

## LETTER

## A unified model explains commonness and rarity on coral reefs

Sean R. Connolly,<sup>1,2\*</sup> Terry P. Hughes<sup>1</sup> and David R. Bellwood<sup>1,2</sup>

<sup>1</sup>ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Qld., Australia

<sup>2</sup>Marine Biology and Aquaculture, College of Science and Engineering, James Cook University, Townsville, Qld., Australia

\*Correspondence: E-mail: sean.connolly@jcu.edu.au

### Abstract

Abundance patterns in ecological communities have important implications for biodiversity maintenance and ecosystem functioning. However, ecological theory has been largely unsuccessful at capturing multiple macroecological abundance patterns simultaneously. Here, we propose a parsimonious model that unifies widespread ecological relationships involving local aggregation, species-abundance distributions, and species associations, and we test this model against the metacommunity structure of reef-building corals and coral reef fishes across the western and central Pacific. For both corals and fishes, the unified model simultaneously captures extremely well local species-abundance distributions, interspecific variation in the strength of spatial aggregation, patterns of community similarity, species accumulation, and regional species richness, performing far better than alternative models also examined here and in previous work on coral reefs. Our approach contributes to the development of synthetic theory for large-scale patterns of community structure in nature, and to addressing ongoing challenges in biodiversity conservation at macroecological scales.

### Keywords

Aggregation, biodiversity, coral reefs, macroecology, species-abundance distribution, Taylor's power law, unified theories.

*Ecology Letters* (2017) 20: 477–486

### INTRODUCTION

Patterns of commonness and rarity are fundamental characteristics of ecological assemblages. In general, within communities, most species are very rare, and a few species constitute most of the community (McGill *et al.* 2007). Between communities, individuals of a given species are usually distributed nonrandomly, with more regions of high and low abundance than one would expect by chance (Holt *et al.* 2002). The strength of this aggregation varies between species, with common species typically distributed less unevenly than rare species (Taylor 1961). These spatial and interspecific abundance patterns exhibit both commonalities and differences among taxa and ecosystems that ecologists have sought to explain for more than a century (Hubbell 2001; McGill *et al.* 2007). They also have important implications for ecosystem functioning (Smith & Knapp 2003; Loreau 2010; Mouillot *et al.* 2013), and for how communities may respond to environmental change (He & Hubbell 2011).

The ecological importance of spatial and interspecific patterns of abundance has made explaining such patterns a core criterion for testing proposed 'unified theories of biodiversity'. Such theories are formulated to explain patterns of abundance and species richness based on a limited set of biological or statistical axioms. Sometimes, these axioms consist of a small subset of ecological processes believed to be responsible for ubiquitous ecological patterns. Neutral theory of biodiversity, for instance, assumes that species are ecologically and demographically equivalent, and thus all abundance variation is driven by demographic stochasticity – chance variation in numbers of births, deaths and dispersal events among species and local communities (Hubbell 2001). Alternatively, unified theories may use non-biological algorithms for generating statistical expectations. Maximum Entropy Theory of Ecology

(METE), for instance, uses Lagrange multipliers to derive maximum entropy distributions for abundance patterns, such as species-abundance distributions, by constraining predicted distributions to have the same mean value (i.e. the same number of individuals and number of species) as the empirical data (Harte 2011). Models derived from unified theories often produce compelling agreement with data when individual patterns, such as species-abundance distributions, are considered in isolation (McGill *et al.* 2007). However, they have been less successful when used to explain multiple patterns simultaneously (Adler 2004; McGill *et al.* 2005; Dornelas *et al.* 2006; Xiao *et al.* 2015). In addition, when tested against individual patterns, alternative models often make very similar predictions (for instance, truncated lognormal abundance distributions strongly resemble log-series distributions); in such cases, the model with fewer parameters may well be favoured even if more comprehensive sampling, or testing against multiple patterns, would yield a different result.

