

The statistical mechanics of community assembly and species distribution

Colleen K. Kelly¹, Stephen J. Blundell², Michael G. Bowler², Gordon A. Fox³, Paul H. Harvey¹, Mark R. Lomas⁴ and F. Ian Woodward⁴

¹Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; ²Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK; ³Department of Biology, University of South Florida, 4202 E. Fowler Ave, Tampa, FL 33620, USA; ⁴Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Summary

Author for correspondence:

Colleen Kelly

Tel: +44 1874 625226

Email: colleen.kelly@zoo.ox.ac.uk

Received: 25 January 2011

Accepted: 8 March 2011

New Phytologist (2011) **191**: 819–827

doi: 10.1111/j.1469-8137.2011.03721.x

Key words: biotic resistance, diversity and productivity, idiosyncrasy, naturalized species, *n*-dimensional hypervolume, neutrality, niche dynamics.

- Theoretically, communities at or near their equilibrium species number resist entry of new species. Such 'biotic resistance' recently has been questioned because of successful entry of alien species into diverse natural communities.
- Data on 10 409 naturalizations of 5350 plant species over 16 sites dispersed globally show exponential distributions both for species over sites and for sites over number of species shared. These exponentials signal a statistical mechanics of species distribution, assuming two conditions. First, species and sites are equivalent, either identical ('neutral') or so complex that the chance a species is in the right place at the right time is vanishingly small ('idiosyncratic'); the range of species and sites in our data disallows a neutral explanation. Secondly, the total number of naturalizations is fixed in any era by a 'regulator'.
- Previous correlation of species naturalization rates with net primary productivity over time suggests that the regulator is related to productivity.
- We conclude that biotic resistance is a moving ceiling, with resistance controlled by productivity. The general observation that the majority of species occur naturally at only a few sites, and only a few species occur at many sites, now has a quantitative (exponential) character, offering the study of species' distributions a previously unavailable rigor.

Introduction

The effects, accidental or otherwise, that humans may have on natural systems are a classic source of insight into the fundamental processes governing those systems (Darwin, 1859; Elton, 1958). Here we use distributions of plant species' naturalization to characterize the factors determining entry of new species into a standing species complement, the fundamental building block of natural communities ('naturalized' denotes merely having an established population; 'invasive' (pest) species are a subset of naturalized species, but not all naturalized species become pests). Ecological theory of community assembly predicts that mature communities – those at or near their equilibrium species number – will resist the entry of new species. Such 'biotic resistance' is proposed to occur either through *in situ* coevolution filling all available niche space, or by ecological sorting to find the combination of species best able to exploit available resources. The resulting complex matrix of

interactions is supposed to leave little niche space in the existing community into which a newcomer may easily insert itself, thus regulating community diversity (Elton, 1958; Hutchinson, 1959; MacArthur, 1965; May, 1973; Pimm, 1991; Tilman, 2004).

Biotic resistance has been interpreted on a practical level to mean that highly diverse communities are protected from invasion by species not currently a part of the community, and small-scale manipulations under natural conditions largely support this expectation (Levine, 2000; Kennedy *et al.*, 2002). However, two patterns of species' naturalization at greater geographic scales and incorporating longer time-spans seem to contradict these observations: first, regional inventories of species occurrences show highly diverse communities readily invaded by naturalized species (Lonsdale, 1999; Stohlgren *et al.*, 1999; Sax, 2002); secondly, successful naturalizations are not being offset by the concomitant extinctions of native species that would be expected if niche filling regulates community assembly (Sax & Gaines, 2003).

These observations challenge the idea that complex interactions regulate the successful entry of new species into natural communities, and pose the question as to what, then, determines the entry of a species into a community.

In this study we exploit the global-scale 'natural experiment' created by the escalation of species naturalizations over the last century. We employ these data to examine the large-scale patterns of species naturalizations and community assembly through the high-power lens of statistical mechanics. Statistical mechanics uses probability theory to provide a framework relating the properties of large numbers of individual units to the bulk properties of the whole, revealing emergent properties that give an insight into the regulators of individual behaviour not available from considering one or a few individuals independently. Statistical mechanics underlies much of the realm of the physical sciences, but also has been useful in problems such as the distribution of wealth (Drăgulescu & Yakovenko, 2001) and the ubiquitous lognormal distribution of individuals among species in ecological communities (Pueyo, 2006; Pueyo *et al.*, 2007; Dewar & Porté, 2008; Harte *et al.*, 2008; Bowler & Kelly, 2010).

On collating for other purposes a data set comprising all naturalized species over 16 sites dispersed globally (Woodward & Kelly, 2008), a family of exponential distributions fell naturally out of the data. The results of our approach to these patterns produces new insights into several fundamental ecological processes. First, we have devised an analytical explanation of community assembly able to incorporate naturally all the above observations. From this, we are able to conclude that biotic resistance exists, but as a moving ceiling regulated by some external factor; combining these findings with the earlier work (Woodward & Kelly, 2008), we infer that external factor to be net primary productivity (NPP) or some process innately linked to NPP. Secondly, we have identified a quantitative (exponential) character to the general observation that the majority of species are of restricted distribution and only a few are widespread. In our model, this pattern is an emergent property deriving from the fundamental nature of niches themselves and does not require the operation of any particular trait of any particular niche. Lastly, the simple exponential distributions make possible analytical tools carrying with them a degree of rigor not previously available to the comparative study of species' distributions (see Gotelli *et al.*, 2009).

