

Appendix from Jabot and Chaves, “Analyzing Tropical Forest Tree Species Abundance Distributions Using a Nonneutral Model and through Approximate Bayesian Inference”

(Am. Nat., vol. 178, no. 2, p. E37)

Test of Deviation from Neutrality

The neutrality test developed here detects in which cases Hubbell’s model cannot explain the observed species abundance distribution. Hubbell’s model makes use of two parameters: θ describes the regional diversity, while m describes the immigration rate into the local community. Increasing θ or m leads to local communities with larger species richness S (fig. A1A). Jabot and Chave (2009) showed that in neutral communities of same richness (hence with parameters θ and m located close to a level line in fig. A1A), Shannon’s index H is a unimodal function of θ (fig. A1B; fig. 1 in Jabot and Chave 2009). The range of H values typically produced by neutral simulations does not encompass all possible values from 0 to $\ln(S)$, hence two zones of H values are outside of the neutral range and correspond to communities where neutrality can be rejected based on species abundance distribution (fig. A1B). Lower H values than expected under neutrality can be reproduced with our model with positive δ while larger H values can be reproduced with negative δ .

Our neutrality test contains four steps. First, the neutral parameters θ and m are estimated from the species abundance distribution of a particular data set. This estimation is performed with the software TeTame (Jabot et al. 2008, <http://www.edb.ups-tlse.fr/equipe1/chave/tetame.htm>). This program uses the likelihood formula of Etienne (2005) to estimate the neutral parameters by maximum likelihood. In some cases, there are two different local maxima L_1 and L_2 for this likelihood in the parameter space (Etienne et al. 2006). In such cases, two different estimates for the parameters are found by the software TeTame, with respective weights of evidence w_1 and w_2 given by (Burnham and Anderson 2002):

$$w_1 = \frac{L_1}{L_1 + L_2}, \quad (\text{A1})$$

$$w_2 = \frac{L_2}{L_1 + L_2}. \quad (\text{A2})$$

Second, the estimated values of θ and m are used to simulate neutral communities. When there are two local maxima L_1 and L_2 in the likelihood profile, the first parameter set is used for $1,000 \times w_1$ simulations, and the second parameter set for the remaining $1,000 \times w_2$ simulations. This choice of using both parameter estimates is done for two reasons. First, pragmatically, in some cases $L_1 = L_2$, or they are so close that their difference is missed numerically. Second, even in cases where one maximum is clearly the global maximum likelihood ($L_1 < L_2$ or $L_2 < L_1$), we want to be sure that the data set cannot be reproduced by any of these locally best fit parameterization of the neutral model. Using only the global maximum or one of the two equal local maximum would produce similar results (data not shown). Simulations make use of the algorithm proposed by Etienne (2005). In the simulations, the only inputs are θ , m , and J , the size of the community that we take equal to the one of the tested data set. Simulations produce communities of variable species richness S . Since we want to test whether best-fit neutral communities produce Shannon’s index H comparable to what is observed in real data and since H is correlated with species richness S , we need to control for variation in S in the simulations. We do this by retaining only simulations which have the same species richness as in the tested data set. We run the simulations until 1,000 such compatible communities are simulated.

Third, for each retained simulated community, we compute Shannon’s index H . Hence, we obtain a null distribution of H values depicting the range of equitability values produced by the neutral model that best fits the

tested data set. Fourth, the H value computed in the tested data set is compared to the null distribution of H values to test whether the data set is well described by this neutral model.

Details on the Simulation Algorithm Used in the ABC Procedure

Step 1: Simulation of a Large Neutral Community of Size J_n

This simulation step is achieved with the algorithm proposed by Etienne (2005). This algorithm sequentially draws the J_n community members and determines which species they belong to as a function of the previous draws and of the parameters θ and $I = [m \times (J_n - 1)] / (1 - m)$ (eq. [7]). The procedure consists of using a double label for each individual k : one for its species s_k and one for its ancestry a_k . Then, the probability $p_l(j)$ for the j th drawn individual of belonging to ancestry l given the previous draws $(a_1, \dots, a_{j-1})$ and the probability $q_m(j)$ for the j th drawn individual of belonging to species m given the previous draws $(s_1, \dots, s_{j-1}, a_1, \dots, a_j)$ can be expressed as (Etienne 2005)

$$p_l(j) = 0 \quad (\text{A3})$$

if $l > A + 1$,

$$p_l(j) = \frac{I}{I + j - 1} \quad (\text{A4})$$

if $l = A + 1$, and

$$p_l(j) = \frac{\sum_{k < j} 1(a_k = l)}{I + j - 1} \quad (\text{A5})$$

if $l < A + 1$, where A is the number of different ancestry labels in the $(j - 1)$ draws, and $1(x)$ is a function that equals 1 if x is true and 0 otherwise.