An alternative to constructing predictions from a restrictive set of axioms, as existing unified theories do, is to begin with distributions or functional relationships that are known to be ubiquitous in nature, and that can be generated by a broad range of realistic processes (Lawton 1999). Provided that the assumptions required to produce the component relationships are not mutually inconsistent, one could then ask whether a tractable, parsimonious unification of these distributions and functional relationships (hereafter termed a 'unified model') can simultaneously predict multiple macroecological patterns more successfully than unified theories can. If so, the unified model can provide a benchmark against which unified theories can be compared.

Here, we present a new model of metacommunity biodiversity, which unifies four well-known and common macroecological patterns, and we test it on reef-building corals and

labrid fishes, which encompass some of the most critical functional groups on coral reefs (Bellwood *et al.* 2004). We analyse surveys of fifteen meta communities of reef corals and labrid fishes, encompassing three habitat types in five different regions in from the tropical western and central Pacific (Connolly *et al.* 2005; Hughes *et al.* 2014). Coral reefs are highly species-rich ecosystems, consisting of extensive networks of habitat patches connected by dispersal. This makes them ideal systems for testing models of metacommunity structure (Connolly *et al.* 2005, 2009, 2014; Dornelas *et al.* 2006; Volkov *et al.* 2007; Bode *et al.* 2012). Specifically, we assess our model's ability to simultaneously capture species-abundance distributions within local communities, species differences in the strength of aggregation, patterns of species associations (i.e. community similarity), and species richness. We evaluate the model's relative fit, compared to alternative models for intraspecific aggregation, for spatial mean–variance scaling of abundances, and for the metacommunity species-abundance distribution. We also conduct out-of-sample tests of the model's ability to predict species richness at local and metacommunity scales. The unified model generates predictions that are statistically indistinguishable from observed abundance patterns of corals and reef fishes, and performs considerably better than any alternative models that we examine, or that have been proposed to date.

### A unified model

Our goal is to formulate a model that specifies the probability for a distribution of abundances among species, and among sites, in a metacommunity,  $\Pr(\mathbf{R})$ , where  $\mathbf{R}$  is the sites  $\times$  species-abundance matrix. We begin by choosing a distribution to describe the spatial pattern of an individual species' abundances among sites in that metacommunity,  $\Pr(\mathbf{r})$ . Then, we specify how the parameters that characterise that spatial abundance pattern are themselves distributed among species. Finally, we specify the full distribution for the sites-by-species matrix  $\mathbf{R}$ .

To describe how a species' abundances vary across sites, we use a negative binomial distribution. Empirically, this distribution characterises aggregation very well (Engen *et al.* 2008; Zillio & He 2010). Theoretically, the negative binomial can arise from a broad range of different mechanisms used to explain aggregation (Kendall 1949; Boswell & Patil 1970; Nachman 2000). According to the negative binomial, the probability that a species  $i$ , with mean abundance  $\lambda_i$ , has a vector of abundances  $\mathbf{r}_i$  in samples from  $N$  sites is:

$$\Pr(\mathbf{r}_i|k_i, \lambda_i) = \prod_{j=1}^N \left( \frac{\Gamma(r_{ij} + k_i)}{\Gamma(k_i)\Gamma(r_{ij} + 1)} \left( \frac{k_i}{k_i + m_j \lambda_i} \right)^{k_i} \left( \frac{m_j \lambda_i}{k_i + m_j \lambda_i} \right)^{r_{ij}} \right) \quad (1)$$

where  $m_j$  is a measure of sampling effort at site  $j$  (i.e.  $m_j \lambda_i$  is the expected abundance of species  $i$  in the sample from site  $j$ ),  $r_{ij}$  is the number of individuals of species  $i$  in the sample from site  $j$  (i.e. the  $j$ th element of the abundance vector  $\mathbf{r}_i$ ), and  $k_i$  is the shape parameter of the negative binomial distribution.  $k_i$  is an inverse measure of aggregation: as  $k_i$  increases to infinity, a negative binomial sample tends to a poisson (random) sample.