Materials and Methods

We collated data on 10 409 naturalizations of 5350 unique plant species over 16 sites dispersed globally, determining the number of sites at which each unique species occurred. We also recorded the number of species in common between sites, grouping sites first into all possible pairwise

combinations, next into all possible triplet combinations, and finally into all possible quadruplet combinations.

Because species naturalization is largely tabulated at the country scale, our study is at this scale. Site selection was dictated by the availability of naturalized species lists including all known established alien pteridophyte, gymnosperm and angiosperm species, and not restricted to invasive pests. The 16 sites meeting these criteria and included in the study are: Chile (Castro *et al.*, 2005), the Czech Republic (Pysek *et al.*, 2002), Estonia (Anonymous, 2007c), Galapagos (Tye, 2001), Hawai'i (Wester, 1992), Israel (Dafni & Heller, 1990), Japan (Anonymous, 2007a), Latvia (Anonymous, 2007b), New Zealand (Healy & Edgar, 1980; Webb *et al.*, 1988; Edgar & Connor, 1999), Poland (Bobel & Król, 2007), Singapore (Corlett, 1988), Swaziland (Braun), Switzerland (Wittenberg, 2005), Taiwan (Wu *et al.*, 2004), the UK (Preston *et al.*, 2006), and Wyoming (Rice, 2007). Subspecies were subsumed under the name of their parent species in determining the number of unique species.

In order to investigate the possible effect of dispersal on the observed distributions, we performed a Mantel test of correlation between geographical distance and number of species shared between pairs of sites (Table 1) using the R-package MANTEL module (with 9999 permutations) (Casgrain & Legendre, 2001).

Results

Three important properties were revealed by our treatment of the data. First, species naturalizations show an exponential distribution of the number of naturalized species $S(n)$ found at n sites (Fig. 1). To correspond to the analyses illustrated in Figs 2 and 3, the exponential is fitted to $n \geq 2$ using maximum likelihood; the relationship is $S(n+1) = 0.59S(n)$ or $S(n) = S_0 e^{-0.52n}$ where the fitted parameter S_0 is 2343 and the coefficient in the exponent (-0.52) is uniquely related to the number of naturalized species summed over sites, which we term 'the alien footprint', defined by $M_1 = \sum nS(n)$ (Supporting Information Notes S2).

Secondly, there is no correlation between the number of naturalized species common to a pair of sites and the separation of those sites (Fig. 2). The comparison of matrices of distance between two sites and the number of species shared pairwise showed no relationship between the two factors ($P > 0.22$). Some correlation might be present for distances ≤ 5000 km, but, if so, it is not sufficient to affect the overall conclusion that, at the global scale, the proportion of sites sharing a large number of species does not depend on distance. The number of shared species for each site pairing is given in Table 1. What species are shared between which sites may depend on many particularities; overall, there is no correlation with separation (Notes S1).

Table 1 Number of species shared between sites (authorities given in the Materials and Methods section)

	Chile	Czech Republic	Estonia	Galapagos	Hawaii	Israel	Japan	Latvia	New Zealand	Poland	Singapore	Swaziland	Switzerland	Taiwan	UK
Chile	0														
Czech Republic	225	0													
Estonia	100	372	0												
Galapagos	22	47	21	0											
Hawai'i	128	147	62	99	0										
Israel	28	44	24	19	41	0									
Japan	248	413	211	68	25	52	0								
Latvia	75	299	340	15	47	21	162	0							
New Zealand	294	442	256	55	264	42	479	199	0						
Poland	33	188	142	9	28	12	78	127	94	0					
Singapore	7	10	3	27	57	10	61	3	18	3	0				
Swaziland	43	54	30	48	94	28	89	23	93	14	17	0			
Switzerland	48	189	118	13	43	16	114	98	146	78	0	0			
Taiwan	59	81	40	57	158	31	197	24	103	17	54	62	23	0	
UK	134	487	295	34	109	35	281	207	437	129	4	56	168	46	0
Wyoming	103	147	106	8	68	7	160	94	164	27	1	15	34	35	76

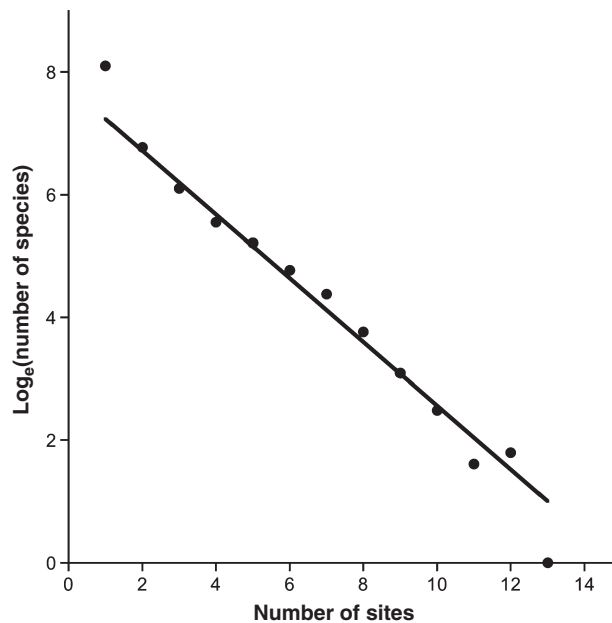


Fig. 1 Number of species as a function of number of sites. The number of naturalized species $S(n)$ falls exponentially with the number of sites n at which each is found. To correspond to the analyses illustrated in Figs 2 and 3, the exponential is fitted to $n \geq 2$ using maximum likelihood, with goodness of fit assessed using the appropriate one-sample δ -corrected Kolmogorov–Smirnov analysis ($P > 0.20$; Khamis, 2000).

