalent requires construction of the multivariate distribution for species abundances under the density-dependent model, as well as evidence that this distribution is mathematically equal to that derived in ref. 6. If so, the two models would indeed be indistinguishable. If the models are not mathematically equivalent, they may still be statistically equivalent in most practical situations. Proving statistical equivalence requires a case-by-case comparison of the models' exact likelihood functions, as outlined here.

Our analysis provides evidence in favour of the dispersal-limited model, and contradicts the claim of Volkov *et al.*[3]. However, the authors' density-dependent model in its current formulation lacks a suitable sampling theory with an exact-likelihood function, and hence precludes any definitive answer about the fundamental or practical equivalence of alternative models. Although Volkov *et al.* have made a useful contribution by emphasizing the value of simple models in ecology, further theoretical developments aimed at exploring

the contribution of density dependence to patterns of species diversity are needed[9,10].

Jérôme Chave*, David Alonso†, Rampal S. Etienne‡
*Laboratoire Evolution et Diversité Biologique, UMR5174 CNRS, Université Paul Sabatier, 31062 Toulouse, France
e-mail : chave@cict.fr
†Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA
‡Community and Conservation Ecology Group, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

1. Harms, K. E., Wright, S. L., Calderón, O., Hernandez, A. & Herre, E. A. *Nature* **404,** 493–495 (2000).
2. Packer, A. & Clay, K. *Nature* **404,** 278–281 (2000).
3. Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. *Nature* **438,** 658–661 (2005).
4. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography.* (Princeton Univ. Press, Princeton, New Jersey, 2001).
5. Alonso, D. & McKane, A. J. *Ecol. Lett.* **7,** 901–910 (2004).
6. Etienne, R. S. *Ecol. Lett.* **8,** 253–260 (2005).
7. Etienne, R. S. & Alonso, D. *Ecol. Lett.* **8,** 1147–1156 (2005).
8. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference* 2nd edn (Springer, New York, 2002).
9. Engen, S. & Lande, R. *Math. Biosci.* **132,** 169–183 (1996).
10. Diserud, O. & Engen, S. *Am. Nat.* **155,** 497–511 (2000).

THEORETICAL BIOLOGY

Volkov et al. *reply*

Replying to: J. Chave, D. Alonso and R. S. Etienne *Nature* **441,** doi:10.1038/nature04826 (2005)

We have demonstrated that information on relative species abundance (RSA) cannot, without additional information, be used to discriminate among biological explanations for different RSA patterns[1]; but Chave *et al.* claim that our conclusion is premature[2]. Here

we show that their analysis was not carried out in a consistent manner and that density dependence gives an equally valid mechanistic explanation for RSA patterns in addition to, and independently of, dispersal limitation.
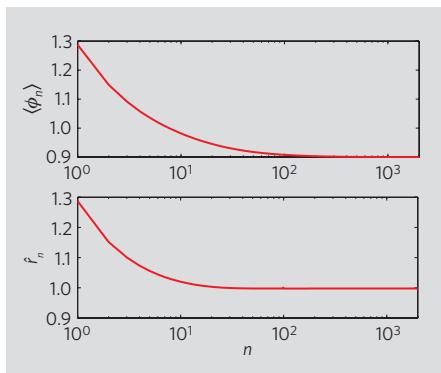
Consider a very simple model of immigra-

Figure 1 | **Species abundance.** Top, plot of $\langle \phi_n \rangle$, the average number of species with abundance n, against n for the model described in ref.2, using the biodiversity parameter $\Theta = 50$, the immigration parameter $m = 0.1$ and a population of 20,000. Bottom, plot of $\hat{r}_n$, the density-dependent birth rate–death rate ratio, against n, deduced from the RSA data.

tion, in which the birth and death rates for a species with n individuals is given by $b_n = b^* n + c^*$ and $d_n = d^* n$, respectively, or the equivalent per capita rates are $b_n/n = b^* + c^*/n$ and $d_n/n = d^*$, where b^* and d^* represent the intrinsic per capita birth and death rates, respectively, and c^* is a constant that captures the effects of immigration from a surrounding metacommunity that has a uniform distribution of species. This model can equally well be envisaged as arising not from immigration but from a density-dependent birth mechanism, so it is impossible to determine the biology

underlying the resulting RSA data without additional information.