Eqn 1 contains two unknown parameters: the mean abundance  $\lambda_i$  and the shape parameter  $k_i$ . To move from a probability distribution of abundances for a single species across sites, to a probability distribution for all species among sites in a metacommunity (i.e. a multi-site species-abundance distribution), we must characterise how these two parameters vary among species. Our first step is to assume that among species variation in metacommunity mean abundance follows a lognormal distribution. This distribution commonly fits species-abundance distributions very well, particularly at large scales (Connolly *et al.* 2014). Also, the lognormal can arise in a variety of ecologically plausible ways: from differences between species traits, such as variation in species' demographic rates (Engen *et al.* 2002), niche hypervolumes (Sugihara *et al.* 2003; Connolly *et al.* 2014), or responses to environmental fluctuations (Engen & Lande 1996b). The lognormal can also be derived as a generic mathematical approximation to departures from pure demographic stochasticity (Pueyo 2006). If the metacommunity species-abundance distribution is lognormal with parameters  $\mu$  and  $\sigma^2$ ,  $f(\lambda|\mu, \sigma^2)$ , and intraspecific variation in abundance follows a negative binomial, then the probability that an arbitrarily chosen species  $i$  has a vector of abundances  $\mathbf{r}_i$  across  $N$  sites is:

$$\Pr(\mathbf{r}_i|k_i, \mu, \sigma^2) = \int_{\lambda=0}^{\infty} \left[ \prod_{j=1}^N \left( \frac{\Gamma(r_{ij} + k_i)}{\Gamma(k_i)\Gamma(r_{ij} + 1)} \left( \frac{k_i}{k_i + m_j \lambda} \right)^{k_i} \left( \frac{m_j \lambda}{k_i + m_j \lambda} \right)^{r_{ij}} \right) \right] f(\lambda|\mu, \sigma^2) d\lambda \quad (2)$$

(Because each species' mean abundance is drawn from a common metacommunity species-abundance distribution, and we integrate over that distribution to calculate  $\Pr(\mathbf{r})$ , we drop the species-specific subscript  $i$  from the parameter  $\lambda$ .)

Having characterised how mean abundance  $\lambda$  is distributed across species, it remains to characterise how the shape parameter  $k$  of eqn 1 varies among species. To do this, we employ Taylor's power-law scaling of the mean and spatial variance of abundance, which can be generated by common stochastic demographic processes (Keeling 2000; Engen *et al.* 2008; see also Eisler *et al.* 2008 for alternative explanations for power-law scaling). Taylor's power-law is:

$$v = a\lambda^b \quad (3)$$

where  $v$  is a species' spatial variance of abundance, and  $a$  and  $b$  are parameters. The variance of the negative binomial distribution is

$$v = \lambda + \lambda^2/k \quad (4)$$

Setting the right-hand sides of eqns 3 and 4 equal to each other, and solving for  $k$ , we obtain:

$$k(a, b, \lambda) = \frac{\lambda}{a\lambda^{b-1} - 1} \quad (5)$$

With the right-hand side of eqn 5 substituted for  $k_i$  in eqn 2, only four free parameters remain: the mean and variance of log abundance in the metacommunity,  $\mu$  and  $\sigma^2$ , and the power-law scaling parameters  $a$  and  $b$  that specify a species' spatial variance, given its mean abundance:

$$\Pr(\mathbf{r}_i|a, b, \mu, \sigma^2) = \int_{\lambda=0}^{\infty} \left[ \prod_{j=1}^N \left( \frac{\Gamma(r_{ij} + \frac{\lambda}{a\lambda^{b-1}-1})}{\Gamma(\frac{\lambda}{a\lambda^{b-1}-1}) \Gamma(r_{ij} + 1)} \right) \left( \frac{\frac{\lambda}{a\lambda^{b-1}-1}}{a\lambda^{b-1}-1 + m_j\lambda} \right)^{\frac{\lambda}{a\lambda^{b-1}-1}} \left( \frac{m_j\lambda}{a\lambda^{b-1}-1 + m_j\lambda} \right)^{r_{ij}} \right] f(\lambda|\mu, \sigma^2) d\lambda \quad (6)$$

Last, we assume that species' abundances are distributed among sites independently of one another. That is, whether a species is present or absent from a site, or, if present, whether it is common or rare, is statistically independent of the species identities of the other individuals that are present there. This assumption of statistical independence, an assumption shared by all unified theories of biodiversity (McGill 2010), can occur if species' abundances are ecologically independent of one another. It may also occur if species interactions are negligibly weak, sufficiently diffuse, or sufficiently variable in nature among locations that any particular interactions have no detectable effects on species' relative abundances across sites (e.g. Callaway 1997).

