The following resources related to this article are available online at
www.sciencemag.org (this information is current as of February 20, 2009):

A correction has been published for this article at:
http://www.sciencemag.org/cgi/content/full/sci;316/5822/201a

Updated information and services, including high-resolution figures, can be found in the online
version of this article at:
http://www.sciencemag.org/cgi/content/full/314/5800/812

A list of selected additional articles on the Science Web sites **related to this article** can be
found at:
http://www.sciencemag.org/cgi/content/full/314/5800/812#related-content

This article has been **cited by** 26 article(s) on the ISI Web of Science.

This article has been **cited by** 10 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/314/5800/812#otherarticles

This article appears in the following **subject collections**:
Ecology
http://www.sciencemag.org/cgi/collection/ecology

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce**
this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

20. See the supporting material on *Science* Online for specific information about the data.
21. G. R. Huss, in *Astrophysical Implications of the Laboratory Study of Presolar Materials*, T. J. Bernatowicz, E. K. Zinner, Eds. (American Institute of Physics Conference Proceedings 402, Woodbury, NY, 1997), pp. 721–746.
22. SiC X grains make up at most 45 parts per billion of Murchison (*18*). The typical Murchison X grain has an enrichment of 38,000ε in ^{138}Ba (*7*) and has a Ba concentration of approximately twice that of CI Ba (9 ppm) (*8*). This will contribute only 0.3 ppm excess in the whole rock of Murchison.
23. Q. Yin, S. B. Jacobsen, K. Yamashita, *Nature* **415**, 881 (2002).
24. Q. Yin, S. B. Jacobsen, J. Blichert-Toft, P. Telouk,

F. Albarede, *Lunar Planet. Sci.* **XXXII**, A2128 (2001).
25. Excesses in these X grains as compared to Earth are 38,080ε for ^{138}Ba/^{136}Ba and 22,230ε for ^{96}Zr/^{94}Zr in grain 113-3 and 3310ε for ^{95}Mo/^{96}Mo in grain 113-2 (*7*). Adding between 0.1 and 0.001% of these components to an isotopically normal sample can give the observed variations seen in bulk Murchison (0.35ε in ^{138}Ba/^{136}Ba, 48ε in ^{96}Zr/^{94}Zr, and 1.80ε in ^{95}Mo/^{96}Mo) (*20, 23*).
26. C. L. Harper Jr., S. B. Jacobsen, *Nature* **360**, 728 (1992).
27. G. Caro, B. Bourdon, J. L. Birck, S. Moorbath, *Geochim. Cosmochim. Acta* **70**, 164 (2006).
28. S. B. Jacobsen, C. L. Harper Jr., *Geochim. Cosmochim. Acta* **60**, 3747 (1996).
29. This work was done under a NASA Origins of Solar System Grant (NAG5-13164) and a NASA Cosmochemistry Grant

From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity

Bill Shipley,[1]* Denis Vile,[1,2] Éric Garnier[2]

We developed a quantitative method, analogous to those used in statistical mechanics, to predict how biodiversity will vary across environments, which plant species from a species pool will be found in which relative abundances in a given environment, and which plant traits determine community assembly. This provides a scaling from plant traits to ecological communities while bypassing the complications of population dynamics. Our method treats community development as a sorting process involving species that are ecologically equivalent except with respect to particular functional traits, which leads to a constrained random assembly of species; the relative abundance of each species adheres to a general exponential distribution as a function of its traits. Using data for eight functional traits of 30 herbaceous species and community-aggregated values of these traits in 12 sites along a 42-year chronosequence of secondary succession, we predicted 94% of the variance in the relative abundances.

Ecological models of community assembly have traditionally been conceived through the lens of population ecology (*1–4*). A classic example is the Lotka-Volterra system of equations in which the per capita growth rate of each of *S* species is linked through an *S*-by-*S* "community" matrix of interaction coefficients. This quickly becomes impractical. Interaction coefficients are difficult to measure and the number of binary interactions, even if unaffected by environmental variation, increases with the square of the number of species in the regional pool. Although such models can qualitatively describe community assembly in very simple empirical systems involving a few unicellular species in well-mixed and simple environments (*3, 5, 6*), they have not been able to do this in more complex field situations involving many species. Empirical attempts to predict relative abundance from plant traits

without any theoretical context have largely failed (*7*). A more recent demographic approach is to assume that the per capita growth rates or average individual fitness of all species in a metapopulation are equal, leading to neutral models of community structure (*4, 8*). Neutral models cannot predict which species will have which relative abundance; indeed, they imply that this is impossible because the dynamics are completely random. Furthermore, their community-level patterns (*9*) and the assumption of average individual fitness of all species in the metacommunity (*10*) have been rejected by empirical data.

