

Explaining the excess of rare species in natural species abundance distributions

Anne E. Magurran*† & Peter A. Henderson†‡

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

‡ Pisces Conservation Ltd, IRC House, The Square, Pennington, Lymington, Hampshire SO41 8GN, UK

† The authors contributed equally to this work

The observation that a few species in ecological communities are exceptionally abundant, whereas most are rare, prompted the development of species abundance models^{1–3}. Nevertheless, despite the large literature on the commonness and rarity of species inspired by these pioneering studies, some widespread empirical patterns of species abundance resist easy explanation⁴. Notable among these is the observation⁵ that in large assemblages there are more rare species than the log normal model predicts^{6,7}. Here we use a long-term (21-year) data set, from an estuarine fish community, to show how an ecological community can be separated into two components. Core species, which are persistent, abundant and biologically associated with estuarine habitats, are log normally distributed. Occasional species occur infrequently in the record, are typically low in abundance and have different habitat requirements; they follow a log series distribution. These distributions are overlaid, producing the negative skew that characterizes real data sets.

Concern about rapid biodiversity loss has intensified the need to understand community structure. It is generally accepted that most distributions of species abundance in large assemblages tend towards the log normal^{13,7,8}. However, whereas the ‘canonical hypothesis’³ was the focus of debate in the past half-century^{8,9}, it is negative skew that now captures attention. Two neutral theories predict species abundance distributions that replicate the negative skew observed in empirical data sets. The unified theory of biodiversity and biogeography⁶ develops a new species abundance distribution, the zero-sum multinomial, in which the degree of negative skewness is a function of community size and immigration

rate. The self-similarity model¹⁰ predicts that the change in species richness with area is constant across all spatial scales. Both approaches assume the ecological equivalence of all species in the community. A process of multi-dimensional niche subdivision has been postulated⁹ that also leads to negative skew⁵. We argue that the observed patterns can be more parsimoniously explained by dividing an assemblage into two components—persistent and occasional species—and without invoking neutrality.

Here we examine the relative abundance of species in an exceptionally large data set—a 21-year investigation (with monthly sampling) of a fish community at Hinkley Point in the Bristol Channel, UK, in which $S = 80$ species and $N > 96,000$ individuals were recorded (see Methods). Our data clearly show that the maximum abundance of a species in the year in which it is most abundant is a function of the number of years for which it has been recorded (Fig. 1). The commonness and rarity of species in the assemblage is thus related to their permanence. Three species (sprat (*Sprattus sprattus*), sand goby (*Pomatoschistus minutus*) and whiting (*Merlangius merlangus*)) continuously dominate and together account for an average 70% of total abundance (by weight or by number). Adding the other 28 core species brings the total to 99%. The 49 infrequent species thus contribute only 1% of total abundance over the 21 years of the study. Importantly, the fish at Hinkley Point fall into two distinct groups: a core of persistent (more than 10 years in the record) and usually—but not invariably—abundant species and a set of occasional (less than 10 years in the record) and typically non-abundant species (Fig. 1). This allows us to decompose the empirical species abundance distribution (Fig. 1b) into two groups of species. The distribution of the occasional species, which are not continuously present in an assemblage but may sometimes breed there, follows a log series distribution (Fig. 1d)², whereas the persistent species are log normally distributed (Fig. 1c)^{3,8}. We employ three different methods to demonstrate that the division of species into two groups is not arbitrary.

First, we use a diversity statistic to reveal the shift from the log series to the log normal model as we work through the data set. The average value of Simpson’s diversity index, D , will remain approximately constant, once $S > 10$, if species are distributed according to the log series model⁸. However, if a log normal distribution pertains, D will increase with S . We therefore expect Simpson’s index to track the transition from log series to log normal as persistent species are progressively included in the analysis. Figure 2 confirms this. The switch is to the right of the natural break in our empirical data set

