

Predicting species abundance distributions by simultaneously using number and biomass as units of measurement

Peter A. Henderson¹ & Anne E. Magurran²

¹*Pisces Conservation Ltd., IRC House, The Square, Pennington, Lyminster, Hants., England SO41 8GN and Dept. of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*

²*Gatty Marine Laboratory, School of Biology, University of St Andrews, St Andrews, Fife, Scotland KY16 8LB*

The authors contributed equally to this work

The universal observation that some species in an ecological community are common, but many more are rare, is neatly encapsulated in a species abundance distribution (SAD)¹. However, the shape of the distribution can depend on the currency used to measure abundance ². Here we show how the SADs for numerical abundance and biomass are related and how this relationship can be used to predict the form of the SAD. When plotted in log numerical abundance, log biomass space, species points lie within an approximately triangular area the limits of which are set by body size range, and the upper limit of abundance in both metrics. Under the simplifying, but reasonable, assumption that the observed scatter of species within this region is random, the shape of the SAD is immediately derived from simple geometrical considerations. For the SAD of numerical abundance this is a power curve. The biomass SAD can be either a power curve or, more frequently, a unimodal curve, which can approximate a log normal. This log triangular random placement model serves as a null hypothesis against which actual communities can be compared. Data from two intensively surveyed local communities indicate that it can give a good approximation, with species scattered within a triangle. Further, we can predict the consequences, for the SAD, of size-

selective sampling protocols. We argue that mechanistic models of SADs must be able to account for the relative abundance of species in alternative currencies. Moreover, this approach will shed light on niche packing and may have application in environmental monitoring.

Species abundance distributions capture an essential feature of ecological communities, in that they depict a few common and many rare species. However, while this pattern of commonness and rarity is so pervasive that it has been dubbed a law of ecology ¹ it is also a pattern that defies easy explanation. Over 40 hypotheses ^{1,3}, including both statistical ⁴ and biological ⁵ models, have been proposed in the 75 years since Motomura ⁶ first suggested that species abundances are a product of niche apportionment. The challenge is not simply to replicate species abundance distributions seen in nature, which many of the existing models do convincingly, albeit on the basis of different and sometimes incompatible assumptions, but to make distinct and testable predictions ¹.

One aspect of species abundance distributions that is beginning to attract more attention, and which has the potential to shed light on the underlying mechanisms, is the relationship between the shape of the distribution and the currency used to measure species importance. Although abundance is usually expressed as numerical abundance (number of individuals), because these are the units in which the taxa (eg birds and trees) that predominate in such analyses are typically recorded, other measures are possible. Biomass is an alternative, and preferred by some ecologists eg 3, 7, 8 as it is assumed to provide a more direct measure of resource use; energy flow is correlated to body mass of individuals as metabolism scales with body mass to the 0.75 power 9, 10. The shape of a species abundance distribution can depend on the units chosen 11. For example, species abundance distributions of fish and coral are lognormal at local scales if biomass is used, but it is only at large geographic scales that a lognormal distribution

of individuals becomes apparent 2. To date there has been no compelling explanation for the discrepancies between distributions measured using different currencies. Moreover, the debate about whether numerical abundance and biomass provide equivalent insights into the way in which species subdivide resources remains unresolved 12-15.

Using some simple, and testable, assumptions we are able to predict the shape of the species abundance distribution for both biomass and numerical abundance, as well as the circumstances under which the two distributions will be similar. Similar analyses would be possible for other currencies such as cover. We focus on local scales (where community processes such as competition are likely to be important), on well-sampled communities (so that sampling effects, which can have a large influence on the perceived shape of the abundance distribution, are minimised) and discrete time periods (to avoid the confounding effects of autocorrelation 16 and turnover 17).

When species data for the two currencies are displayed on a log-log scatter plot they will be contained in a triangular-shaped polygon (Figure 1). The boundaries of this region are defined as follows. The maximum and minimum biomass for a singleton species must lie within the body mass range. Increases in the upper boundary of the region will track increases in numerical abundance, until maximum observed biomass for a single species is reached. Likewise, the minimum biomass, multiplied by numerical abundance, sets the lower boundary of the region. In short it is the range of body size, maximum numerical abundance, and maximum biomass, that constrain the distribution of species in this space.