With this final assumption, the probability distribution of abundances among species and sites is fully specified by the unified model:

$$\Pr(\mathbf{R}) = \prod_{i=1}^S \Pr(\mathbf{r}_i|a, b, \mu, \sigma^2) \quad (7)$$

where  $\Pr(\mathbf{r}_i|a, b, \mu, \sigma^2)$  is eqn 6.

## METHODS

### Study system

The coral and fish surveys that we conducted spanned c. 10 000 km from Indonesia to French Polynesia (Fig. 1a). For corals, within each of five regions in the central Indo-Pacific (Indonesia, Papua New Guinea, Solomon Islands, Samoa and French Polynesia), twelve sites for each of three habitat types (crest, flat and slope) were sampled. At each site, species-level coral composition quantified by means of the line-intercept method along ten 10-m transects. For fishes, the same habitat types and regions were sampled, except that Micronesia replaced the Solomon Islands due to civil unrest in the latter region. Eight sites were surveyed in each region. At each site, two timed-swim (20-min) belt transects were run: one 5 m wide, to identify and count fish > 10 cm length, and one 1 m wide, to count fish < 10 cm length. To equalise sampling effort, the fish counted on the wide transects were subsampled, with each fish given a 20% chance of appearing in the sample for analysis. In keeping with our previous analyses (Connolly *et al.* 2005, 2009; Dornelas *et al.* 2006), we treat sites from the same habitat type and region as belonging to the same metacommunity. This protocol avoids confounding within-habitat aggregation effects with among-habitat (beta diversity) or biogeographical (gamma-diversity) variation in species composition.

### Testing the unified model

Before fitting the unified model (eqn 7) to these data, we must modify it to account for the fact that any species appearing in

a community matrix,  $\mathbf{R}$ , by definition, is observed at least once (i.e. it does not have abundance zero everywhere). Thus, our likelihood is:

$$L(\mathbf{R}|a, b, \mu, \sigma^2) = \prod_{i=1}^S \frac{\Pr(\mathbf{r}_i|a, b, \mu, \sigma^2)}{1 - \Pr(\mathbf{0}|a, b, \mu, \sigma^2)} \quad (8)$$