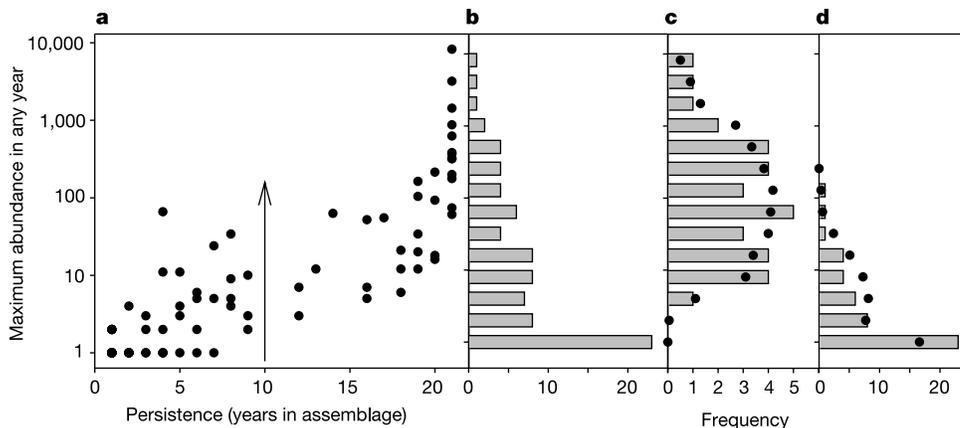


Figure 1 The pattern of abundance and persistence in the fish community of Hinkley Point, Bristol Channel. The data are for a 21-year time series of monthly samples. **a**, The number of years for which each fish was observed, plotted against the maximum abundance in any one year. A discontinuity (indicated by the vertical arrow) allows the core and occasional species to be defined as those present in >10 and <10 years.

b, The abundance distribution for all species. **c**, The abundance distribution of the core species; the frequency of each abundance class predicted with a log normal model is shown as a dot ($\chi^2_{[6]} = 0.88, P = 0.99$). **d**, The abundance of the occasional species; the frequency of each abundance class predicted with a log series model is shown as a dot ($\chi^2_{[4]} = 4.24, P = 0.39$).

because this analysis checks the signal of the underlying model as the proportion of core species increases. Thus, it is only when species that occur in 14 or more years are included that the log normal distribution drowns out the log series distribution created by the abundances of the less-persistent species.

Second, core and occasional species should be biologically distinguishable. We assigned species to habitat categories using descriptions in refs 11 and 12. Fish predominantly associated with muddy substratum, estuaries or with anadromous/catadromous life histories were allocated to the estuarine group; those that prefer rock, sand, gravel or weed substratum or are found in deeper water were placed in the non-estuarine group. Typical examples in the estuarine group are sprat, whiting, European seabass (*Dicentrarchus labrax*) and flounder (*Platichthys flesus*). Non-estuarine species include the pilchard (*Sardina pilchardus*), Ballan wrasse (*Labrus bergylta*) and 15-spined stickleback (*Spinachia spinachia*), which are typically found in deep water, rocky and weedy habitats respectively. A χ^2 test revealed a highly significant association between the biological and empirical allocation of species (estuarine core species, $n = 25$; estuarine occasional species, $n = 8$; non-estuarine core species, $n = 8$; non-estuarine occasional species, $n = 38$ (1 species not assigned); $\chi^2 = 26.9$, $P < 0.001$).

Last, the arrival of occasional species can be modelled as a stochastic event often related to unusual weather conditions such as storms or unusually settled and warm periods. However, not all arrivals are independent of each other, because particular climatic conditions will favour certain categories of fish. Overall, a Poisson process can approximate the pattern of arrival of each of these species. It was this process that Fisher invoked when he proposed his log series model. If a species follows a Poisson distribution, the variance to mean ratio will not be significantly different from 1.0 (ref. 13). Figure 3 demonstrates that this ratio successfully separates

the species into the persistent and occasional categories previously identified by using the empirical species abundance distribution. As long as a reasonable time series is available, this approach provides a tractable and objective solution to the difficult challenge of fitting a mixture of two distributions to species abundance data. Occasional species can be identified by using the variance/mean ratio, then assigned to the log series distribution and removed, and a log normal fitted to the remaining species distribution.

Although the precise location of the split point varies depending on the method used, each of these independent approaches is consistent in demonstrating that the species that comprise the empirical distribution fall into two categories. In practice, the exact position of the division is relatively unimportant from a model-fitting perspective because both the log series and the log normal are robust against moderate shifts in relative abundance^{14,15}. As the time series lengthens we expect the modes of the two distributions to move apart and the division zone to become broader and clearer. Snapshot surveys, which record species abundances at one point in time, obscure these complexities.