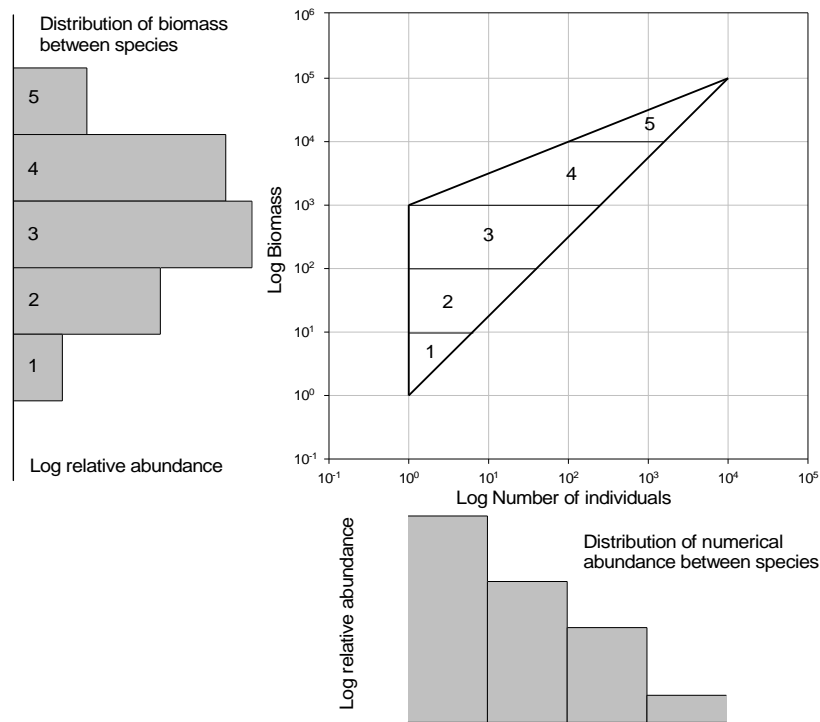


Figure 1 The derivation of the shape of the SAD in terms of biomass and number based on the assumption that the distribution of species within natural communities are scattered at random within a triangular region. In this example we assume an obtuse triangle, which is probably the commonest pattern in well-sampled communities.

As numerical abundance and biomass are typically weakly correlated (8, 18), we assume that species are distributed at random within the triangular region. It follows that the shape of the triangle will determine the species abundance distribution for both biomass and numerical abundance (Figure 1). Figure 2 shows the range of distributions that will arise from different biomass-numerical abundance couplets. We assume for simplicity that on a log-log scale the points are bounded within a straight-sided region. The shape of the SAD is easy to visualise as simply the distance across the polygon

perpendicular to the x axis for the number SAD and the y axis for the biomass SAD. In the case of Figure 2 A both biomass and number SADs follow a power relationship of the form:

$$n_s = ka^b$$

where n is the number of species, a is the size class in terms of number of individuals or biomass and K and b are constants. It is notable that in Figure 2A frequency of species per class decreases with increasing numerical abundance but increases when abundance is measured in terms of biomass. In the case of Figure 2B the SAD peaks at an intermediate biomass per species. Although the diagram represents this peak as the intercept of 2 straight lines, in practice this would form a peaked curve because of the random placement of the individual species points within the triangle. This curve might, in some circumstances, be indistinguishable from a log normal. Figure 2C represents the situation where there is a lower cutoff in the mass of individuals sampled. This could occur, for example, when the mesh size of a net allows small individuals to escape capture or when stems below a certain diameter are excluded from a survey of trees. This cutoff results in a numerical abundance SAD with a peak at the lowest abundance class, similar to a log normal distribution with a veil line close to the modal class, or to a log series distribution. Finally, Figure 2D shows the region in which the random points would be scattered given a sampling cut-off in biomass for both the maximum and minimum size of individuals sampled. It is assumed that the unsampled large individuals can only occur in small numbers. Such approaches lead to SADs with a peaked distribution in both mass and number.