$$q_m(j) = 0 \quad (\text{A6})$$

if $m > S + 1$,

$$q_m(j) = \frac{\theta}{\theta + A} \quad (\text{A7})$$

if $m = S + 1$ and $a_j = A + 1$,

$$q_m(j) = \frac{\sum_{k < j} 1(s_k = m) \times 1(a_k \notin \cup_{n < k} a_n)}{\theta + A} \quad (\text{A8})$$

if $m < S + 1$ and $a_j = A + 1$, and

$$q_m(j) = 1(s_{k \min(j)} = m) \quad (\text{A9})$$

if $m < S + 1$ and $a_j < A + 1$, where S is the number of different species labels in the $(j - 1)$ draws, and $k \min(j) = \min \{k/a_j \in \cup_{n \leq k} a_n\}$.

Step 2: Simulation of a Nonneutral Community of Size $J \ll J_n$

The nonneutral community of size J is initialized by a random draw from the larger neutral community of size J_n . This local community is used as an initial condition, from which the nonneutral dynamics starts. In this nonneutral dynamics, each individual dies at rate $N_i^{-\delta}$, where N_i stands for the local abundance of species i , and is replaced by the descendant of an individual chosen at random in the larger (fixed) neutral community of size J_n . To speed up this dynamics, a tenth of the individuals die simultaneously in the local community and are then replaced simultaneously. One hundred such death-recruitment cycles are performed (i.e., each individual dies and is replaced 10 times on average) to ensure that the dynamic nonneutral equilibrium is reached (see below).

Details of the Approximate Bayesian Computation (ABC) Method

Local nonneutral communities were simulated using a forward-in-time algorithm for a time sufficient to reach a dynamic equilibrium. Several assumptions were made: (a) number of iterations in the forward-in-time dynamics, (b) size of the large neutral community, and (c) number of simultaneous deaths in the forward dynamics. Sensitivity of our results to these assumptions is discussed below.

a) We simulated 100 cycles of death of one-tenth of the community (thus 10 generation cycles). We also ran ABC simulations with 200 cycles of death and computed the correlation between ABC δ estimates in the Panama Canal watershed (PCW) data set with 100 cycles of death and those with 200 cycles of death. We obtained an R^2 of 98%.

b) We then tested that $J_n = 10,000$ is sufficiently large for the initial neutral metacommunity. We performed ABC simulations with $J_n = 20,000$ individuals. We computed the correlation between ABC δ estimates in the PCW data set with large neutral community sizes 10,000 and 20,000. We obtained an R^2 of 96%.

c) To speed up ABC simulations, we assumed that 10% of the community died simultaneously during the forward dynamics. We also performed ABC simulations with only one-twentieth of the community dying at the same time (and did 200 such death cycles instead of the 100 so that the total community turnover is the same). We computed the correlation between the ABC δ estimates in the PCW data set with this alternative disturbance rate and with the original one. We obtained a R^2 of 97%.

Why Can the ABC Method Infer δ But Not the Three Parameters Jointly?

Hubbell's model fails to explain the evenness of the species abundance distribution in some plots. This is why our nonneutral parameter δ can be estimated from such data. For each (J, S) combination observed in the PCW plots, we simulated 100 communities with randomly drawn parameters θ , m , and δ and inferred the δ value by ABC (see "Methods" in the main text). We then computed the correlation coefficient R^2 between simulated and inferred δ values. These R^2 values measure the estimation efficiency of the ABC method for each (J, S) combination. We found that the estimation efficiency of our method is fairly good for sample sizes of J above 200 individuals and does not increase much with sample size once this threshold is reached (fig. A2). Furthermore, this estimation presents little bias: the slope of the regression between simulated and inferred values has a median value of 0.79 and ranges from 0.61 to 0.93, while the intercept has a median value of -0.02 and ranges from -0.09 to 0.02 (excluding the plot where $R^2 = 0.27$). Note also that this small contraction of inferred values toward zero and this small bias toward negative δ values make our finding of positive δ values in real plots conservative. Similar results were obtained using (J, S) values corresponding to the ones in the large plots.

