Species abundance distribution results from a spatial analogy of central limit theorem

Arnošt L. Šizling[a,b,1], David Storch[a,c], Eva Šizlingová[d], Jiří Reif[c,e], and Kevin J. Gaston[b]

[a]Center for Theoretical Study, Charles University and Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic; [b]Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom; Departments of [c]Ecology and [d]Philosophy and History of Science, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic; and [e]Institute for Environmental Studies, Faculty of Science, Charles University in Prague, Benátská 2, 128 01 Praha 2, Czech Republic

Edited by James H. Brown, University of New Mexico, Albuquerque, NM, and approved March 2, 2009 (received for review October 8, 2008)

The frequency distribution of species abundances [the species abundance distribution (SAD)] is considered to be a fundamental characteristic of community structure. It is almost invariably strongly right-skewed, with most species being rare. There has been much debate as to its exact properties and the processes from which it results. Here, we contend that an SAD for a study plot must be viewed as spliced from the SADs of many smaller nonoverlapping subplots covering that plot. We show that this splicing, if applied repeatedly to produce subplots of progressively larger size, leads to the observed shape of the SAD for the whole plot regardless of that of the SADs of those subplots. The widely reported shape of an SAD is thus likely to be driven by a spatial parallel of the central limit theorem, a statistically convergent process through which the SAD arises from small to large scales. Exact properties of the SAD are driven by species spatial turnover and the spatial autocorrelation of abundances, and can be predicted using this information. The theory therefore provides a direct link between SADs and the spatial correlation structure of species distributions, and thus between several fundamental descriptors of community structure. Moreover, the statistical process described may lie behind similar frequency distributions observed in many other scientific fields.

log-normal distribution | spatial autocorrelation | spatial turnover

Many models have been proposed to explain the general observation that the majority of species are rare, and to predict the major properties of the species abundance distribution (SAD) (1). Some assume a particular biological process, such as sequential niche division among species (2), stochastic population dynamics driven by simple rules and constraints (3, 4), or spatial rules imposed on species geographic distributions (5–7). These models can produce quite realistic SADs, often close to lognormal distributions. However, the ubiquity of the SAD pattern (i.e., its independence of particular taxon specifics and other biological settings) indicates that the processes responsible are much more general, and perhaps of a statistical rather than a biological nature (7). Indeed, similar patterns have also been observed in many nonbiological systems (8).

It has been suggested (9) that the approximately log-normal shape of the SAD might result from a random multiplicative process acting on abundances (i.e., an additive process acting on their logarithms). Purely multiplicative processes cannot, however, be responsible for the multiple SADs that are observed simultaneously at several spatial scales (10). The reason is that the SAD of an assemblage on a study plot (whose bounds may be arbitrary or ecological) is necessarily spliced from the SADs of subassemblages occurring in nonoverlapping subplots covering that plot (6, 11, 12). Because abundances for the whole plot arise by summing the abundances of the subassemblages across all of the subplots, an additive process acting on abundances must also play a role. In fact, many models of the SAD explicitly or implicitly comprise additive processes (4, 13, 14). However, this has never been clearly identified as the major mechanism responsible for the shape of the SAD. Here, we show that it is the additive process itself that represents the clue to understanding the universally reported shape of SADs, regardless of any model-specific dynamics.

Suppose that the SAD for an assemblage on a plot (SAD*p*) is comprised from those of the subassemblages on nonoverlapping subplots (SAD*s-p*). We can ask how the properties of the SAD*p* depend on the properties of the SAD*s-p*, and to what extent it is affected solely by the process through which the SAD*p* arises. We will explore the possibility that the SAD*p* is independent of the SAD*s-p* for the smallest subplots (initial-SAD*s-p*), because the statistical process giving rise to the SAD*p* outweighs the contribution of the particular initial-SAD*s-p*. This situation would be similar to the process that lies behind the central limit theorem (CLT) [introduced in 1733 by de Moivre and proved in 1901 by Lyapunov (15)]. According to the CLT, the normal (Gaussian) distribution arises by addition of many mutually independent variables with finite variances regardless of their original distribution.