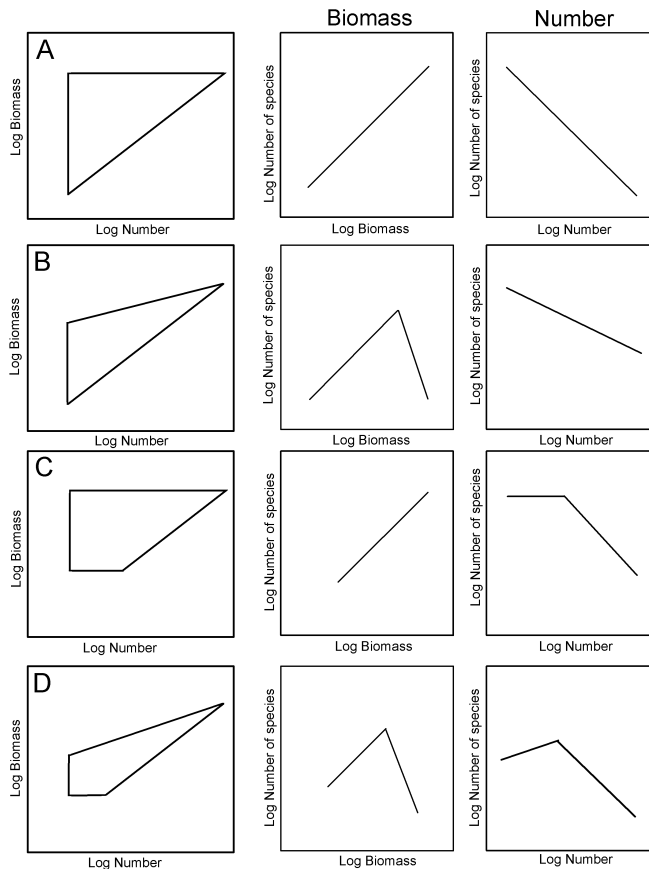


Figure 2 The predicted shape of the SAD in terms of both biomass and number for different log Number – Log Biomass areas. It is assumed that the position of the individual species abundances is scattered at random within the polygons.

We tested these assumptions using two well-surveyed local communities. In the first of these, the Milford Haven Benthic community (Figure 3 A), species are distributed within an approximately right-angled triangle as in Figure 2A. This is related to the fact that animals across a wide size range, - 10⁻³ to 10³ g wet weight - could be sampled. Given a right-angled triangle and a random distribution of species within the triangle, numerical and biomass SADs are described by a power curve (as in Figure 2A). The second example is from an intensive 2-month study of an estuarine fish community of Bridgwater Bay in winter (Figure 3B). In this case, the upper constraint

line is far from horizontal. In part, this can be explained by the inability of the sampling method to catch large fish such as sharks. The largest fish in the sample was below 103 g in weight, whereas fish >105 g occasionally occur in the area. The species lie within an obtuse triangle similar to Figure 2B.

To predict the form of the SADs we have assumed that the species are randomly distributed within the log biomass log number triangle. This assumption is supported by the observation that the positions of individual species within the triangle can vary greatly through time and show low between species correlation. Figure 4 shows the position of the nine commonest fish species in 24 monthly samples collected from Bridgwater Bay between 1987 and 1989. The points are joined in temporal order. Only Pout-Dab and Pout-Poor cod species pairs show an absolute Pearson correlation, r , in abundance > 0.5 for both number and biomass, the majority of the 36 species pairs show low positive or negative correlations (Figure 4). As the 5% critical value for $|r| = 0.404$, only 4 (11.1%) of the species pairs had r -values significantly different from 0 for both biomass and number. Movements within the triangle are particularly extensive for the commonest species sprat and whiting. These plots demonstrate that even over small periods of time species do not maintain fixed places within the number-biomass triangle. While individual trajectories are non-random, as most remain within a restricted subspace within the triangle, the net result of all the movements for species with low cross-correlations in abundance is that the distribution of points at any sampling instant may appear close to random.