Although δ can be estimated, we cannot jointly estimate θ and m (see Etienne et al. 2006; Jabot and Chave 2009). The joint posterior distribution of these last two parameters forms a ridge (similar to the one in fig. A1A), and the position of θ and m on this ridge cannot be estimated. Adding new summary statistics did not change this, reflecting the fact that a species abundance distribution alone is insufficient to jointly infer all the model parameters. In a recent paper where we presented an ABC approach to infer the two neutral parameters θ and m (Jabot and Chave 2009), we argued that phylogenetic information was necessary to accurately estimate these parameters. We showed that the statistics of phylogenetic tree shape B_1 (Shao and Sokal 1990) was positively correlated with $\ln(\theta)$ and provided key information to locate the best-fit neutral parameters on their ridge of high likelihood values. Here, we consider smaller communities than in our previous article (J of the order of 500, compared to 20,000 in Jabot and Chave 2009). For such sample sizes, phylogenies do not contain any more valuable information, since the variance in phylogenetic tree shape due to the stochasticity of the neutral model becomes far greater than the variation of phylogenetic tree shape linked to parameter values (fig. A4). We thus cannot use phylogenetic information here to improve parameter estimation.

Output of the Nonneutral Model Fit to the 10 Large Smithsonian Tropical Research Institute Center for Tropical Forest Science Plots

To visualize graphically the improvement in fit brought by our nonneutral model, we simulated 100 nonneutral communities with parameter values drawn in the posterior distribution for each of the 10 large plots, retaining only those with a species richness S equal to the one observed in the plot. The average ranked species

abundance distribution and associated standard deviations were computed for each plot (fig. A3). The fit is better than the one of the neutral model, although three plots (Lambir, Lenda, and Yasuni) are still not very well modeled by our nonneutral model.

Location of the Published Data Used in This Article

The first data set (large plots) can be downloaded at <http://www.sciencemag.org/content/suppl/2006/06/08/1124712.DC1/1124712SupportData.zip>. The second data set (small plots) can be downloaded at <http://www.sciencemag.org/content/suppl/2002/01/24/295.5555.666.DC1/ConditWebTable.xls>

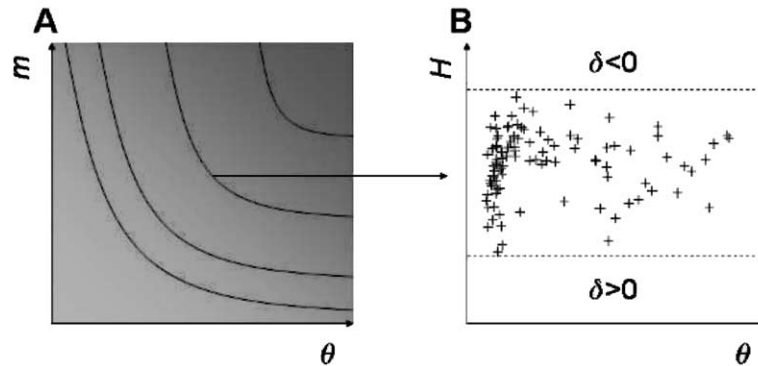


Figure A1: Principle of the neutrality test. *A*, Schematic representation of the variation in local species richness in neutral communities (represented by the level of gray) as a function of the neutral parameters θ and m . Black curves represent parameter regions leading to constant richness. *B*, Shannon's index H values typically encountered along a line of constant richness as a function of parameter θ . Each cross represents the H value obtained in a neutral community with a particular θ value. H values below the inferior dotted line or above the superior dotted line are very unlikely to be encountered in neutral communities. In contrast, communities simulated with our nonneutral model have a wider range of H values, depending on the δ value used in the simulations. Communities simulated with positive (respectively, negative) δ values are likely to occur below the inferior (respectively, above the superior) dotted line.