The process through which an SAD*p* arises, being spliced (see *Materials and Methods*) from many initial-SAD*s-p*, is, however, different, because it is necessarily spatially determined. This means that the abundances of each of two adjacent subplots to be joined are dependent on each other, and that some species are missing from some subplots. The SAD then arises by summing pairs of abundances of the species common to both joined subplots, and appending abundances of the species missing from one subplot at each recurrent step. The resulting distribution is thus shaped by the spatial correlation structure, which is exemplified by species spatial turnover and the spatial autocorrelation of abundances. Positive correlation between the abundances of given species in neighbouring subplots leads to elongation of the right-hand tail of the SAD, because eventual high abundance in one subplot is probably added to similarly high abundance in another. This elongation occurs even if abundances are not correlated (for abundances are positive and thus only the right-hand tail can grow), but the stronger is the autocorrelation the faster the tail grows, regardless of the exact nature of that autocorrelation (Fig. 1). However, species spatial turnover leads to the addition of species occurring in only one of two joined subplots, which produces a prevalence of rare species in the spliced SAD. These two effects combined thus lead to a right-skewed abundance distribution.

Results

We simulated the process described above (for details, see *Materials and Methods*), varying its inputs in terms of the shape

Author contributions: A.L.S., D.S., E.S., and J.R. analyzed data; and A.L.S., D.S., E.S., J.R., and K.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: sizling@cts.cuni.cz.

This article contains supporting information online at www.pnas.org/cgi/content/full/0810096106/DCSupplemental.

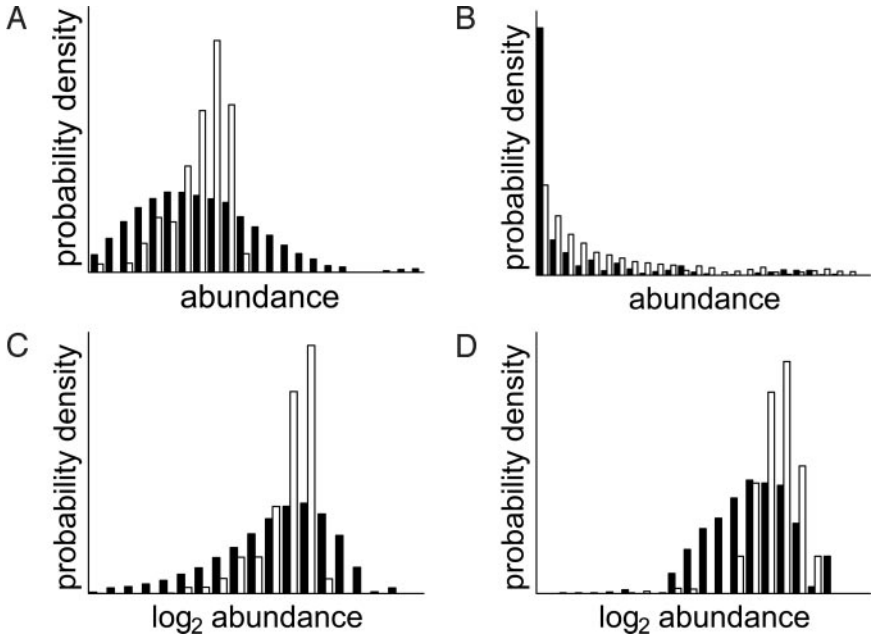

Fig. 1. Probability density functions presented as histograms of distributions arising through the splicing (above: abundance classes on arithmetic scale; below: abundance classes on logarithmic scale). Distributions of uncorrelated (*A* and *C*) and autocorrelated (*B* and *D*) abundances for high (black) and low (white) spatial turnovers (for Settings see *Materials and Methods*). Low correlation and turnover (white plot in *A*) approach the standard Central Limit Theorem and produces a nearly Gaussian distribution. High correlation and turnover (black plot in *B*) elongates the right tail, producing a right-skewed, almost lognormal (black plot in *D*), distribution. The positive skewness of the distribution is thus promoted by spatial turnover and autocorrelation.