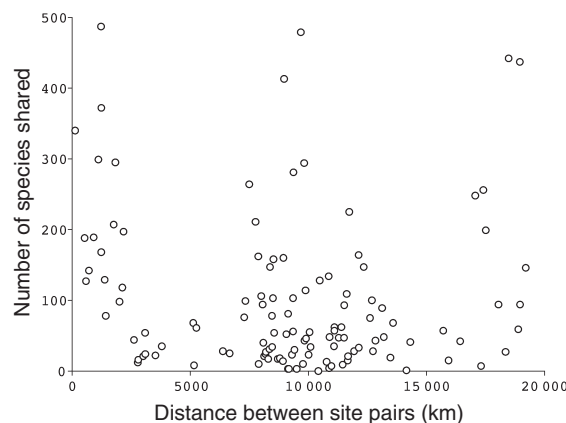


Fig. 2 Number of species shared pairwise between sites relative to distance between sites. Distance between two sites compared with number of species shared pairwise shows no relationship between the two factors. Some correlation might be present for distances ≤ 5000 km, but the main conclusion that, at the global scale, the proportion of sites sharing a large number of species does not depend on distance is unaffected by this possibility. The number of shared species for each site pairing is given in Table 1.

Finally, the number of pairs of sites sharing a given number of naturalized species falls exponentially with the number of shared species (Fig. 3a). The observed distribution is an exponential $y = 60e^{-0.01x}$, fitted to the individual values in Fig. 2 using maximum likelihood, with goodness

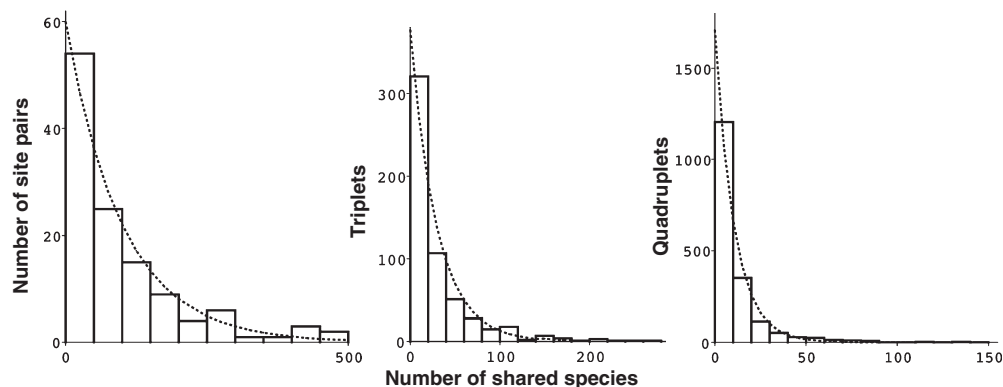


Fig. 3 Distribution of groups of sites over shared species. In (a), the y-axis represents the number of pairs of sites sharing the number of species counted along the x-axis. The data are binned and the exponential is calculated from $S(n)$. (b) and (c) are similarly for triplets and quadruplets of sites. The ubiquity of exponentials at every level of site grouping corroborates the robustness of our findings over alternative explanations of species shared between sites (details in Supporting Information Notes S1: Dispersal; Fig. S1).

of fit assessed using the appropriate one-sample δ -corrected Kolmogorov–Smirnov analysis ($P > 0.20$; Khamis, 2000). As we show in our discussion of the distribution of sites over species, it is also predicted from $S(n)$, assuming only that the distribution is exponential. In particular, the coefficient in the exponent is given by the number of pairs (120) divided by what we refer to as the ‘overlap measure’ $M_2 = \sum n(n-1)S(n)/2$. The exponentials illustrated in Fig. 3(b,c) are similarly predicted.

The model

No reasonable model of exponential attenuation of propagules spreading stepwise (through dispersal and transport) appears able to produce the exponential distribution of Fig. 1; even extremely contrived models are at odds with the lack of distance correlations shown in Fig. 2 (see Notes S1: Dispersal; Fig. S1). There is, however, an explanation that includes naturally the complexity of the biological world and this lack of correlation: the statistical behaviour of complex systems involving large numbers of components yields exponential distributions of the kind observed in Figs 1 and 3. Such systems function subject to certain constraints, in this case of biological or environmental origin. The techniques of statistical mechanics are mostly employed in the physical sciences but have also found some applications in ecology (Pueyo, 2006; Shipley *et al.*, 2006; Pueyo *et al.*, 2007; Dewar & Porté, 2008; Harte *et al.*, 2008). Ecologists have long been familiar with the attempt by MacArthur (1957, 1960) to account for species abundance distributions in terms of a statistical model known as the ‘broken stick’. MacArthur postulated that a finite resource (the stick) is partitioned at random into a given number of pieces, taken to represent species with the abundances given by the lengths. This postulate leads to an exponential distribution of species abundance as the most probable configuration subject to those constraints and can

be obtained using just the techniques we discuss following Eqn 1. The model is not correct for most species abundance distributions but does serve as a model for the distribution of alien species over the sites at which they are naturalized, a very different ecological problem.