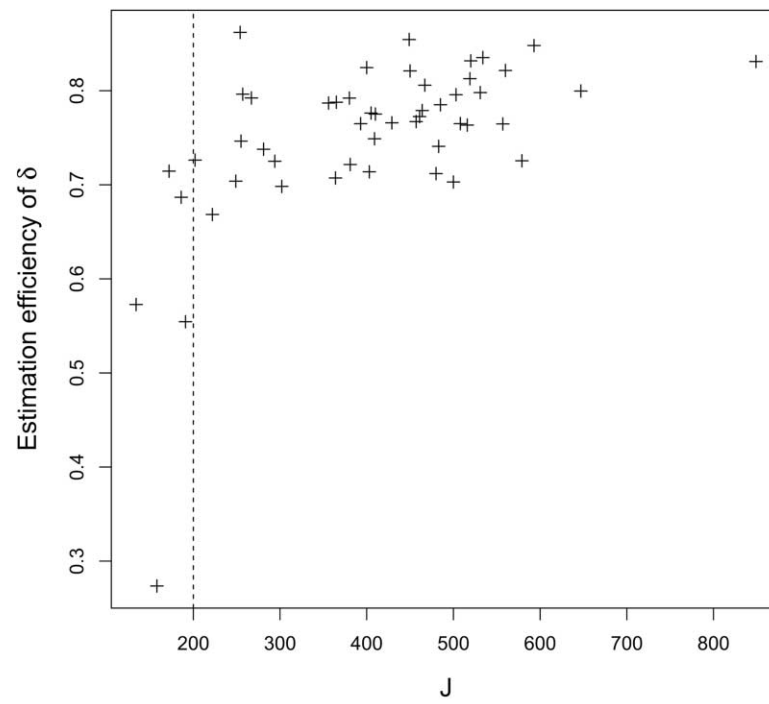


Figure A2: Effect of sample size J on the estimation efficiency of δ in the Panama Canal watershed data set.

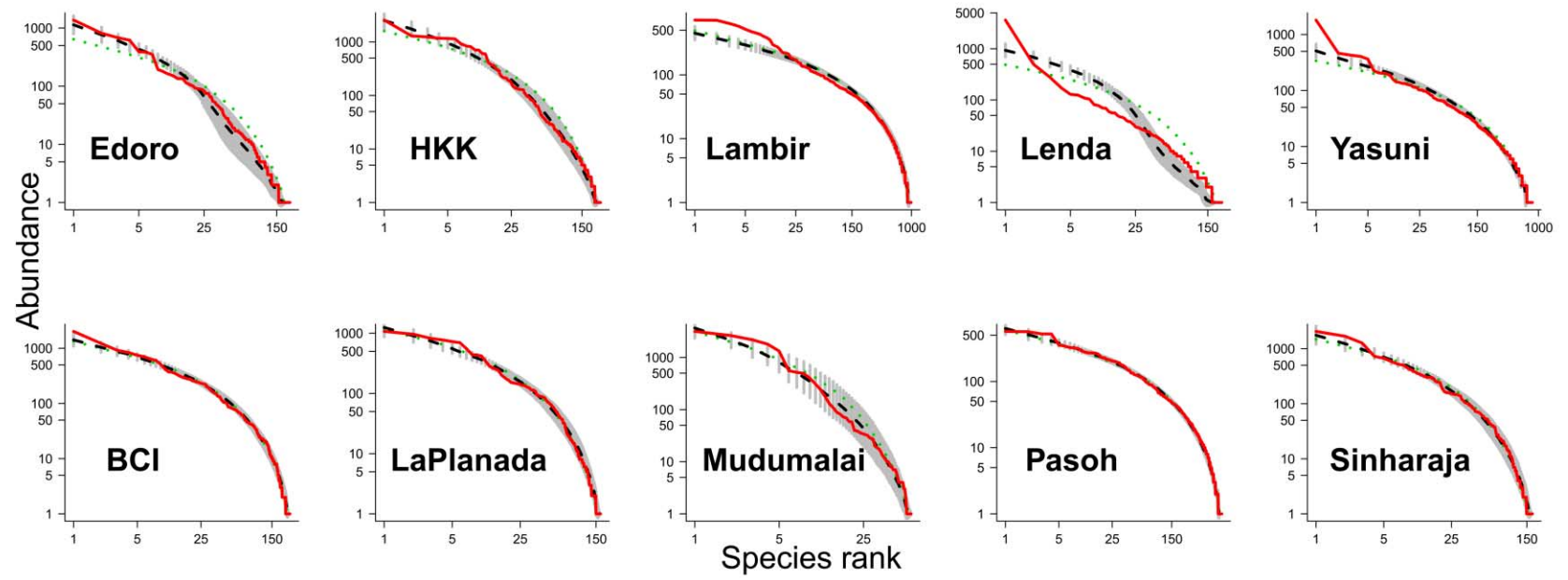


Figure A3: Fit of the nonneutral model for the 10 large Smithsonian Tropical Research Institute Center for Tropical Forest Science plots. The red lines stand for observed rank abundance curves. The black dashed lines represent nonneutral species rank abundance curves averaged over 100 simulations with parameters drawn in the posterior distributions. Gray bars represent standard deviations. The green dotted lines represent neutral species rank abundance curves averaged over 100 simulations with best-fit parameters.