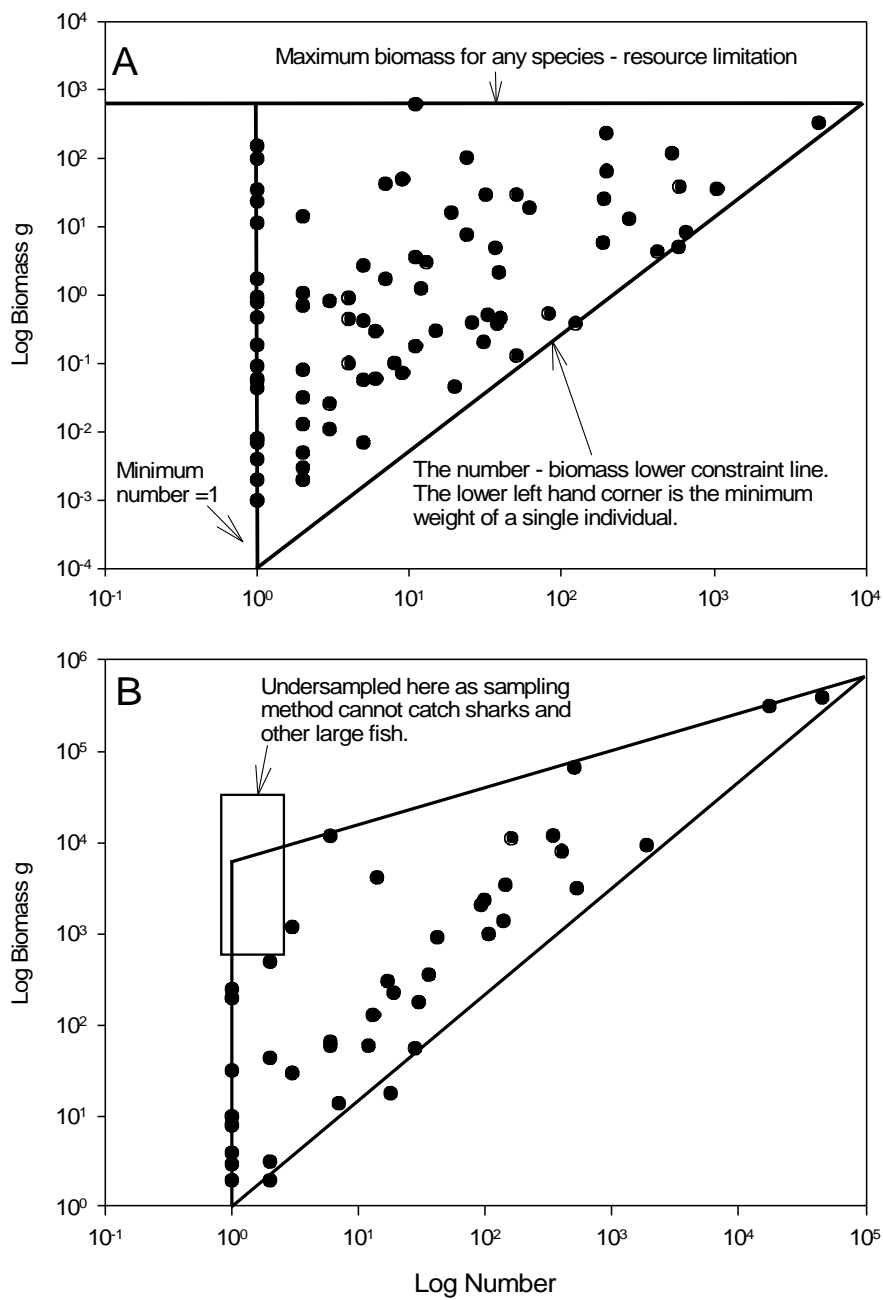


Figure 3 The relationship between biomass and numerical abundance for two well sampled communities. A. The Milford Haven sub-tidal benthic community sampled by grab. B The Bridgwater Bay fish community sampled using a nuclear power station cooling water intake.