The logical structure of our investigation is that we started with the hypothesis that a simple argument in statistical mechanics accounts for the observed exponential distribution of species over sites. We identified the necessary general conditions and constraints and found them to account also for the exponential distribution of pairs of sites over numbers of species held in common (Fig. 3a). We were then able to predict successfully the exponential distributions shown in Fig. 3(b,c), further supporting our hypothesis of the nature of our original observation.

Suppose we have S objects (of as yet unspecified nature) assigned to classes such that the class labelled n contains s_n objects. The number of ways of arranging S objects over the different classes so as to achieve a configuration $\{s_n\}$, characterized by numbers in each class $s_1, s_2, \dots, s_m, \dots$ is simply

$$W = \frac{S!}{\prod s_n!}, \quad \text{Eqn 1}$$

(where \prod denotes the continued product.)

The quantity W is proportional to the probability of finding this configuration $\{s_n\}$, provided that each arrangement has equal weight; without further conditions, every object has the same probability of being found in every class. If this is not true, an additional weight factor can be introduced (a ‘prior’) and the form of that prior is determined by the nature of the problem to be addressed (Bowler & Kelly, 2010; Haegeman & Etienne, 2010).

MacArthur’s broken stick To match the mathematics of Eqn 1 to reality, it is necessary to specify the nature of the S objects and the classes labelled by n . In MacArthur’s broken

stick, the objects are species and the class labelled by n is the class of all species with population n individuals (see also Pueyo *et al.*, 2007). The environmental constraints to be applied are first that there are a given number of species S (the number of pieces into which the stick is to be broken) and secondly that $\sum ns_n$ is fixed – this is the length of the stick; the sum of all the pieces is equal to the original length. The most probable of the configurations $\{s_n\}$ is found by maximizing Eqn 1 subject to the constraints – an operation which is mathematically well defined – where the constraints are on the number of species and available resources to be subdivided. The solution is

$$s_n = s_0 \exp(-\beta n) \quad \text{Eqn 2}$$

(the parameters of this exponential are determined by the values of the constraints; see Notes S2). If the stick is broken randomly then the distribution of species with population n as a function of n is exponential, provided that Eqn 1 contains the essential underlying biology and the constraints are the only ones that matter in this problem. For most guilds, Eqn 2 is not an acceptable species abundance distribution (Rosenzweig, 1995). The above conditions are not sufficient for this problem and indeed the papers by Pueyo *et al.* (2007), Dewar & Porté (2008) and Harte *et al.* (2008) are attempts to identify additional assumptions or constraints beyond those of MacArthur in order to produce a log series distribution, and the biological nature of such additions (Notes S4).

The distribution of species over sites In our problem of the distribution of naturalized species over a number of sites, the objects in Eqn 1 are naturalized species and the classes are defined by the number of sites at which a species is to be found. Thus here we identify s_m general in Eqn 1, with the number of species found at n sites, $S(n)$. The most probable configuration $\{s_n\}$ is obtained by maximizing Eqn 1 with respect to all s_m subject to conditions dictated by the nature of this problem. The first is that a given number of species S is involved and the second is that the sum $\sum nS(n) = M_1$ is fixed – this is the analogue of the length of MacArthur's stick and it is the total number of alien establishments over the 16 sites available to us. It is of such importance that we have given it a name; the 'alien footprint' introduced earlier. When Eqn 1 is maximized with respect to all $S(n)$ subject to these constraints, the most probable distribution of species over the number of sites n at which they are found is given by

$$S(n) = S_0 \exp(-\beta n). \quad \text{Eqn 3}$$

The mathematical constraint on the number of naturalized establishments $M_1 = \sum nS(n)$, which is also the sum of site diversities, implies a biological constraint. The rate at

which the exponential decreases is controlled by the mean alien footprint $\bar{n} = \sum s_n n / S = M_1 / S$, the number of sites reached averaged over all species, and in Eqn 3 the value of β is determined by the value of $1/\bar{n} (\beta = -\ln(1 - 1/(\bar{n} - 1)))$; see Fig. 1. The constants S_0 and β are obtained from Eqn 3 by evaluating the sums over $S(n)$ and over $nS(n)$. This is discussed in greater detail in Notes S2.)

Thus the observed exponential in Fig. 1 is reproduced by two ecological assumptions. The first assumption is that the alien footprint has a fixed value (it is a conserved quantity) and the value is determined by ecological constraints. The second assumption is that the nature of the world is such that Eqn 1 is indeed proportional to the probability of finding some specified configuration; there is a sense in which species are equivalent. Any other ecological forces then do not affect this distribution.