As a second example, consider the dispersal limitation model studied by Volkov *et al.*[3] using the biodiversity parameter $\Theta = 50$ and the immigration parameter $m = 0.1$. Figure 1 shows the RSA data for this model obtained through the analytical expression derived in ref. 3, as well as a graph of the density-dependent birth–death rate ratio $\hat{r}_n$ plotted against n, deduced[1] from this RSA data ($\hat{r}_n = [(n+1)/n][b_n/d_{n+1}] = [(n+1)/n][\langle \phi_{n+1} \rangle/\langle \phi_n \rangle]$, where $\langle \phi_n \rangle$ represents the mean number of species with abundance n).

Assume now that we were given the RSA data (Fig. 1, top) and asked what we could learn from it. Clearly, the dispersal-limitation model from which the data were derived provides a perfect description of the data, so it might be tempting to conclude that the mechanism underlying the data is in fact dispersal limitation with the selected model parameters. However, a symmetric density-dependence mechanism that gives $\hat{r}_n$ versus n (Fig. 1, bottom) would also yield exactly the same RSA data. Furthermore, $\hat{r}_n$ does not constrain the individual birth and death rates, but only their ratio. This provides even greater flexibility in the range of models and mechanisms that are able to fit the data exactly.

Turning to the narrower issue of the comparison between two simplified models, the table presented by Chave *et al.*[2] is misleading because their analysis is inconsistent. Our assumption[1] was that the RSA data were a

measure of $\langle \phi_n \rangle$, the average RSA, whereas Chave *et al.*[2] consider the multivariate probability distribution for ϕ_n and maximize it for the snapshot in question. To compare the two models consistently, we computed $\langle \phi_n \rangle$ by generating 100,000 realizations of each plot with the estimates by Chave *et al.*[2] of the parameters Θ and m. The results for four forests (forests with a large value of m are very slow in their convergence) shown in Table 1 do not support the claim by Chave *et al.*[2].

In summary, any RSA data set is exactly equivalent to an effective symmetric density-dependence model. The key unknown is what led to this effective density dependence: did it arise from a dispersal-limitation model, by some other mechanism, or is there intrinsic symmetric density dependence at play? It is not possible, in principle or in practice, to distinguish between these options given just the RSA data[1].

Igor Volkov*†, Jayanth R. Banavar*,
Fangliang He‡, Stephen P. Hubbell§||,
Amos Maritan¶

*Department of Physics, 104 Davey Laboratory, and †Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
e-mail: jayanth@phys.psu.edu
‡Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada
§Department of Plant Biology, The University of Georgia, Athens, Georgia 30602, USA
||The Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama
¶Dipartimento di Fisica 'G. Galilei', Universitá di Padova, 35131 Padova, Italy

1. Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. *Nature* **438,** 658–661 (2005).
2. Chave, J., Alonso, D. & Etienne, R. S. *Nature* **441,** doi:10.1038/nature04826 (2006).
3. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. *Nature* **424,** 1035–1037 (2003).

Table 1 | **Model comparisons**

Plot	L_1	L_2	L_3
BCI, Panama	−314.5	−315.0	−314.0
Pasoh, Malaysia	−363.8	−365.3	−363.7
Lambir, Malaysia	−390.1	−391.2	−390.5
Sinharaja, Sri Lanka	−258.8	−258.5	−258.9

A comparison is shown between the analysis by Chave *et al.*[2] of the dispersal-limitation model (L_1), the density-dependent symmetric model (L_2) and the dispersal-limitation model proposed in ref. 3 (L_3) (the smaller the absolute value of L, the maximum log-likelihood value, the better the quality of the fit). Both dispersal-limitation models produce almost identical fits and are statistically comparable to the density-dependent model.