Southwood¹⁶ noted that terrestrial insect assemblages are continually challenged by a flow of migrants and suggested that the balance between transient and core species will determine whether a species abundance distribution is log normal or log series in character. Hubbell⁶ also concluded that the distribution of species abundance is a function of immigration rate. Our results extend these insights by showing that the occasional and persistent species leave different signatures on the species abundance distribution. However, we depart from Hubbell on the neutrality argument. As we have shown, there are clear differences in the ecological requirements of species in the two components of the distribution. Explanations of species abundance distributions cannot therefore be divorced from biology^{4,7,9}. Species are also lost over longer time scales: in our study area the eel (*Anguilla anguilla*) is currently declining and might shortly become an occasional visitor rather than a member of the core community. These longer-term changes might explain why the log normal distribution of the core species retains some negative skew.

Brown *et al.*¹⁷ contend that systems open to colonization are better able to withstand natural or human induced perturbations because immigrants can replace missing species and tolerate changed conditions. This is why islands, with their reduced flow

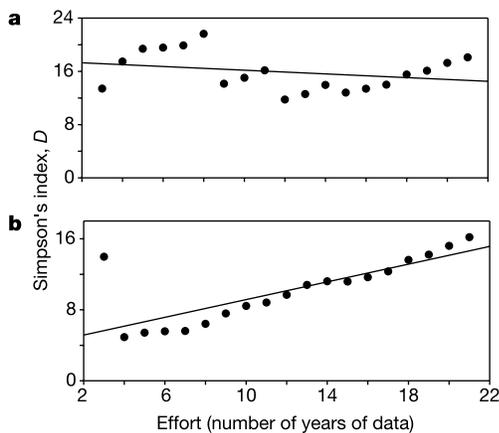


Figure 2 Switch from log series to log normal model revealed by Simpson's index. We calculated the value of Simpson's index, D (reciprocal form¹⁴), in relation to increasing effort (number of years of data) for 10 progressively larger subsets of the fauna. These were chosen to span the transition between the occasional and core species identified by the empirical species abundance distribution. **a**, D remained constant with effort when calculated for all species found in 13 or fewer years in the record ($F_{1,16} = 2.98$, slope, $b = -0.212$, $P > 0.1$). This is consistent with a log series distribution⁸. **b**, When the analysis was run with species present in 14 or fewer years, a strong positive relationship, indicative of a log normal model⁹, emerged ($F_{1,16} = 1105.9$, $P < 0.001$, $b = 0.66$), confirming the shift in species abundance distribution. Analyses that included species with longer or shorter representation in the time series gave consistent results (≤ 6 years: $F_{1,16} = 4.85$, $P = 0.04$, $b = -0.315$; ≤ 7 years: $F_{1,16} = 3.78$, $P = 0.07$, $b = -0.266$; ≤ 8 years: $F_{1,16} = 3.75$, $P = 0.07$, $b = -0.193$; ≤ 9 years: $F_{1,16} = 3.34$, $P = 0.09$, $b = -0.217$; ≤ 12 years: $F_{1,16} = 3.61$, $P = 0.07$, $b = -0.229$; ≤ 16 years: $F_{1,16} = 171.2$, $P < 0.001$, $b = 0.40$; ≤ 17 years: $F_{1,16} = 326.8$, $P < 0.001$, $b = 0.40$; ≤ 18 years: $F_{1,16} = 473.2$, $P < 0.001$, $b = 0.49$).

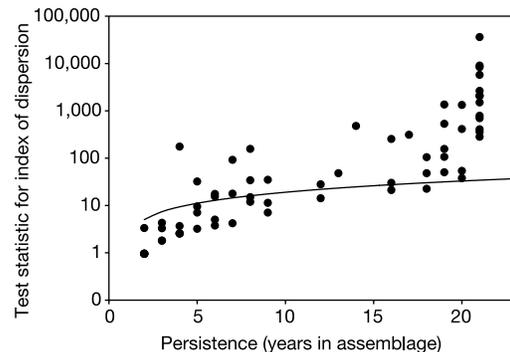


Figure 3 Using the variance/mean ratio to decompose the distribution. In accordance with ref. 13 we calculated, for each species, an index of dispersion based on the ratio of the variance to the mean. This is plotted against the number of years for which the species was present in the assemblage. The line shows the 2.5% confidence limit for the χ^2 distribution: species that fall below this line follow a Poisson distribution, and are randomly dispersed in time. Those that lie above the line are not. The 14 species that occurred in a single year are excluded from this analysis because their dispersion in time has no variance. However, these species would, by Fisher's logic, be assigned to the log series distribution. Nine species in our occasional grouping were mis-classified by this analysis. These species were found in clusters of years and their incidence increases close to our empirical split point. Only three core species are Poisson distributed.