of the initial-SAD*s-p*, and using observed levels of species spatial turnover (measured as the proportion of species common to both neighbouring subplots, i.e., the Jaccard index, *J*) and of spatial autocorrelation of abundances (determined by Pearson's correlation coefficient, *r*) (see *Materials and Methods*). We proceeded in a step-by-step manner, splicing pairs of neighbouring initial-SAD*s-p* in the first step, then (the second step) splicing pairs of neighbouring SAD*s-p* that resulted from the first step, and so on up to the SAD*p* of the whole plot.

Three different simulation experiments were performed, each beginning with a differently shaped initial-SAD*s-p* (left-skewed, regular, right-skewed). We checked whether all of the simulations had reached a particular shape of the distribution, whether these shapes were the same regardless of initial-SAD*s-p*s, and ultimately compared the resulting distributions from each of the 3 simulations with the observed SAD*p* of 2 well-resolved spatial datasets. These latter comprised (*i*) trees within a tropical study plot on Barro Colorado Island (ref. 16 and 17 and http://ctfs.si/edu/datasets/bci) (see *Materials and Methods*), and (*ii*) central European birds mapped on a transect through the whole of the Czech Republic (7) (see *Materials and Methods*). All of the observed and simulated SAD*p* and SAD*s-p* to be compared were standardized to the same mean abundance (i.e., $a_{st} = a/\bar{a}$, where a_{st} is the standardized abundance, a is a raw abundance, and $\bar{a}$ is mean abundance), and veiled by minimum observed values. The SAD*s-p* to be spliced were neither standardized nor veiled.

Both datasets revealed close similarity to the respective SAD*s-p* resulting from the convergent processes (Figs. 2 and 3). None of (*i*) a rank plot (Figs. 2 bottom row and 3*C*), (*ii*) the maximum distance between simulated and observed cumulative distributions [Kolmogorov–Smirnov statistics (KS)] (see *SI Appendix*, Fig. S1) or (*iii*) the skewness of the SAD*p* of log-transformed abundances (Fig. 3 *A* and *B*) revealed disagreement between observation and the SAD*p* resulting from the simulated splicing from the 200th step on. Visually, the simulations followed the usually reported shape (i.e., sigmoid and almost symmetric rank-log-abundance plot) from the 20th step on (for steps of 50 and 100 see Fig. 2, second and third rows). A nonparametric DKW test (18) based on Kolmogorov–Smirnov statistics could not reject agreement between modeled and observed SAD*p* in any case, whereas for the earlier steps the agreement was rejected at $P < 0.01$ (see *Materials and Methods*). The difference between SAD*p* for tropical trees and central European birds (Fig. 3*C*) was accurately predicted by the difference in species spatial turnover, *J*, and spatial autocorrelation of abundances, *r*. The probabilistic process of splicing of SAD*s-p* in neighbouring subplots, modeled by our simulations, thus represents a realistic mechanism for the emergence of observed SADs.

Discussion

We have demonstrated a universal principle that inevitably applies if summing variables irregularly distributed in space or time, and thus inevitably affects the SAD. This principle is similar to the CLT, which states that sums of the same numbers of mutually independent variables approach a bell-shaped distribution. We argue that sums of various numbers of mutually independent or dependent, positive variables approach a right-skewed distribution, which is more or less symmetric on a logarithmic scale. The crucial difference between the CLT and our principle, i.e., "various numbers of variables," corresponds to the fact that some species are missing in some samples, whereas the potential mutual dependence of variables corresponds to spatial intraspecific correlation between abundances of two adjacent plots. The mutual dependence is not, however, necessary, because it only determines how heavy is the right tail of the distribution (Fig. 1). Applying this simple principle to abundance data, we get realistic SADs. Because missing observations (either really missing or missing because of the limitations of the method of observation) and/or their mutual dependence is rather common across all fields of science, we would not be surprised if this principle governed many other asymmetric distributions observed there (8).