The distribution of sites over species Fig. 3(a) shows the distribution of the number of pairs of sites over the number of naturalized species common to both and is again an exponential. This distribution follows from the subsidiary ecological assumption that Eqn 1 is proportional to the probability of finding some specified configuration of pairs (s_n in Eqn 1) over n species in common; that there is a sense in which sites are equivalent. A constraint equivalent to the length of MacArthur's stick is wholly determined by the exponential distribution $S(n)$ for $S(n)$ species over n sites; it is that the sum of the number of pairs with m alien species in common multiplied by that number m of common species is constrained. This sum is easily evaluated. Pairs can be chosen from a set of n sites in $n(n-1)/2$ different ways and so a species found at n sites will also be found at $n(n-1)/2$ pairs of sites. Summing over all n yields a total overlap measure M_2 given by $M_2 = \sum n(n-1)S(n)/2$ which counts up all pairs of sites and sums the number of common species over all pairs and is thus the required sum. Because $S(n)$ is already determined, M_2 is fixed, essentially by the same biological constraints that limit M_1 . The coefficient in the exponential in Fig. 3(a) is given by the number of pairs (120) divided by the overlap measure M_2 .

Our hypotheses now allow us to predict that the distribution of triplets of sites over the number of species common to all three will also be exponential and with parameters given by the numbers $S(n)$ for the distribution of species over the number of sites. The quantity $M_3 = \sum n(n-1)(n-2)S(n)/6$ is the analogue of M_2 and is again fixed; we obtain an exponential distribution with coefficient equal to the number of triplets (560) divided by M_3 . This is shown in Fig. 3(b), together with the exponential for quadruplets, with a coefficient given by the number of quadruplets (1820) divided by the analogous moment M_4 , in Fig. 3(c). The calculated exponentials are again in agreement with the data and support our hypotheses.

Finally, consider the distribution of single sites over the number of naturalized species. The mean number of species per site is given by $M_1/N(\sum nS(n)/N)$, the alien footprint divided by the number of sites, and this singlet distribution is also exponential under the statistical assumptions. With only 16 sites the distribution is not very well defined by the data, but maximum likelihood and a Kolmogorov–Smirnov test (Khamis, 2000) show them to be consistent with being drawn from the exponential ($P > 0.2$).

The distributions of sites over species and of various multiplets of sites holding species in common all contain information. The exponentials observed show that in every case the distribution corresponds to the maximum amount of missing information (the distributions most likely to be encountered) after fixing the alien footprint.

The ecological implications of the success of our assumptions in reproducing the observed distributions are first that the overlap measures M_n are fixed (already ensured by the form of the distribution of species over sites) and secondly that sites are (without constraints) indifferent to the classes of the number of species, that pairs of sites are similarly indifferent to the classes of the number of species held in common, and so on; Eqn 1 is applicable to all these classes. Thus, within our model, these distributions imply that sites are in some sense equivalent, just as the distribution of species over sites implies an equivalence of species.

Discussion

While it is widely observed that, in nature, species are generally restricted in distribution and relatively few species are widely distributed (Pielou, 1979; Brown, 1995; Gaston, 2003), no definitive quantitative pattern of species distribution has previously been revealed (Gotelli *et al.*, 2009). The number of species $S(n)$ at n sites might fall with n in many different ways, and the exponential observed here is new information concerning the ecological processes responsible.

The distribution of species naturalizations contains an analogue of the mean energy term kT in the theory of gases in the mean alien footprint, the mean number of sites reached over all species. With a fixed number of naturalized species, increasing the mean number of sites per species (\bar{n}) dictates an increase in the average number of species per site; for a given number of species distributed among a given number of sites, the sum of sites over naturalized species is equal to the sum of species over sites. For naturalized species we suggest that the determinant of this fixed number of alien establishments, an ecological analogue of thermodynamic temperature T , may be associated with productivity, in the light of the relationship between plant species naturalization rate and increasing NPP over time (Woodward & Kelly, 2008). An increase in productivity would then increase this 'ecological temperature', to produce

a new most probable exponential in which species are found at more sites, and more species are found per site. This would be so regardless of whether the total number of naturalized species increases or not (where the number of species does not increase, the analogue is heating a box of gas from outside; where the number of species increases, the analogue is an injection of hotter gas into the box).

The fixed nature of the number of naturalizations does not imply that no further naturalizations are possible; such an extreme interpretation is not necessary. The model is not likely to be perfect and the world is not likely to be in equilibrium. There are also stochastic effects with a sample of only 16 sites. Finally, we envisage the lid on the total number of naturalizations being raised as global climate changes.

The relevant point is that the dynamic 'relaxes' rapidly into the (quasi) equilibrium configuration, achieving a new maximum number of species within the time scale over which changes in productivity occur. Evidence for this may be found in the observed exponential itself and, independently, in the close tracking of NPP by local (site) naturalization rates shown in Woodward & Kelly (2008) using a large proportion of the data included here (Notes S3: The nature of equilibrium). In this picture, biotic resistance is best portrayed as a moving ceiling responding to generally increasing productivity levels; the apparent 'failure' of biotic resistance is rather a reflection of its innate character. In such a picture, the current escalation of species naturalizations, carrying with it potentially destructive invasive weeds (Rejmánek & Randall, 2004; Ricciardi & Kipp, 2008), will continue as long as NPP continues to increase, a phenomenon generally attributed to ongoing global climate change and potentially tied into increasing concentrations of atmospheric carbon dioxide (Woodward & Kelly, 2008).