We adopted the view of community assembly as a process of species sorting, leading to the concept of community assembly by environmental filters (*11–13*). A pool of *S* species determines those that can disperse to a site. A sorting of these species by the environment, including the environment as modified by the interacting species, determines the relative abundances of those that actually occur at a site. This is simply a result of the nonrandom demographic processes of dispersal, growth, survival, and reproduction of individuals within and between species, as

determined by the functional traits they possess, when the interspecific variations of such functional traits are greater than their intraspecific variations.

Species sorting by the environment is a process of selection over short (ecological) time scales, resulting in changes in species composition over an environmental gradient. The breeder's equation of quantitative genetics (*14*) (Eq. 1b or its multivariate equivalent) describes how the mean value of a quantitative trait *j* of genotype *i*, occurring with proportion p_{ik} at site *k* in a population of *n* genotypes at time *x*, changes over the time period during which a selection event occurs. The amount of change in the mean trait value $(\overline{t_j})$ after the selection event is a function of its heritability (h^2, the slope of a regression of offspring trait values on midparent trait values) and the force of selection [$S_j(x)$] for that trait at time *x*:

$$\overline{t_j(x)} = \sum_{i=1}^{n} p_{ik}(x)t_{ij} \qquad (1a)$$

$$\left[\overline{t_j(1)} - \overline{t_j(0)} \right] = h^2(0)S_j(0) \qquad (1b)$$

Equation 1 is usually applied to selection between genotypes of a single species, but it also applies even if the genotypes occur in different species. Because intraspecific variances of our functional traits are much smaller than their interspecific variances (*15*), we assumed that such trait values can be approximated as species-specific attributes. We also assumed that intraspecific genotypic evolution of the trait values during ecological time is insignificant relative to preexisting interspecific variation. Given these two assumptions, selection will primarily occur between species, and the heritability will be close to unity, although this is not a requirement of the model. Iterating over time gives

$$\overline{t_j(x)} = \sum_{k=0}^{x-1} h^2(k)S_j(k) + \overline{t_j(0)} \qquad (2)$$

The relative abundance of each species in the regional pool, and therefore the value of $\overline{t_j(0)}$,

[1]Département de Biologie, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada. [2]Centre d'Écologie Fonctionnelle et Évolutive, CNRS, UMR 5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France.

*To whom correspondence should be addressed. E-mail: bill.shipley@usherbrooke.ca

will be the same for all sites in the region. Once community assembly begins, any differences between sites in $\overline{t_j(x)}$—called "community-aggregated" traits (*16*)—will reflect different selective pressures over time or across environmental gradients. These community-aggregated traits reflect the constraints on community assembly imposed by the process of species sorting.

Once plants are at a site, the biomass (abundance) of each will be proportional to the total amount of resource units that each species is able to capture at that point in time. The total number of such captured resources $N_k(x)$ at site k at time x, and the number that are captured by each species $n_{ik}(x)$, defines the abundance structure of the community: $\{n_{1k}(x), n_{2k}(x), \ldots, n_{Sk}(x)\}$. Defining relative abundance as $p_{ik}(x) = n_{ik}(x)/N_k(x)$ describes community structure independent of site productivity.

Assume first a strict equivalence of species such that each species is equally likely to disperse and capture resources. There would be no species sorting and community assembly would be purely random. The number of different ways (W) that the N resource units, comprising the living biomass at a site, could be partitioned into a particular community structure is

$$W = \frac{N!}{n_1!n_2!\ldots n_S!} \quad (3)$$

Site and time indices are suppressed for simplicity. N will be less than the number of atoms of carbon and mineral elements in the biomass because these must respect the stoichiometry of the tissues, but it will be very large. Taking logarithms, Sterling's approximation $[\log(n!) \xrightarrow{n\to\infty} n\log(n)-n]$

can be rearranged to form the following equations:

$$\ln(W) = N\ln(N) - \sum_{i=1}^{S} [n_i \ln(n_i)] \quad (4a)$$

$$\frac{\ln(W)}{N} = -\sum_{i=1}^{S} p_i \ln(p_i) \quad (4b)$$

The most likely community composition is the one maximizing W and, given large N, this is equivalent to maximizing Shannon's (*17*) index of information entropy. If all species are not strictly equivalent, because their traits allow individuals of some species to capture more resources, then not all partitions will be equally likely to occur. The species sorting, based on functional traits, results in particular community-aggregated trait values, and therefore excludes partitions that contradict such values. A maximally random community (**p**) constrained by these community-aggregated values (Eqs. 1 and 2) are the values of p_i that maximize W conditional on all $\overline{t_j}$ plus normalization ($\sum_{i=1}^{S} p_i = 1$). This is found, using the method of Lagrangian multipliers, by defining a new system of equations: $f(p_i) =$

$$\ln(W) - \lambda_0 (1 - \sum p_i) - \sum_{j=1}^{T} \lambda_j \left(\overline{t_j} - \sum_{i=1}^{S} p_i t_{ij} \right)$$

and solving for the $T+1$ values of λ_j when the partial derivates are set to zero. The general solution to this problem (*18*), from the maximum entropy formalism, is Eq. 5, where $\hat{p}_i$ is the predicted relative abundance of species i, possessing the T functional traits $\mathbf{t_i} = \{t_{i1},\ldots,t_{1T}\}$. The maximum likelihood values of the nine estimated values of λ in our study were obtained using the Improved Iterative Scaling algorithm of Della Pietra *et al.* (*19*, *20*).
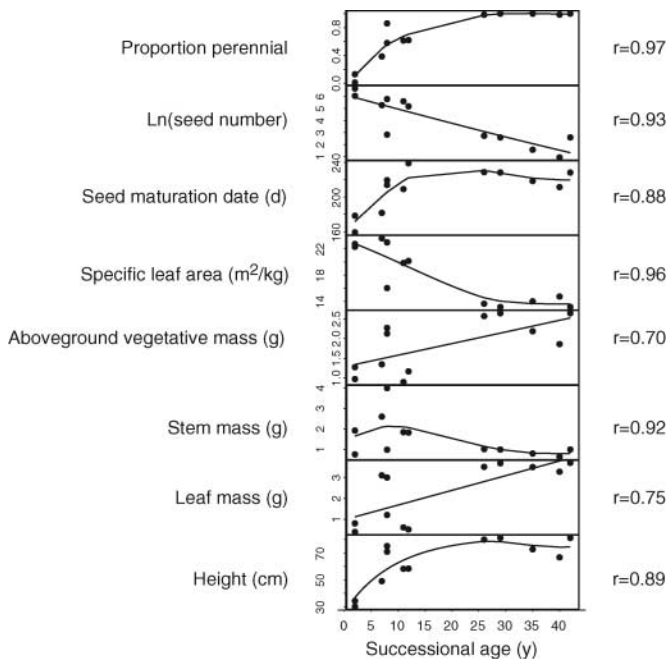

Fig. 1. Each point represents one of 12 sites in a 42-year chronosequence of secondary succession in southern France after vineyard abandonment. The dependent variables are the community-aggregated trait values of the vegetation at that site: proportional perenniality (0 = all biomass belongs to annual species, 1 = all biomass belongs to perennial species), Ln-transformed number of seeds per plant, julian days until seed maturation, specific leaf area (m²/kg), aboveground prereproductive vegetative dry mass (g), stem dry mass (g), leaf dry mass (g), postreproductive height (cm). Lines show the predicted values from a cubic-spline regression and the Pearson correlation coefficient (*r*) relates observed and predicted values.
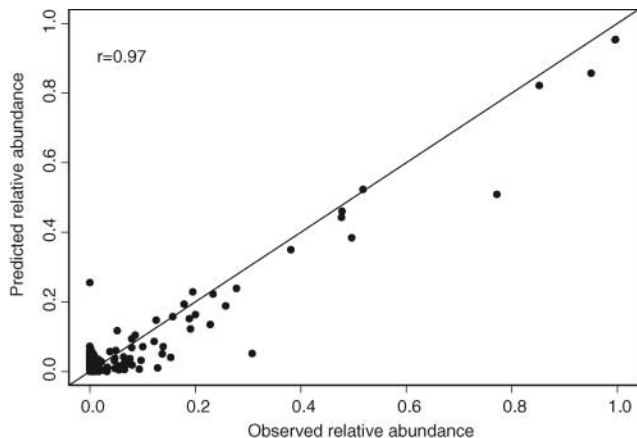

Fig. 2. Observed and predicted relative abundances of 30 species in 12 sites during a 42-year chronosequence of secondary succession. Predicted values are those obtained by maximizing the Shannon entropy conditional on the measured community-aggregated trait values of each site, shown in Fig. 1.