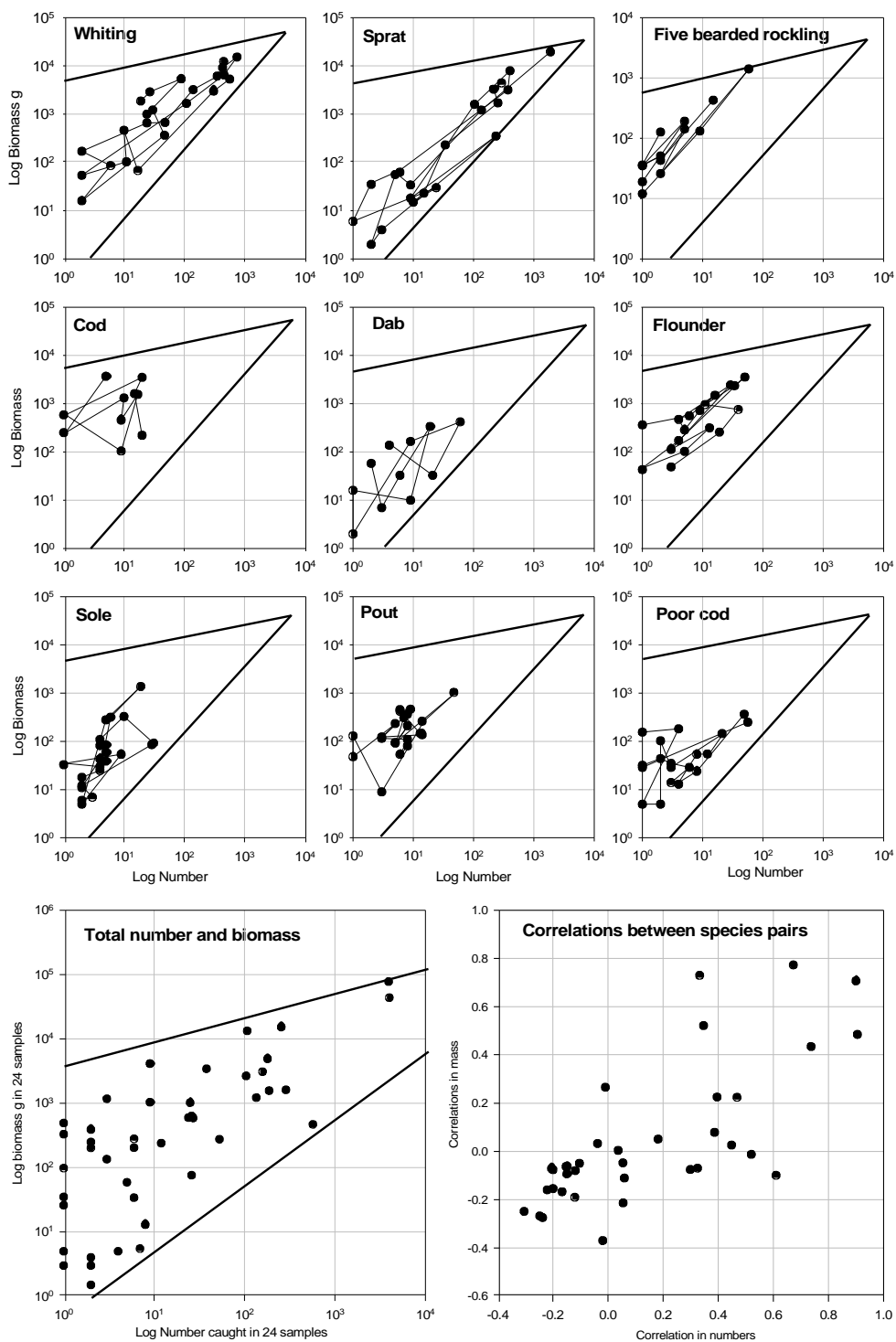


Figure 4 The changing relationship between biomass and numerical abundance for the common fish caught in Bridgwater Bay over 24 consecutive monthly

samples. The plots are for the most abundant species in terms of biomass or number. The upper line reflects the maximum biomass any species or individual can achieve in a sample. The largest individual observed in these samples was a conger eel of 2400 g