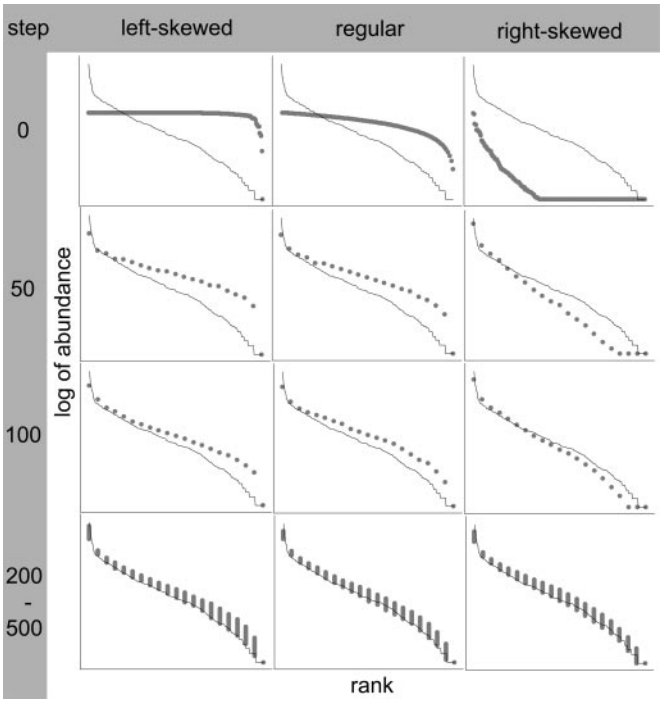

Fig. 2. Comparison of observed (full line) and simulated (gray circles) rank-log-abundance plots for tropical trees. Simulations are shown for 3 different initial-SAD*s-p* (left-skewed, regular, right-skewed) (columns) and for various steps of SAD*s-p* and SAD*p* (rows - steps 0, 50, 100, and 200–500). The observed SAD*p* is apparently indistinguishable from the fully converged simulated distributions, regardless of the initial-SAD*s-p*. Plotted distributions are standardized (mean abundance = 1) and veiled by minimum observed values for comparison.

The fit of our prediction was obtained using the simplified assumption that both of the spatial parameters are constant over all steps (i.e., all spatial scales). This is clearly unrealistic, because at least spatial turnover has been reported to be scale dependent (19). However, by parameterizing the process using measurements extracted from the whole plot, we set the process by the parameters crucial for the final convergent stage. If considering only a small part of the transect data, we should not assume that the observed SAD has yet converged, but we might still assume agreement between the observed abundance distribution and simulated SAD*s-p* at some particular step of the process. That is exactly what happened for all of the initial-SAD*s-p* and, surprisingly, for various settings (see *Materials and Methods*) of the parameters (Fig. 4 and *SI Appendix*, Fig. S2). The process is thus so pervasive that it predicts the observed shape, whatever the initial-SAD*s-p*, even for smaller areas with an SAD*s-p* that does not represent a complete convergent stage.

Having demonstrated this universal principle, it is possible to see why so many models that have been proposed (1) produce quite realistic SADs. All of the spatial models include the existence of species spatial turnover and most of them spatial autocorrelation. Various mechanisms then only tune their exact values to fit a model to data. For instance, manipulating the proportion of newly arriving individuals (13, 20) or the proportion of newly established species (21) effectively leads to specific levels of species turnover and spatial autocorrelation, and so it is not surprising that it affects the shapes of resulting SADs. Many similar processes effectively produce species turnover at several spatial scales, which is, according to our theory, the proximate driver of observed SADs.