$$\hat{p}_i = \frac{e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)}}{\sum_{i=1}^{S} \left[e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)} \right]} \quad (5)$$

We conducted our study in 12 vineyards that had been abandoned between 2 and 42 years previously, within a 4-km² area of southern France. The aboveground dry biomass of all species in each of four plots measuring 0.25 by 0.25 m was used to estimate relative abundances for each site. Details of the sites and the methods of vegetation sampling are given in (*16*). We measured eight functional traits on 30 species, representing at least 80% of the total biomass

over the chronosequence, and calculated the community-aggregated traits as defined in Eq. 1a. Details of the measurements of these traits and their values are given in (21). Figure 1 shows the eight community-aggregated traits of these 12 sites during the chronosequence. Using these observed community-aggregated traits, the maximum entropy estimates predicted 94% of the observed variation in the actual relative abundances of these 30 species in the 12 sites (Fig. 2).

For generality, it is preferable to predict the community-aggregated traits, given information on environmental conditions, rather than measuring them. Therefore, we predicted the community-aggregated traits in each site from its successional age using cubic-spline regression (Fig. 1). Using these predicted values, we obtained the predicted community structure for each age (Fig. 3, B and D). The observed successional dynamics of each species, smoothed using a cubic-spline smoother, are shown in Fig. 3, A and C; we also show this on a $\log_e$ scale (Fig. 3, C and D) to emphasize the rare species. The observed and predicted successional dynamics are quite close ($r = 0.96$).

Curiously, given the historical dominance of the demographic Lotka-Volterra equations, Volterra recognized the difficulties of this ap-proach and even considered a statistical mechanistic approach (22). Very few authors have followed his lead (23–31). Classical statistical mechanics ignores, but does not deny, the detailed dynamics of atomic colli-sions and assumes a random allocation of atoms to energy states whose total (and there-fore mean) energy is constrained. Our model ignores, but does not deny, the details of resource allocation, population dynamics, sto-chastic processes, and species interactions. It assumes a constrained random allocation of resource units to species; the constraints are generated by natural selection and are quanti-fied by the community-aggregated trait values. The relative abundance of each species in a species pool is therefore a function of how closely its functional traits agree with the community-aggregated traits.

Determining whether our model is gener-ally successful will require further testing. If successful, it could provide a quantitative link between functional ecology, community ecol-ogy, and biogeography, while still being applicable in the field. Furthermore, it could predict which species could successfully invade new environments and how plant communities will change after changes in environments. Because it includes compo-nents of both niche-based and neutral models,

it could also point to a way of reconciling these two different conceptions of community as-sembly. Realizing these potentials will require further empirical testing, a consistent quan-tification of major environmental gradients, and a demonstration of generality in the pat-terns of community-aggregated traits along such gradients.