Our central premise in obtaining the most probable distributions by maximizing Eqn 1 is that, without the specified conditions, every object (species or sites or groups of sites) has the same probability of being in any class. From the observed exponential distribution of naturalized species over sites, we infer that every species in our data set has the same *a priori* probability of being in any class and all arrangements corresponding to a given configuration are equally probable, similarly for sites over species. One ecological model of this would be that every species is identical and further that every site is identical; the wide range of environments and species comprising our data set and the reported variety of mechanisms for individual cases of naturalization (Mack *et al.*, 2000; Mitchell *et al.*, 2006) disallow this assumption.

A reasonable basis for the observed distributions and the consequent inference of independence in the action of the component species and sites is provided by the concept of idiosyncrasy (Pueyo *et al.*, 2007). Idiosyncratic species are each governed by a single ecological model, of which there

are very many – each has a unique, highly complex niche. This implies that any species plucked at random has the same probability of ending in the class characterized by that species being found at n sites. Like Hutchinson's (1957) classic ' n -dimensional hypervolume', idiosyncratic niches contain the full range of factors permitting a species to persist at a site, environmental conditions, competitors, consumers, infectious diseases and mutualists as well as resources. With this definition, the distribution of naturalized species over the number of alien sites reached (n) is given by an exponential once we maximize the number of equivalent configurations with Eqn 1. Similarly, the distribution of (idiosyncratic) pairs of sites over classes defined by the number of species in common is given by an exponential once the number of configurations is maximized. That potential species (and potential niches) are so varied is the underlying assumption of the idiosyncratic model of species abundance, so that 'the bits of information which are different in different (ecological) models cancel out' (Pueyo *et al.*, 2007).

Previous applications of statistical mechanics to community assembly have focused on the lognormal distribution of individuals over species within a guild (ecologically similar taxa) of a single community, and so have not had information necessary to discriminate between neutral and idiosyncratic explanations (although recent analyses have demonstrated that even highly similar co-occurring species cannot be assumed to meet the fundamental neutrality criterion of species interchangeability; Kelly *et al.*, 2008, 2010; Leibold, 2008). The relation between our treatment of naturalized species, Pueyo *et al.* (2007) and other recent works employing statistical mechanics in ecology (Pueyo, 2006; Dewar & Porté, 2008; Harte *et al.*, 2008; Bowler & Kelly, 2010) is discussed in Notes S4: Statistical mechanics in ecology.

Independent evidence ties our findings directly into the fundamental nature of community assembly: free-living heteroflagellate communities show a similar exponential distribution of species across sites (Lee & Patterson, 1998), as do tree species from the tropical deciduous forest of México (Trejo & Dirzo, 2002). This is not particularly surprising: the relationship between productivity and diversity in naturalized species reported in Woodward & Kelly (2008) suggests general correlation of species diversity with productivity and the determinants of productivity (Mittelbach *et al.*, 2001; Hawkins *et al.*, 2003; Gillman & Wright, 2006; Krefl & Jetz, 2007). The natural inference is of a similarly general directionality between productivity and diversity, an inference in accord with recent theoretical treatments relating diversity to both complexity and productivity (Tokita, 2004, 2006; Dewar & Porté, 2008; Harte *et al.*, 2008). At smaller scales, the reverse has been observed, with productivity apparently causally affected by diversity (Flombaum & Sala, 2008). Scale dependence in

the directionality of the relationship is an intuitively satisfying integration of these differences, with productivity determining the population process of species entry as proposed in Tilman (2004), along the major axis of the relationship, and filtering of species (sampling) through subsequent species interactions affecting the variation at any particular point along that axis as in Flombaum & Sala (2008).

In conclusion, the primary result of our treatment of species naturalization is a new view of the fundamental processes governing community assembly and diversity – identifying the significance of the alien footprint, the implications of a causative role for productivity and the rapidity with which equilibrium species number can be reached – but it also generates subsidiary insights. Regardless of the extent to which an assumption of idiosyncrasy holds, the data of Figs 1–3 make it most unlikely that any single pronounced signature will reveal species that can easily naturalize; while there may be geographically or taxonomically local generalities, no one solution will be universal, consistent with recent reviews of empirical species naturalization studies (Mack *et al.*, 2000; Mitchell *et al.*, 2006). The implication of species idiosyncrasy also provides an explanation in the same vein for the general observation that the majority of species have a restricted distribution and few species are widespread over many sites; this pattern is an emergent property deriving from the fundamental nature of niches themselves, and does not require the operation of any particular trait of any specific niche (cf. Brown, 1984, 1995). We have shown here a quantitative (exponential) character to that general observation, making possible an analytical tool carrying with it a degree of rigor not previously available to the comparative study of species' distributions.

Acknowledgements

We thank Oliver Pybus for his comments on an earlier draft of the manuscript, Chris Preston for providing the UK Alien Species list and David Patterson for giving us access to his heteroflagellate data. We thank S. Pueyo and two anonymous referees for their comments on the manuscript. CKK was supported in part by DEB 0713866 and a grant from the National Geographic Society; she thanks the Chamela Biological Station for its hospitality during manuscript development.