Importantly, we need not assume that SADs for real assemblages have actually emerged because of the large number of steps of the process described above. However, we argue that this process encapsulates the major feature of the emergence of observed SAD*p*, which is the splicing of SAD*s-p* in neighbouring subplots. In reality, the spatial scale of the initial-SAD*s-p* may correspond to the spatial requirements of an individual, i.e., home range of an animal or the spatial requirement of a plant. The shape of such an initial-SAD*s-p* may be driven by that of the species-body size distribution (22), and thus may be much less extreme (i.e., closer to the shape resulting from the convergent process) than those used in our simulations. The process thus might actually require a much smaller number of steps to reach full convergence.

Another possibility is that an SAD really originates from many steps of splicing, starting with initial-SAD*s-p*s for extremely small patches. The "abundance" of a species in these small patches would then be represented by the probability of species occurrence, and the "true" SAD would be a frequency distribution of these probabilities. Because the probability of occurrence corresponds to the reciprocal of the size of a species' home range, the SADs might still be linked with the species-body size distribution. Both interpretations of the initial-SAD*s-p* have the potential to link our theory with the factors that affect landscape properties enabling species coexistence (productivity and habitat complexity) and species energetic and resource requirements at very local scales. According to our theory, only at the very local scale are biologically important processes taking place, whereas the patterns observed at large scales are dominated by a statistical process. The theory thus has the potential to separate statistical and biological effects. Importantly, we do not need to assume any particular "fundamental" scale (comprising initial-SAD*s-p*) from which the patterns on other scales are derived; the convergent process leads to the observed SAD shape regardless of the scale we begin with, given a sufficient number of splicing steps.

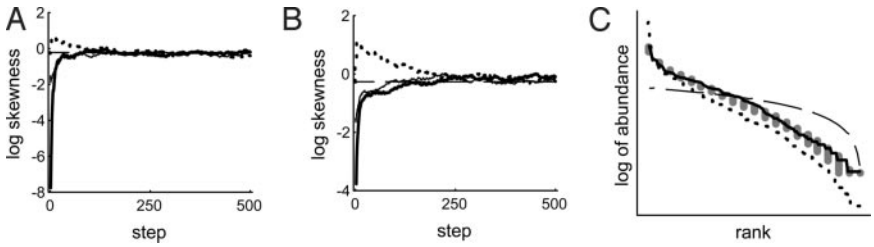

Fig. 3. Convergence of the shape of the SAD. (*A* and *B*) Convergent series of skewnesses of SAD*s-p* of log-transformed abundances starting with left-skewed (bold line), regular (thin line), and right skewed (dotted line) initial-SAD*s-p*. Each series is parameterized by species spatial turnover and spatial autocorrelation of (*A*) tropical tree and (*B*) central European bird data. Dashed lines show the observed skewnesses. (*C*) Rank-log-abundance plots of the central European bird data (thin line), and a result of the respective convergent series (gray circles; steps 300–500), which started as the regular initial-SAD*s-p* (dashed). For contrast see the tropical tree data (dotted line). SADs are standardized (mean abundance = 1), and veiled by minimum observed values.