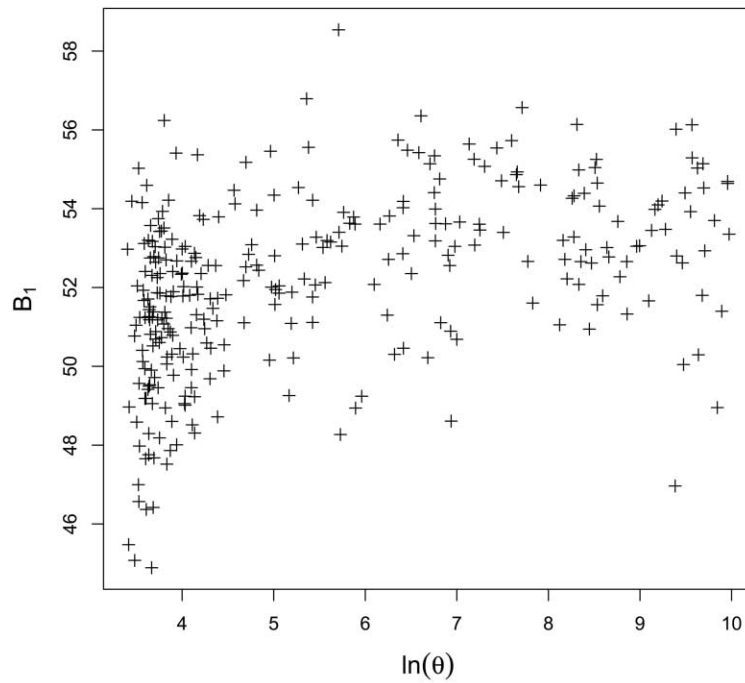


Figure A4: Phylogenetic tree shape (measured by B_1) no longer contains information on θ for small sample size $J = 500$; B_1 values in 300 simulated neutral communities of size $J = 500$ and species richness $S = 100$. There is no significant positive relationship between $\ln(\theta)$ and B_1 , contrary to what happens for larger sample sizes of $J = 20,000$ individuals (see Jabot and Chave 2009).

Table A1. Test of deviation from neutrality and values of the parameter δ in each plot of the Panama Canal watershed

Plot	J	S	H	P value	δ
P 01	400	63	3.13	<.001	.68
P 02	409	84	3.90	.42	-.04
P 03	365	74	3.82	.49	-.16
P 04	450	94	4.06	.50	-.16
P 05	364	71	3.43	.01	.57
P 06	480	78	3.62	.12	.27
P 07	380	93	3.95	.13	.26
P 08	560	94	3.54	<.001	.68
P 09	503	107	3.91	.008	.47
P 10	403	78	3.65	.089	.34
P 11	449	75	3.43	.009	.52
P 12	520	74	3.33	.009	.57
P 13	647	60	2.44	<.001	.79
P 14	381	92	3.93	.14	.26
P 15	457	91	3.91	.29	.16
P 16	467	90	3.69	.014	.52
P 17	461	63	3.02	.002	.70
P 18	429	86	3.89	.48	.07
P 19	519	89	3.66	.016	.45
P 20	534	90	3.70	.037	.39
P 21	405	78	3.76	.38	.12
P 22	508	75	3.37	.009	.56
P 23	579	60	2.70	<.001	.79
P 24	557	60	2.95	.012	.62

Table A1 (Continued)