of colonists, are especially vulnerable to environmental change. Marine and estuarine assemblages are already affected by climate change; this data set tracks the recent expansion of range and population size in fish that are close to the northern limit of their range in British waters such as sole (*Solea solea*¹⁸), seabass¹⁹ and trigger fish (*Balistes carolinensis*). Conversely, it also shows the decline in species that reach the southern limit of their range in British waters, for example sea snail (*Liparis liparis*²⁰) and dab (*Limanda limanda*²¹). If conditions alter sufficiently we predict that new core species, drawn from the pool of occasional species, will replace the existing ones. Temporal components of species abundance also have important implications for conservation planning, as recognized by the latest generation of reserve selection algorithms, which incorporate information on species permanence²². □

Methods

Fish samples were collected from the cooling-water filter screens at Hinkley Point B Nuclear Power Station, situated on the southern bank of the Bristol Channel in Somerset, England. The power station intakes are placed in front of a rocky promontory within Bridgwater Bay; to the east are the extensive Stert mud flats with an intertidal area of ~40 km². The water intakes are placed between -1 and -5 m MLWS (mean low water springs), so the fish are sampled from water of between 8 and 18 m depth. Full descriptions of the intake configuration and sampling methodology are given in refs 18 and 23. Quantitative sampling began in 1980 when 24-h surveys of the diurnal pattern of capture were undertaken in October and November. From these surveys it was concluded that samples collected during daylight were representative of the 24-h catch²⁴, and monthly quantitative sampling began in January 1981. The total volume of water sampled per month, which has not varied over the entire 21-year period, is 3.24×10^5 m³. To standardize for tidal influence, all sampling dates are chosen for tides halfway between springs and neaps, with sampling starting at high water (normally about 12:00). Fish are collected hourly from two filter screens for a 6-h period, identified to species, measured and the number of individuals recorded. Since 1987 the standard lengths (SLs) of all captured fish have been recorded to the nearest millimetre. The filter screens have a solid square mesh of 10 mm and start to retain fish >25 mm SL. A 100% retention for many species occurs at SLs > 40 mm. For fish such as sprat, whiting and pout the screens retain all fish captured with a SL greater than ~60 mm (ref. 25). The sampling method therefore catches adults and juveniles older than 6 months for all known British marine fish.

Received 25 October 2002; accepted 7 March 2003; doi:10.1038/nature01547.

- Motomura, I. On the statistical treatment of communities [in Japanese]. *Zool. Mag., Tokyo* **44**, 379–383 (1932).
- 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).
- Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 254–283 (1948).
- Brown, J. H. Towards a general theory of biodiversity. *Evolution* **55**, 2137–2138 (2001).
- Nee, S., Harvey, P. H. & May, R. M. Lifting the veil on abundance patterns. *Proc. R. Soc. Lond. B* **243**, 161–163 (1991).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 2001).
- Gaston, K. J. & Blackburn, T. M. *Macroecology* (Blackwell Science, Oxford, 2000).
- May, R. M. In *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 81–120 (Harvard Univ. Press, Cambridge, Massachusetts, 1975).
- Sugihara, G. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.* **116**, 770–787 (1980).
- Harte, J., Kinzig, A. & Green, J. Self-similarity in the distribution and abundance of species. *Science* **284**, 334–336 (1999).
- Lythgoe, J. & Lythgoe, G. *Fishes of the Sea* (Blandford, London, 1971).
- Wheeler, A. *The Fishes of the British Isles* (Macmillan, London, 1969).
- Krebs, C. J. *Ecological Methodology*, 2nd edn (Harper & Row, New York, 1999).
- Magurran, A. E. *Ecological Diversity and its Measurement* (Princeton Univ. Press, Princeton, New Jersey, 1988).
- Southwood, R. & Henderson, P. A. *Ecological Methods* (Blackwell Science, Oxford, 2000).
- Southwood, T. R. E. The Croonian Lecture 1995. Natural communities: structure and dynamics. *Phil. Trans. R. Soc. Lond. B* **351**, 1113–1129 (1996).
- Brown, J. H., Whitham, T. G., Ernest, S. K. M. & Gehring, C. A. Complex species interactions and the dynamics of ecological systems: long term experiments. *Science* **293**, 643–650 (2001).
- Henderson, P. A. & Seaby, R. M. H. On the factors influencing juvenile flatfish abundance in the lower Severn Estuary. *Neth. J. Sea Res.* **32**, 321–330 (1994).
- Henderson, P. A. & Corps, M. The role of temperature and cannibalism in interannual recruitment variation of bass in British waters. *J. Fish Biol.* **50**, 280–295 (1997).
- Henderson, P. A. & Seaby, R. M. H. Population stability of the sea snail at the southern edge of its range. *J. Fish Biol.* **54**, 1161–1176 (1999).
- Henderson, P. A. On variation in the dab, *Limanda limanda*, recruitment: a zoogeographic study. *J. Sea Res.* **40**, 131–142 (1988).
- Rodrigues, A. S. L., Gaston, K. J. & Gregory, R. D. Using presence-absence data to establish reserve selection procedures that are robust to temporal species turnover. *Proc. R. Soc. Lond. B* **267**, 897–902 (2000).
- Henderson, P. A. & Holmes, R. H. A. On the population dynamics of dab, sole and flounder within Bridgwater bay in the lower Severn Estuary, England. *Neth. J. Sea Res.* **27**, 337–344 (1991).