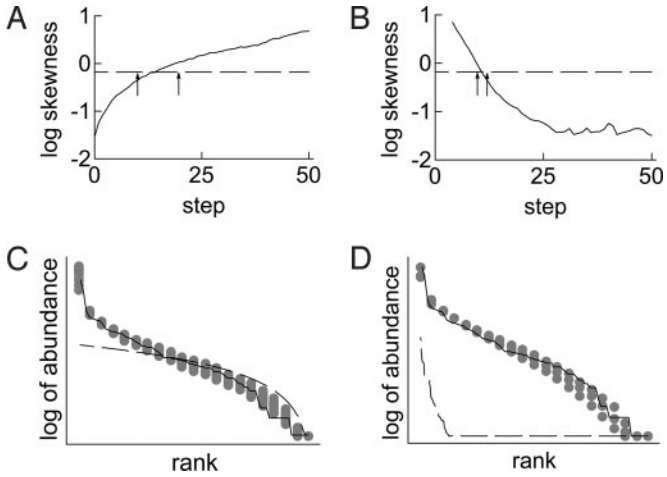

Fig. 4. Simulated series of skewnesses of SAD*s-p* of log-transformed data starting with (*A*) regular and (*B*) right-skewed initial-SAD*s-p* and Central European bird data for 1/8 of the transect. The observed value (dashed line) is crossed in between steps 10 and 20 in *A*, and 10 and 12 in *B* (spaces between arrows). The simulated rank-log-abundance plots for these steps (gray circles) are plotted in *C* and *D*, respectively. Observed SAD*p* is marked by a solid line, initial-SAD*s-ps* are dashed. Series vary in their parameters (for settings see *Materials and Methods*), but their values have little effect on the agreement between data and simulations (*SI Appendix*, Fig. S2).

Our approach comprises purely bottom-up processes leading from SADs at local scales to convergent SAD at large spatial scales. This approach is in contrast to the top-down attempts to derive particular shapes of SADs by spatial sampling of given regional SAD (23), and to prevailing macroecological consideration of regional patterns as those determining local ones (24). Biologically relevant processes may actually act at regional scales or rather at many scales interacting together. Even then, the purely statistical bottom-up process we describe has in most cases an overwhelming influence on the shapes of regional SADs, because it acts whenever there are particular local distributions (of any shape), and nonzero spatial species turnover between subplots.

Our theory provides a direct link between SADs on the one hand and species spatial turnover and autocorrelation on the other, i.e., between several fundamental descriptors of community structure. Many such links have already been determined (7), and the mathematical connections to other macroecological patterns have been demonstrated (e.g., the species-area relationship) (25). Here, we have shown that abundance patterns can be derived using three assumptions: (*i*) that most species do not occur everywhere, (*ii*) that species abundances are positive (a trivial, but critical detail), and (*iii*) that these abundances are spatially autocorrelated. These assumptions represent quite universal biological observations, and thus it is understandable that they universally lead to the observed shape of the SAD.

According to our theory, the approximately log-normal shape of SADs, universally found in species assemblages, is a consequence of a purely statistical limiting process parameterized by species spatial turnover. The exact parameters of each particular SAD are then given by the structure of species' spatial distributions, and an SAD thus reflects the spatial distribution of habitats and (meta)population and metacommunity dynamics. Therefore, as in the case of other macroecological patterns (7), the overall shape reflects a universal statistical process, but the details and particular parameters reveal biology and can bring important information about the structure and dynamics of ecological communities.

Materials and Methods

Splicing. This is a newly introduced term for an operation over probability distribution functions, which comprises summing and concatenating (appending) mutually dependent variables; the standard term "convolution" is related only to summation of (mutually independent) variables. The analytical expansion of the splicing is "f_1 spliced with f_2"$\equiv \pi_1 f_1 + \pi_2 f_2 + J f_1 *_c f_2$, where $\pi_1 + \pi_2 + J = 1$, and $*_c$ is a correlated convolution.

Simulation. It was a step-by-step process, each step with 3 inputs ((*i*) a pair of identical distributions given by *S* real positive numbers; (*ii*) Jaccard index, *J*; and (*iii*) a pair of real numbers $\{\sigma_{min}; \sigma_{max}\}$, which set up the spatial autocorrelation of abundances), and one output (a distribution given by *S* real positive numbers). Each step consisted of (*i*) drawing two sets of $S \times J$ abundances (those for species common to the two subplots) from the distributions input; (*ii*) making random pairs of these abundances $\{a_1,a_2\}$ so that $\sigma_{min}a_1 \leq a_2 \leq \sigma_{max}a_1$ (if the inequality cannot be met, the a_2 that is nearest to the constraints $\sigma_{min}a_1$ and $\sigma_{max}a_1$ is attributed to the a_1) and appending $a_1 + a_2$ to the distribution in the output; (*iii*) drawing $S \times (1 - J)$ abundances (those for species that occur only in one of two subplots) from a distribution input, and appending them to the distribution in the output. The parameter $S = 5{,}000$. Note that drawing from a distribution given by a set of particular values does not mean that only those values can be drawn. (For procedure and picture guide see *SI Appendix*, *Guide and Procedures*). For utility to run the procedure, see www.cts.cuni.cz/wiki/ecology:start.