References

- Anonymous. 2007a. *Alien species recognized to be established in Japan* [WWW document]. URL <http://www.env.go.jp/en/topic/as/041110.pdf>.
- Anonymous. 2007b. *Latvijas svēszemju sugu datu bāze* [WWW document]. URL http://biodiv.lvgma.gov.lv/cooperation/invaz/i-netam_invazivie.xls.

- Anonymous. 2007c. *Vöörliikide andmebaa* [WWW document]. URL <http://loodus.keskkonnainfo.ee/voorliigid/>.
- Bobel E, Król W. 2007. *Alien species in Poland* [WWW document]. URL <http://www.iop.krakow.pl/ias/Baza.aspx>.
- Bowler MG, Kelly CK. 2010. *The general theory of species abundance distributions* [WWW document]. URL <http://arxiv.org/abs/1002.5008>.
- Braun K. *Swaziland's Alien Plants Database* [WWW document]. URL <http://www.kbraunweb.com/alienplants/speciesstatus.asp>.
- Brown JH. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124: 255–279.
- Brown JH. 1995. *Macroecology*. Chicago, IL, USA: University of Chicago Press.
- Casgrain P, Legendre P. 2001. *The R package for multivariate and spatial analysis, version 4.0 d5*. Montreal, Canada: Département de Sciences Biologiques, Université de Montréal.
- Castro SA, Figueroa JA, Muñoz-Schick M, Jaksic FM. 2005. Minimum residence time, biogeographical origin, and life cycle as determinants of the geographical extent of naturalized plants in continental Chile. *Diversity and Distributions* 11: 183–191.
- Corlett RT. 1988. The naturalized flora of Singapore. *Journal of Biogeography* 15: 657–663.
- Dafni A, Heller D. 1990. Invasions of adventive plants in Israel. In: di Castri F, Hansen AJ, Debussche M, eds. *Biological invasions in Europe and the mediterranean basin*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 135–160.
- Darwin C. 1859. *On the origin of species. A facsimile of the first edition*. Cambridge, MA, USA: Harvard University Press.
- Dewar RC, Porté A. 2008. Statistical mechanics unifies different ecological patterns. *Journal of Theoretical Biology* 251: 389–403.
- Drăgulescu A, Yakovenko VM. 2001. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A* 299: 213–221.
- Edgar E, Connor HE. 1999. *Flora of New Zealand. V. Gramineae*. Lincoln, New Zealand: Manaaki Whenua Press.
- Elton C. 1958. *The ecology of invasions by animals and plants*. London, UK: Chapman & Hall.
- Flombaum P, Sala OE. 2008. Higher effect of plant species diversity on productivity in natural than artificial ecosystems. *Proceedings of the National Academy of Sciences, USA* 105: 6087–6090.
- Gaston KJ. 2003. *The structure and dynamics of geographic ranges*. Oxford, UK: Oxford University Press.
- Gillman LN, Wright SD. 2006. The influence of productivity on the species richness of plants: a critical assessment. *Ecology* 87: 1234–1243.
- Gotelli NJ, Anderson MJ, Arita HT, Chao A, Colwell RK, Connolly SR, Currie DJ, Dunn RR, Graves GR, Green JL *et al.* 2009. Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters* 12: 873–886.
- Haegeman B, Etienne RS. 2010. Entropy maximization and the spatial distribution of species. *American Naturalist* 175: E74–E90.
- Harte J, Zillio T, Conlisk E, Smith AB. 2008. Maximum entropy and the state-variable approach to macroecology. *Ecology* 89: 2700–2711.
- Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan J-F, Kaufman DM, Kerr JT, Mittelbach G, Oberdorff T, O'Brien EM *et al.* 2003. Energy, water and broad-scale geographic patterns of species richness. *Ecology* 84: 3105–3117.
- Healy AJ, Edgar E. 1980. *Flora of New Zealand. III. Adventive cyperaceous, petalous & spathaceous monocotyledons*. Wellington, New Zealand: P. D. Hasselberg.
- Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology* 22: 415–427.
- Hutchinson GF. 1959. Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist* 93: 145–159.
- Kelly CK, Bowler MG, Joy JB, Williams JN. 2010. *Temporal niche dynamics of closely related tree species on Barro Colorado island, Panama* [WWW document]. URL [ArXive http://arxiv.org/abs/1008.2527](http://arxiv.org/abs/1008.2527).
- Kelly CK, Bowler MG, Pybus OG, Harvey PH. 2008. Phylogeny, niches and relative abundance in natural communities. *Ecology* 89: 962–970.
- Kennedy TA, Naeem S, Howe KM, Knops JMH, Tilman D, Reich PB. 2002. Biodiversity as a barrier to ecological invasion. *Nature* 417: 636–638.
- Khamis HJ. 2000. The two-stage δ -corrected Kolmogorov-Smirnov test. *Journal of Applied Statistics* 27: 439–450.
- Kreft H, Jetz W. 2007. Global patterns of determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences, USA* 104: 5925–5930.
- Lee WJ, Patterson DJ. 1998. Diversity and geographic distribution of free-living heterotrophic flagellates – analysis by PRIMER. *Protist* 149: 229–244.
- Leibold MA. 2008. Return of the niche. *Nature* 454: 40–41.
- Levine JM. 2000. Species diversity and biological invasions: relating local process to community pattern. *Science* 288: 852–854.
- Lonsdale WM. 1999. Global patterns of plant invasions and the concept of invasibility. *Ecology* 80: 1522–1536.
- MacArthur RH. 1957. On the relative abundance of bird species. *Proceedings of the National Academy of Sciences, USA* 43: 293–295.
- MacArthur RH. 1960. On the relative abundance of species. *American Naturalist* 94: 25–36.
- MacArthur RH. 1965. Patterns of species diversity. *Biological Reviews* 40: 510–533.
- Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10: 689–710.
- May RM. 1973. *Stability and complexity in model ecosystems*. Princeton, NJ, USA: Princeton University Press.
- Mitchell CE, Agrawal AA, Bever JD, Gilbert GS, Huffbauer RA, Klironomos JN, Maron JL, Morris WF, Parker IM, Power AG *et al.* 2006. Biotic interactions and plant invasions. *Ecology Letters* 9: 726–740.
- Mittelbach G, Steiner CF, Scheiner SM, Gross KL, Reynolds HL, Waide RB, Willig MR, Dodson SI, Gough L. 2001. What is the observed relationship between species richness and productivity? *Ecology* 82: 2381–2396.
- Pielou EC. 1979. *Biogeography*. Malabar, FL, USA: Krieger Publishing.
- Pimm SL. 1991. *The balance of nature?* Chicago, IL, USA: University of Chicago.
- Preston CD, Pearman DA, Dines TD. 2006. *New atlas of the British and Irish flora (alien species)*. Oxford, UK: Oxford University Press.
- Pueyo S. 2006. Diversity: between neutrality and structure. *Oikos* 112: 392–405.
- Pueyo S, He F, Zillio T. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters* 10: 1017–1028.
- Pysek P, Sádlo J, Mandák B. 2002. Catalogue of alien plants of the Czech republic. *Preslia* 74: 97–186.
- Rejmánek M, Randall JM. 2004. The total number of naturalized species can be a reliable predictor of the number of alien pest species. *Diversity and Distributions* 10: 367–369.
- Ricciardi A, Kipp R. 2008. Predicting the number of ecologically harmful exotic species in an aquatic system. *Diversity and Distributions* 14: 374–380.
- Rice P. 2007. *Invaders Database System (Wyoming)* [WWW document]. URL <http://invader.dbs.umt.edu>.
- Rosenzweig ML. 1995. *Species diversity in space and time*. Cambridge, UK: Cambridge University Press.
- Sax DF. 2002. Native and naturalized plant diversity are positively correlated in scrub communities of California and Chile. *Diversity and Distributions* 8: 193–210.
- Sax DF, Gaines SD. 2003. Species diversity: from global decreases to local increases. *Trends in Ecology and Evolution* 18: 561–566.

