Supporting Online Material
Manuscript 1115576

Materials and Methods

The data for the Cape Floristic Region (CFR) sites comes from 5x10 meter relevé plots, which had been surveyed using standard Braun-Blanquet survey methods: all vascular plant species had been identified to species and assigned a categorical percent cover score (*S1*). From the complete survey data, "canopy" shrubs and small trees, as well as smaller shrubs and shrublets were selected. These woody plants cover a range of relative sizes roughly equivalent to the range in relative sizes for rain forest canopy and understory trees surveyed by Pitman et al. (*S2, S3*) and Condit et al. (*S4*). All CFR plots are located in fynbos regions in areas undisturbed by land use transformation. Species names were standardized using the reference Cape Flora (*S5*) and, as for the Amazon data (*S3*), individuals not identified to species were removed.

Since neutral theory describes ecological dynamics at the individual level, it was necessary to convert the areal or cover-based Braun-Blanquet scores for the CFR data to numbers of individuals. We used an independent data set that contains both counts of individuals and cover estimates across a range of typical fynbos plots to calibrate the number of individuals per square meter (*S6*). The conversion factors were estimated separately for each growth form classification: trees (0.1 individual/m^2), large shrubs (0.25 individual/m^2), shrubs (4 individuals/m^2), and shrublets (8 individuals/m^2). Sensitivity of parameter estimates to these conversion factors was assessed, and the results presented are qualitatively unaffected across a reasonable range of values. Despite the difference in methodology (cover-based sampling versus counts of individuals), mean plant size and plot size, the total sampling effort for each sub-regional site was approximately equivalent for the two data sets (see Table 1). Both data sets consist of a set of plots distributed non-randomly across identifiable sub-regions to capture the range of typical vegetation sub-types within the sub-region (*S2, S7-9*).

Parameter estimates were obtained using a likelihood derived from the closed-form expression for the ZSM:

$$L\{S_1, \ldots S_n \mid \Theta, m\} = \prod_{n=1}^{J} \Phi_n^{S_n} = \frac{\Theta}{n} \binom{J}{n} \int_{x=0}^{\gamma} \frac{\Gamma_{\gamma-x}}{\Gamma_{x+1}} \frac{\Gamma_{J+n-\gamma-x}}{\Gamma_{\gamma-x}} e^{\frac{-\gamma\Theta}{x}} dx \qquad (1)$$

where $\gamma = \dfrac{(J-1)m}{1-m}$ and J is the (known) sample size (*S10-12*). In this function, Φ_n is the (predicted) probability that a species randomly drawn from the local community sample has abundance n, and S_n is the observed number of species of abundance n (*S10*).

Parameter estimates were obtained both by maximizing the likelihood and by Bayesian posterior simulation. Maximum likelihood estimates were obtained through minimization of the negative log-likelihood. Using the full likelihood to fit the model is preferable to earlier efforts that have used χ^2 deviance measures on $\log_2$ binned abundances (*S11, 13*) because this approach uses the full data set rather than summarized data (*S10*).

For Bayesian inference, a flat beta prior was placed on *m* and an uninformative gamma prior placed on Θ, with parameters selected so that 95% of the posterior mass of

the prior was below 4,000. Samples were thus drawn from the following (unnormalized) posterior distribution, where $L\{S_1, \ldots S_n \mid \Theta, m\}$ is taken from (1):

$$P\{\Theta, m \mid S_1, \ldots, S_n, J\} \propto L\{S_1, \ldots S_n \mid \Theta, m\} \cdot Gamma_{(1, 0.006)}(\Theta) \cdot Beta_{(1,1)}(m) \tag{2}$$

Convergence was assessed through visual inspection of the chains and the Brooks-Gelman-Rubin statistic (*S14*).

Estimates of Θ for the metacommunity neutral model with no migration limitation were obtained by maximizing the likelihood of the metacommunity sampling distribution (S*10*):

$$S_n = \Theta \int_0^1 \exp(-xJ) \; \frac{xJ^n}{n!} \frac{(1-x)^{\Theta-1}}{x} \, dx \tag{3}$$

Comparison between the one and two parameter models were done using a likelihood ratio test, since the models are nested (*S10*).

Supporting References and Notes:

S1. M. J. A. Werger, *Bothalia* **11**, 309-323 (1974).
S2. N. C. A. Pitman, J. W. Terborgh, M. R. Silman, P. V. Nunez, *Ecology* **80**, 2651-2661 (1999).
S3. N. C. A. Pitman *et al.*, *Ecology* **82**, 2101-2117 (2001).
S4. R. Condit *et al.*, *Science* **295**, 666-669 (2002).
S5. P. Goldblatt, J. Manning, *Cape Plants: A Conspectus of the Cape Flora of South Africa* (National Botanical Institute of South Africa, Cape Town, 2000).
S6. P. M. Holmes, D. M. Richardson, B. W. Van Wilgen, C. M. Gelderblom, *Austral Ecol.* **25**, 631-639 (2000).
S7. C. Boucher, *Bothalia* **12**, 455-478 (1978).
S8. H. C. Taylor, *Cederberg Vegetation and Flora* (National Botanical Institute, Cape Town, 1996).
S9. B. E. Van Wyk, P. A. Novellie, C. M. Van Wyk, *Bothalia* **18**, 211-220 (1988).
S10. D. Alonso, A. J. McKane, *Ecol. Lett.* **7**, 901-910 (2004).
S11. I. Volkov, J. R. Banavar, S. P. Hubbell, A. Maritan, *Nature* **424**, 1035-1037 (2003).
S12. M. Vallade, B. Houchmandzadeh, *Phys. Rev. E* **68**, 061902 (2003).
S13. B. J. McGill, *Nature* **422**, 881-885 (2003).
S14. A. Gelman, J. B. Carlin, D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall, New York, 1995).