- Henderson, P. A. & Holmes, R. H. A. Population stability over a ten-year period in the short-lived fish *Liparis liparis* (L.). *J. Fish Biol.* **37**, 605–616 (1990).
- Turnpenny, A. W. H. *The use of multiple regression analysis in formulating operating strategies for power impingement problems* (CEGB publication RD/L/2043N81, Central Electricity Generating Board, Leatherhead, 1981).

Acknowledgements We thank R. Seaby and R. Somes for assistance with fieldwork.

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

Correspondence and requests for materials should be addressed to A.E.M. (e-mail: aem1@st-andrews.ac.uk). The full data set can be downloaded from the Pisces Conservation Ltd website (<http://www.irchouse.demon.co.uk/latestreports.html>).

Decisions about parental care in response to perceived paternity

Bryan D. Neff

Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada

Evolutionary ecologists are attempting to explain how parents make behavioural decisions about how much care to provide to their young^{1–4}. Theory predicts that when genetic relatedness to young is decreased by cuckoldry, for example, parents should reduce their care in favour of alternative broods that provide greater reproductive success^{5–7}. Experimental manipulation of perceived paternity has been used to test the theory^{8,9}, but such studies have generated mixed results^{10–13}. Some manipulations can fail to alter a parent's perceived paternity¹⁴, whereas others may directly affect parental behaviour when, for instance, the manipulation involves capturing the parent^{15–18}. No study has demonstrated parental care adjustment in a manner uncomplicated by experimental design or life history correlates. Here I test the theory using the fact that nest-tending parental male bluegill sunfish (*Lepomis macrochirus*) can assess their paternity using both the visual presence of parasitic cuckolder males during spawning¹⁹, and olfactory cues released by newly hatched eggs^{20,21}. By manipulating both types of cues I show that parental males dynamically adjust their parental care, favouring broods that are apparently most closely related. These results confirm the importance of genetic relatedness in parental care decision-making.

Bluegill are native to lakes and rivers of North America. Males are characterized by a discrete polymorphism in life histories termed 'parental' and 'cuckolder'^{22,23}. In Lake Opinicon (Ontario, Canada), parentals mature at age 7 years and construct nests in a colony during the breeding season²³. Nesting parentals court and spawn with multiple females over the course of a single day and then provide sole care for the developing young in their nests. Parental care involves fanning and defending eggs (which can number in the tens of thousands) until they hatch (2–3 days), and then defending developing fry from predators until the young leave the nest (5–7 days). If a parental abandons his nest before this time, the young do not survive. Parentals do not forage while they are tending their brood and so they lose about ten per cent of their body weight, although they do occasionally cannibalize some of the eggs or fry in their nest²⁴. After the fry have dispersed, parentals return to deeper waters to feed and replenish their energy reserves before re-nesting in a subsequent spawning bout²⁵.

In contrast, cuckolders do not build nests or care for young but instead mature precociously and steal fertilizations in the nests of