- Shipley B, Vile D, Garnier E. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* 314: 812–814.
- Stohlgren TJ, Binkley D, Chong GW, Kalkhan MA, Schell LD, Bull KA. 1999. Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs* 69: 25056.
- Tilman D. 2004. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences, USA* 101: 10854–10861.
- Tokita K. 2004. Species abundance patterns in complex evolutionary dynamics. *Physical Review Letters* 93: 178102 (178101.178104).
- Tokita K. 2006. Statistical mechanics of relative species abundance. *Ecological Informatics* 1: 316–324.
- Trejo I, Dirzo R. 2002. Floristic diversity of Mexican seasonally dry tropical forests. *Biodiversity and Conservation* 11: 2063–2084.
- Tye A. 2001. Invasive plant problems and requirements for weed risk assessment in the Galapagos islands. In: Groves RH, Panetta FD, Virtue JG, eds. *Weed risk assessment*. Collingwood, Australia: CSIRO Publishing, 153–175.
- Webb CJ, Sykes WR, Garnock-Jones PJ. 1988. *Flora of New Zealand, vol iv. Naturalized pteridophytes, gymnosperms and dicotyledons*. Christchurch, New Zealand: DSIR Botany Division.
- Wester LL. 1992. Origin and distribution of adventive alien flowering plants in Hawai'i. In: Stone CP, Smith CW, Tunison JT, eds. *Alien plant invasions in native ecosystems of Hawai'i: management and research*. Honolulu, HI, USA: University of Hawai'i Cooperative National Park Resources Study Unit, 99–154.
- Wittenberg R. 2005. *An inventory of alien species and their threat to biodiversity and economy in Switzerland*. CABI Bioscience Switzerland Centre/ Swiss Agency for Environment, Forests and Landscape. The environment in practice no. 0629. Bern, Switzerland: Federal Office for the Environment.
- Woodward FI, Kelly CK. 2008. Responses of global plant diversity capacity to changes in carbon dioxide concentration and climate. *Ecology Letters* 11: 1229–1237.
- Wu S-H, Hsieh C-F, Rejmánek M. 2004. Catalogue of the naturalized flora of Taiwan. *Taiwania* 49: 16–31.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Patterns expected from a diffusion process.

Notes S1 The failure of dispersal to describe the observed patterns.

Notes S2 Maximization subject to constraints and determination of parameters.

Notes S3 The nature of equilibrium.

Notes S4 Statistical mechanics and maximum entropy in ecology.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.