We have shown that it is possible to predict the shape of the SAD, and the equivalence of SADs based on different currencies, using some simple assumptions. Although ours is a null model in the sense that it defines the boundaries of the space that the species must occupy and then places them at random within in it, it is not neutral in the sense that it strips out biological differences. Indeed, it is becoming clear that SADs should not treat species as identical 19, 20. Our model draws on the observation that communities are composed of species that vary in body size. Sampling considerations, such as the decision to focus on a particular taxonomic group, or the use of selective gear such as plankton nets or light traps, may limit the size range of species included in a SAD 3, 21 To take an extreme example, although elephant, buffalo, termites and ants play major ecological roles in the African savannah community, large mammals and insects are never included in the same SAD. On a less extreme level, SADs for plants rarely encompass both trees and herbs. Our approach means that it should be possible to make predictions about the consequences that different sampling methods will have for the observed shape of the SAD.

The distribution of species within an approximately triangular region within the log biomass – log number space follows from a consideration of ecological constraints. The minimum biomass of an individual, and hence the minimum possible biomass of a species, may be determined by the physical constraints of the system – sediment particle size for example. Maximum biomass will be the upper limit attainable in that system for

a single life style, in other words the maximum extent of the resources that can be pre-empted by a single species. In virtually all empirical species abundance distributions based on numerical abundance singletons occupy the smallest class 3, but if the full community were censused, this class would be composed of species with the minimum viable population size. Similarly the upper limit will be the maximum observed population size. In practice the standard deviation of the distribution of numerical abundance is constrained – this is one of the reasons why lognormal distributions are often canonical 22. We make no assertion that communities are saturated for either biomass or numerical abundance, only that constraints exist. What we do say is that it is the number of orders of magnitude over which the two types of abundance measure are distributed, in conjunction with the range of body size, that will determine whether the species abundance distributions are equivalent.

There is an ongoing debate about the extent to which resource apportionment can be inferred from the distribution of numerical abundance. Some models, for instance Sugihara's 5 sequential breakage model and Hubbell's neutral model 23, are framed in terms of numbers of individuals, and assume that these distributions are indicative of the way that species divide resources. This assumption is supported by a number of empirical and modelling studies (eg 13, 14, 24). A second set of investigations, also drawing on data and allometric theory, indicate that numerical abundance and biomass can provide different insights into assemblage structure (eg 2, 12, 15, 25). Our results confirm that there is no single answer to the question, and show that while there are cases in which numerical abundance is a proxy of biomass (and hence resource use) it does not automatically follow. Our hypothesis that the relationship between the distributions of numerical abundance and biomass are mediated by body size is supported by research on desert rodents. Using a long-term data set, White et al. 15 conclude that changes in the distributions of numerical abundance biomass can result from changes in masses of the individual species – even if overall energetic availability

remains constant. They suggest that shifts in body size are linked to changes in the desert system itself, such as the transition from grassland to shrubland. The idea that there is an ecological space, bounded by the upper and lower limits of biomass and numerical abundance, and constrained by body size, has implications for species packing. Although the numerical abundance of desert rodents increased over time 15 this was matched by a decrease in biomass. Community energy use remained approximately constant and species richness did not vary. In the same way, increased species richness at lower latitudes, could be linked to the increase in the numerical abundance, biomass envelope in tropical systems relative to temperate ones.

Finally our approach will allow researchers to predict changes in the SADs of biomass and numerical abundance in impacted communities relative to undisturbed ones 20. The different responses of numerical abundance and biomass to pollution have already been noted 18. Our method shows how changes in the SADs come about and provides a framework in which the responses of individual species to disturbance can be understood.

Methods

The Milford Haven data set comprises 52 0.1 m² Day grabs collected from subtidal gravel sediment in Milford Haven in June 2007. The individual samples were passed through a 0.1 mm sieve and all the animals in the retained sample were picked under a lens. The individuals in each sample were identified to species and the blotted wet weight measured to the nearest 0.001 g.

For the Hinkley Fish intensive study 89 Hours of sampling from 4 intakes of Hinkley B was undertaken in December 1993. Samples were taken between December 1st and December 23rd. A total of 48,998 individual fish belonging to 35 species were caught. The standard lengths of individual fish were measured to the nearest mm and these were

converted to wet weight in grams using known length – weight relations for fish at Hinkley Point. For the change in species number and biomass through time, 24 monthly samples collected between 1987 and 1989 at Hinkley Point were used. Over these samples every individual fish were weighed to the nearest 1.0 g.