Extracting of the Parameters. The $J = S_{com}/S_{tot}$, where S_{com} is the number of species common to the two (East and West in this case) halves of the observed plot, and S_{tot} is the number of species within the whole observed plot. The σ_{min} and σ_{max} were chosen empirically to meet the observed *r* when running simulations; the *r* is a Pearson's correlation coefficient between abundances of the two halves of the observed region; the species occurring in only one-half were excluded. This applies to both the datasets.

BCI 1983 Data. Data on 307 tropical tree species from the plot of 50 ha on Barro Colorado Island, Panama; all of the dead trees and the trees labeled as ''which not yet entered census'' were excluded.

Transect (April–June) 2004–2005 data. Data on 144 temperate bird species censused within 150m distance around each of 768 points along a linear East–West transect in south Bohemia and Moravia; points were separated by between 300 and 500 m.

Test. A test using the Dvoretzky–Kiefer–Wolfowitz inequality ($P(KS > \varepsilon) \leq 2Exp(-2n\varepsilon^2)$; $\varepsilon > 0$; *P* is the probability that KS oversteps the ε by chance; *n* is a number of samples from the tested distribution; if both the assumed and tested distributions are given by a sample, which is the case, the inequality is an even stronger criterion). KS takes values of 0.07 and 0.1 for steps from 200 on in cases of tropical tree and central European bird data, respectively. If we wanted to reject the agreement of data and simulation using these values, we would need significance levels $P > 0.09$ ($n = 307$) and $P > 0.1$ ($n = 144$), respectively. However, the values KS >0.14 that hold for all of the steps <50 in both cases, are easy to reject at level $P \ll 0.01$. The KS < 0.1 and level needed for rejection $P > 0.37$ ($n = 84$) in test for the Fig. 4.

Settings. Fig. 1: Full bars $J = 60\%$, empty bars $J = 90\%$, regular initial-SAD, histograms show stages 450–500; Fig. 1 *A* and *C*: $\{\sigma_{min}; \sigma_{max}\} = \{0;10^{99}\}$, which produces $\bar{r} \approx 0$; Fig. 1. *B* and *D*: $\{\sigma_{min}; \sigma_{max}\} = \{0.9;1.1\}$, which produces $\bar{r} \approx 0.953$; Figs. 2 and 3*A*: $J = 88.1\%$, $\{\sigma_{min}; \sigma_{max}\} = \{0.9;1.11\}$, which produces $\bar{r} \approx 0.95$ (observed values are: $J = 88.1\%$, $\bar{r} = 0.97$); Fig. 3 *B* and *C*: $J = 77\%$, $\{\sigma_{min}; \sigma_{max}\} = \{0.5;1.7\}$, which produces $\bar{r} \approx 0.84$ (observed values are: $J = 76.4\%$, $\bar{r} \approx 0.81$); Fig. 4 *A* and *C*: $J = 70\%$, $\{\sigma_{min}; \sigma_{max}\} = \{0.9;1.1\}$, $\bar{r} \approx 0.95$; Fig. 4 *B* and *D*: $J = 70\%$, $\{\sigma_{min}; \sigma_{max}\} = \{0.3;100\}$, $\bar{r} \approx 0.195$.