Plot	<i>J</i>	<i>S</i>	<i>H</i>	<i>P</i> value	δ
P 25	593	84	3.80	.57	-.01
P 26	485	76	3.41	.009	.53
P 27	393	61	3.44	.26	.12
P 28	410	63	3.34	.08	.37
P 29	356	65	3.33	.023	.51
P 30	302	64	3.16	<.001	.75
P 31	483	159	4.60	.10	.24
P 32	158	105	4.50	.34	-.46
P 33	191	85	4.10	.20	.27
P 34	172	69	3.79	.046	.47
P 35	186	79	3.95	.027	.46
P 36	254	88	4.04	.18	.26
P 37	257	100	4.20	.08	.29
P 38	267	94	4.21	.49	-.23
P 39	202	64	3.76	.41	-.10
P 40	134	48	3.20	<.001	.76
P 41	222	84	3.84	.001	.65
C 1	281	49	3.19	.17	.29
C 2	255	48	2.95	.008	.67
C 3	249	54	3.26	.041	.48
C 4	294	61	3.26	.008	.62
S 0	464	90	3.84	.18	.24
S 1	531	77	3.74	.49	-.10
S 2	500	68	3.56	.48	.06
S 3	516	75	3.82	.59	-.37
S 4	849	70	3.16	.039	.44
BCI 1	448	93	4.02	.54	-.14
BCI 2	435	84	3.85	.44	.08
BCI 3	463	90	3.81	.12	.28
BCI 4	508	94	3.98	.38	-.02
BCI 5	505	101	3.97	.17	.19
BCI 6	412	85	3.78	.10	.29
BCI 7	416	82	3.84	.42	.05
BCI 8	431	88	3.91	.20	.07
BCI 9	409	90	3.76	.013	.46
BCI 10	483	94	3.89	.18	.22
BCI 11	401	87	3.86	.22	.19
BCI 12	366	84	3.70	.009	.48
BCI 13	409	93	3.98	.21	.09
BCI 14	438	98	4.02	.31	.12
BCI 15	462	93	3.96	.37	.08
BCI 16	437	93	3.92	.18	.21
BCI 17	381	93	3.74	<.001	.61
BCI 18	347	89	3.94	.18	.19
BCI 19	433	109	4.01	.01	.43
BCI 20	429	100	4.08	.46	.09
BCI 21	408	99	3.97	.07	.34
BCI 22	418	91	3.76	.012	.49
BCI 23	340	99	4.06	.10	.26
BCI 24	392	95	3.98	.16	.23
BCI 25	442	105	4.07	.18	.19
BCI 26	407	91	3.95	.33	.13
BCI 27	417	99	3.98	.08	.26
BCI 28	387	85	3.69	.017	.48
BCI 29	364	86	3.69	.005	.58
BCI 30	475	97	3.85	.03	.36
BCI 31	421	77	3.72	.38	.11
BCI 32	459	88	3.78	.06	.29
BCI 33	436	86	3.74	.048	.35
BCI 34	447	92	3.82	.054	.33
BCI 35	601	83	2.64	<.001	.80

Table A1 (Continued)

Plot	<i>J</i>	<i>S</i>	<i>H</i>	<i>P</i> value	δ
BCI 36	430	92	3.85	.07	.32
BCI 37	435	88	3.79	.038	.28
BCI 38	447	82	3.52	.005	.55
BCI 39	424	84	3.53	.001	.63
BCI 40	489	80	3.23	<.001	.74
BCI 41	402	102	4.05	.11	.25
BCI 42	414	87	3.97	.49	-.15
BCI 43	407	86	3.74	.024	.40
BCI 44	409	81	3.71	.098	.31
BCI 45	444	81	3.61	.027	.40
BCI 46	430	86	3.81	.17	.21
BCI 47	425	102	3.92	.015	.46
BCI 48	415	91	3.91	.22	.18
BCI 49	427	91	3.78	.021	.42
BCI 50	432	93	3.91	.16	.21
BCI 1-2-6-7	1,711	139	4.06	.27	.15
BCI 3-4-8-9	1,811	136	4.06	.17	.11
BCI 11-12-16-17	1,585	149	4.02	.02	.33
BCI 13-14-18-19	1,627	155	4.25	.12	.12
BCI 21-22-26-27	1,650	153	4.16	.08	.18
BCI 23-24-28-29	1,483	150	4.10	.029	.26
BCI 31-32-36-37	1,745	135	4.01	.19	.18
BCI 33-34-38-39	1,754	131	3.82	.019	.34
BCI 41-42-46-47	1,671	154	4.20	.08	.14
BCI 43-44-48-49	1,658	138	3.98	.06	.24
BCI 1-10	4,510	170	4.14	.17	.09
BCI 11-20	4,103	187	4.29	.10	.08
BCI 21-30	4,050	183	4.24	.08	.11
BCI 31-40	4,589	167	3.90	.028	.3
BCI 41-50	4,205	177	4.16	.07	.14

Note: *J* = sample size, *S* = species richness, *H* = Shannon's index. Bold numbers indicate *P* values under .05. The plot grouping at Barro Colorado Island was performed among contiguous plots. These data were used to construct figure 2 in the main text.

Literature Cited Only in the Appendix

- Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer, New York.
- Etienne, R. S., A. M. Latimer, J. A. Silander, and R. M. Cowling. 2006. Comment on "Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot." *Science* 311:610b.
- Shao, K. T., and R. R. Sokal. 1990. Tree balance. *Systematic Zoology* 39:266–276.