References and Notes

1. G. F. Gause, *The Struggle for Existence* (Williams and Wilkins, Baltimore, 1934).
2. R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of Species* (Harper and Row, New York, 1972).
3. D. Tilman, *Resource Competition and Community Structure.* (Princeton Univ. Press, Princeton, NJ, 1982).
4. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
5. J. H. Vandermeer, *Ecology* **50**, 362 (1969).
6. J. Passarge, S. Hol, M. Escher, J. Huisman, *Ecol. Monogr.* **76**, 57 (2006).
7. B. R. Murray, P. H. Thrall, A. M. Gill, A. B. Nicotra, *Austral Ecol.* **27**, 291 (2002).
8. G. Bell, *Ecology* **86**, 1757 (2006).
9. B. J. McGill, B. A. Maurer, M. D. Weiser, *Ecology* **87**, 1411 (2006).
10. W. S. Harpole, D. Tilman, *Ecol. Lett.* **9**, 15 (2006).
11. P. A. Keddy, *J. Veg. Sci.* **3**, 157 (1992).
12. E. Weiher, P. A. Keddy, *Oikos* **73**, 323 (1995).
13. J. P. Grime, *Plant Strategies, Vegetation Processes, and Ecosystem Properties* (John Wiley & Sons, New York, ed. 2, 2001).
14. D. A. Roff, *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1997).
15. E. Garnier et al., *New Phytol.* **152**, 69 (2001).
16. E. Garnier et al., *Ecology* **85**, 2630 (2004).
17. C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, IL, 1949).
18. E. T. Jaynes, *Probability Theory: The Logic of Science.* (Cambridge Univ. Press, Cambridge, UK, 2003).
19. S. Della Pietra, V. Della Pietra, J. Lafferty, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 380 (1997).
20. A program to obtain our parameter estimates is available online (http://pages.usherbrooke.ca/jshipley/recherche/my%20programs.htm).
21. D. Vile, B. Shipley, E. Garnier, *Ecology* **87**, 504 (2006).
22. V. Volterra, *Leçons sur la Théorie Mathématique de la Lutte pour la Vie* (Gauthier-Villars, Paris, 1931).
23. M. Plank, *SIAM J. Appl. Math.* **59**, 1540 (1999).
24. M. Plank, *Nonlinearity* **9**, 887 (1996).
25. M. Plank, *J. Math. Phys.* **36**, 3520 (1995).
26. E. H. Kerner, *Gibbs Ensemble: Biological Ensemble* (Gordon and Brench, New York, 1964).
27. D. Lurie, J. Wasenburg, in *Thermodynamics and Regulation of Biological Processes*, I. Lambrecht, A. I. Zotin, Eds. (De Gruyter, Berlin, 1984), pp. 257–271.
28. A. P. Levich, in *Lectures in Theoretical Biology*, T. Tiivel, Ed. (Valgus, Tallinn, Estonia, 1988) pp. 121–127.
29. V. L. Alexeyev, A. P. Levich, *Bull. Math. Biol.* **59**, 649 (1997).
30. A. P. Levich, *Ecol. Modell.* **131**, 207 (2000).
31. B. A. Maurer, *Untangling Ecological Complexity: The Macroscopic Perspective* (Univ. of Chicago Press, Chicago, 1999).
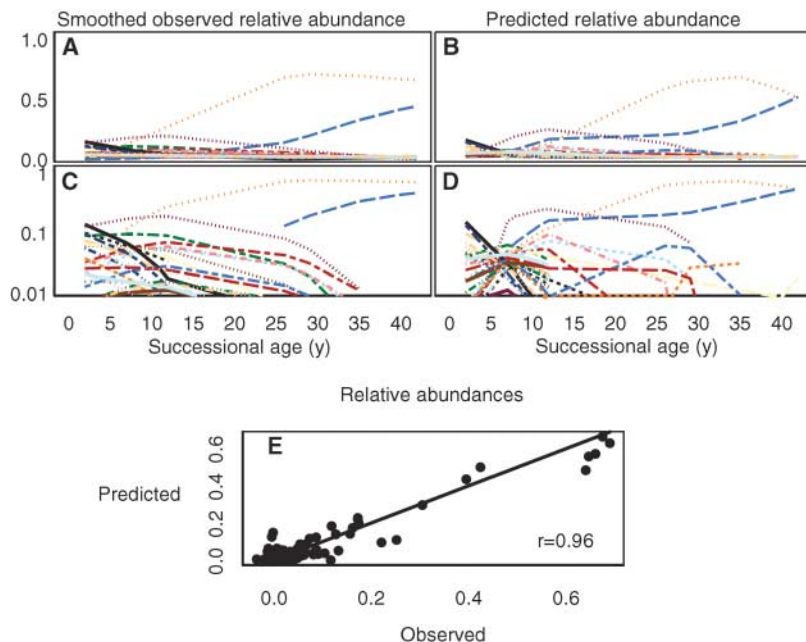32. This research was financially supported by the Natural Sciences and Engineering Research Council of Canada.

Fig. 3. (**A** and **C**) Observed relative abundance of each of 30 species during a 42-year chrono-sequence of secondary succession. Curves for each species were obtained using a cubic-spline regression. (**B** and **D**) Predicted relative abundance of each species, obtained by maximizing the Shannon entropy conditional on the measured community-aggregated trait values of each site, shown in Fig. 1. (A) and (B) show relative abundances on an arithmetic scale and (C) and (D) show the same data on a logarithmic scale in order to emphasize the rarer species. (**E**) The smoothed observed relative abundances of each species at each successional age plotted against the relative abundances predicted by maximizing the Shannon entropy conditional on the measured community-aggregated trait values at each successional age.

ERRATUM

Reports: "From plant traits to plant communities: a statistical mechanistic approach to biodiversity" by B. Shipley *et al.* (3 Nov. 2006, p. 812). In the denominator of Eq. 5 on page 813, the expression in the parentheses should have been set as an exponent of *e*. The correct equation is

$$\hat{p}_i = \frac{e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)}}{\sum_{i=1}^{S} \left(e^{\left(\lambda_0 - \sum_{j=1}^{T} \lambda_j t_{ij}\right)}\right)}$$