ACKNOWLEDGMENTS. We thank Tomáš Herben, Petr Keil, and Ethan White for valuable comments to the text. This work was supported by Marie Curie Fellowship 039576-RTBP-EIF (to A.L.Š.); Czech Ministry of Education (Grants LC06073 and MSM0021620845); Grant Agency of the Academy of Sciences of the Czech Republic (Grant IAA601970801); and a Royal Society-Wolfson Research Merit Award (to K.J.G.). The Barro Colorado Island forest dynamics research project is supported by National Science Foundation (Grants DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, and DEB-7922197 to Stephen P. Hubbell); the Center for Tropical Forest Science; the

Smithsonian Tropical Research Institute; the John D. and Catherine T. MacArthur Foundation; the Mellon Foundation; the Celera Foundation; numerous private individuals; and the hard work of over 100 people from 10 countries over the past 2 decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

1. McGill BJ, et al. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett* 10:995–1015.
2. Tokeshi M (1999) *Species Coexistence: Ecological and Evolutionary Perspectives* (Blackwell, Oxford).
3. Engen S, Lande R (1996) Population dynamic models generating the lognormal species abundance distribution. *Math Biosci* 132:169–183.
4. Hubbell SP (2001) *The Unified Theory of Biodiversity and Biogeography* (Princeton Univ Press, Princeton).
5. Harte J, Kinzig AP, Green JL (1999) Self-similarity in the distribution and abundance of species. *Science* 284:334–346.
6. Šizling AL, Storch D (2007) in Sc*aling Biodiversity,* eds Storch D, Marquet PA, Brown JH (Cambridge Univ Press, Cambridge, UK), pp 77–100.
7. Storch D, et al. (2008) The quest for a null model for macroecological patterns: Geometry of species distributions at multiple spatial scales. *Ecol Lett* 11: 771–784.
8. Nekola JC, Brown JH (2007) The wealth of species: Ecological communities, complex systems and the legacy of Frank Preston. *Ecol Lett* 10:188–196.
9. May RM (1975) in *Ecology and Evolution of Communities,* eds Cody ML, Diamond JM (The Belknap Press of Harvard Univ Press, Cambridge), pp 81–120.
10. Williamson M, Gaston KJ (2005) The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J Anim Ecol* 74:409–422.
11. Šizling AL, Storch D, Reif J, Gaston KJ (2008) Invariance in species-abundance distributions. *Theoretical Ecology* doi 10.1007/s12080–008-0031–3.
12. Storch D, Šizling AL (2008) The concept of taxon invariance in ecology: Do diversity patterns vary with changes in taxonomic resolution? *Folia Geobotanica* doi: 10.1007/s12224–008-9015–8.
13. McGill BJ (2003) Does mother nature really prefer rare species or are log-left-skewed SADs sampling artefact? *Ecol Lett* 6:766–773.
14. Marquet PA, Keymer JE, Cofré H (2003) in Macroecology: Concepts and consequences eds Blackburn TM, Gaston KJ (British Ecological Society and Blackwell Science Ltd, Oxford), pp 64–81.
15. Kallenberg O (1997) *Foundations of Modern Probability* (Springer-Verlag, New York).
16. Condit R (1998) *Tropical Forest Census Plots* (Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas).
17. Hubbell SP, et al. (1999) Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* 283:554–557.
18. Wasserman LA (2004) *All of Statistics: A Concise Course in Statistical Inference* (Springer-Verlag, Berlin).
19. Gaston KJ, Evans KL, Lennon JJ (2007) in *Scaling Biodiversity* eds Storch D, MarquetPA, Brown JH (Cambridge Univ Press, Cambridge), pp 181–222.
20. Dolman AM, Blackburn TM (2004) A comparison of random draw and locally neutral models for the avifauna of an English woodland. *BMC Ecol* 4:8.
21. Zillio T, Condit R (2007) The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* 116:931–940.
22. Allen CR (2006) Patterns in body mass distribution: sifting among alternative hypotheses. *Ecol Lett* 9:630–643.
23. Storch D, Gaston KJ (2004) Untangling ecological complexity on different scales of space and time. *Basic Appl Ecol* 5:389–400.
24. Green JL, Plotkin JB (2007) A statistical theory for sampling species abundances. *Ecol Lett* 10:1037–1045.
25. Harte J, Kinzig AP (1997) On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. *Oikos* 80:417–427.