1. McGill, B. J. et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters* **10**, 995-1015 (2007).
2. Connolly, S. R., Hughes, T. P., Bellwood, D. R. & Karlson, R. H. Community structure of corals and reef fishes at multiple scales. *Science* **309**, 1363-1364 (2005).
3. Magurran, A. E. *Measuring biological diversity* (Blackwell Science, Oxford, 2004).
4. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 254-283 (1948).
5. Sugihara, G. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.* **116**, 770-787 (1980).
6. Motomura, I. On the statistical treatment of communities. *Zool. Mag., Tokyo* (in Japanese) **44**, 379-383 (1932).
7. Tokeshi, M. Species abundance patterns and community structure. *Adv. Ecol. Research* **24**, 112-186 (1993).
8. Saint-Germain, M. et al. Should biomass be considered more frequently as a currency in terrestrial arthropod community analysis. *Journal of Applied Ecology* **44**, 330-339 (2007).
9. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85** (2004).
10. Kleiber, M. *The fire of life* (Wiley, New York, 1962).

11. Thibault, K., White, E. & Ernest, S. K. M. Temporal dynamics in the structure and composition of a desert rodent community. *Ecology* **85**, 2649-2655 (2004).
12. Pagel, M. D., Harvey, P. H. & Godfray, H. C. J. Species-abundance, biomass and resource-use distributions. *Am. Nat.* **138**, 836-850 (1991).
13. Sugihara, G. How do species divide resources? *American Naturalist* **133**, 770-787 (1989).
14. Taper, M. L. & Marquet, P. A. How do species really divide resources? *American Naturalist* **147**, 1072-1068 (1996).
15. White, E. P., Ernest, S. K. M. & Thibault, K. Trade-offs in community properties through time in a desert rodent community. *American Naturalist* **164**, 670-676 (2004).
16. McGill, B. Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecology Letters* **6**, 766-733 (2003).
17. Magurran, A. E. & Henderson, P. A. Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**, 714-716 (2003).
18. Warwick, R. M. A new method for detecting pollution effects on marine macrobenthic communities. *Mar. Biol.* **92**, 557-562 (1986).
19. Alonso, D., Ostling, A. & Etienne, R. S. The implicit assumption of symmetry and the species abundance distribution. *Ecology Letters* in press (2007).
20. MacNally, R. Use of the abundance spectrum and relative-abundance distributions to analyze assemblage change in massively altered landscapes. *American Naturalist* **170**, 319-330 (2007).
21. Southwood, R. & Henderson, P. A. *Ecological methods* (Blackwell Science, Oxford, 2000).

22. Nee, S., Harvey, P. H. & Cotgreave, P. in *Mutalism and community organization: behavioral, theoretical and food-web approaches*. (eds. Kawanabe, H., Cohen, J. E. & Iwasaki, K.) 350-364 (Oxford University Press, Oxford, 1993).
23. Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography* (Princeton University Press, Princeton, 2001).
24. Marquet, P. A., Keymer, J. E. & Cofré, H. in *Macroecology: reconciling divergent perspectives on large scale ecological patterns* (eds. Gaston, K. J. & Blackburn, T. M.) 64-84 (Blackwell, Oxford, 2003).
25. Harvey, P. H. & Godfray, H. C. J. How species divide resources. *Am. Nat.* **129**, 318-320 (1987).

Supplementary Information accompanies the paper on

Acknowledgements We thank Richard Seaby and Robin Somes for assistance with fieldwork and Rowena Henderson for sample sorting and weighing. AM is a member of the NCEAS working group on ‘Tools and fresh approaches for species abundance distributions’ and thanks her colleagues there for many stimulating discussions on SADs.

Competing interests statement The authors declare that they have no competing financial interests. and requests for materials should be addressed to PAH (peter@irchouse.demon.co.uk).

Correspondence and requests for materials should be addressed to PAH (peter@irchouse.demon.co.uk).