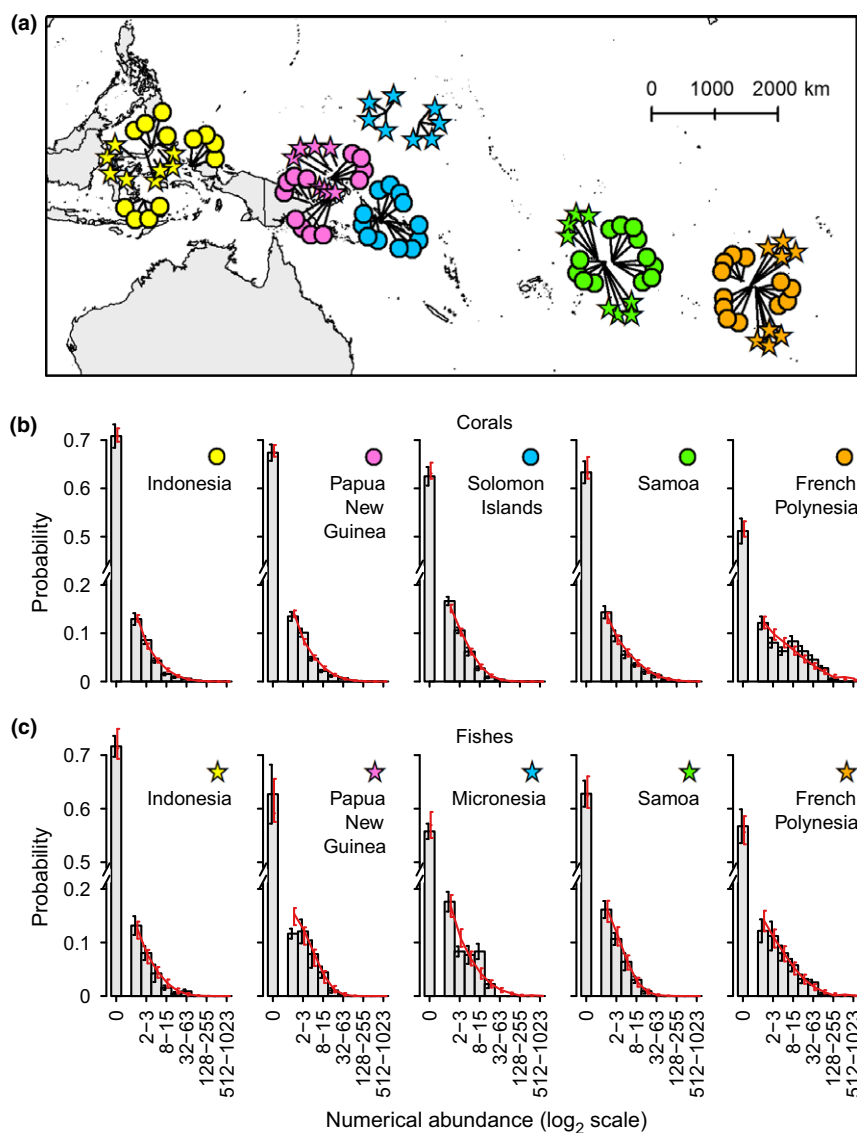
where  $\Pr(\mathbf{r}_i|a, b, \mu, \sigma^2)$  is eqn 6. The denominator zero-truncates the likelihood: it conditions the probability that a species has a particular vector of abundances  $\mathbf{r}_i$  on the fact that the species is observed in at least one of the sites (i.e. it does not have abundance zero at all sites) (See Appendix S1 for details).

We designed a parametric bootstrap algorithm (Efron & Tibshirani 1993) to test the goodness-of-fit of the unified model. The parametric bootstrap involves simulating data from the best-fitting model, so that the simulated data conform to the assumptions of negative binomial aggregation, power-law scaling, lognormal metacommunity abundance distributions and independence of species' abundances from one another. We then ask whether our model fits the empirical data significantly worse than it fits the simulated data (which we know conform exactly to model assumptions). From the parametric bootstrap simulations, we constructed two test statistics for the unified model. To test the aggregation, mean-variance scaling, and metacommunity species-abundance assumptions, we generated an expected distribution of  $\hat{c}$ , a normalised measure of model fit based on the likelihood function (Connolly *et al.* 2009), under the null hypothesis that the model is true, and compared it with  $\hat{c}$  for the empirical data. This test performs better than a previously used parametric bootstrap test, where each site was fitted independently (Appendix S1). To test the independent-species assumption, we calculated Horn's similarity index for each pair of sites from the same metacommunity (Jost *et al.* 2011), and we asked whether the distribution of similarity values from the empirical data was significantly different from that produced by the simulations. (See Appendix S1 for further details.)

The unified model (and the alternatives we consider) predict numbers of unobserved species; however, such species make no contribution to the likelihood (eqn 8), and thus a model may fit the observed data well, regardless of whether it predicts realistic numbers of unobserved species. Therefore, we also tested the model by generating predictions about local and metacommunity species richness. Given the observed number of species in the data for a metacommunity, and the estimated probability that a species is present in the metacommunity, but not observed in the data (which depends on the maximum likelihood estimates  $\hat{a}$ ,  $\hat{b}$ , and  $\hat{\theta}$ ), we can estimate the total number of species in the metacommunity species pool,  $S_{\text{pool}}$ :

$$\widehat{S}_{\text{pool}} = \frac{S_{\text{obs}}}{1 - \Pr(\mathbf{0}|\hat{a}, \hat{b}, \hat{\mu}, \hat{\sigma}^2)} \quad (9)$$

Because this prediction depends upon the number of unobserved species, which is unknown, we assess the plausibility of the unified model estimates by comparing them with those produced by the non-parametric jackknife, which makes no assumptions about the distribution of species abundances (Burnham & Overton 1979; Connolly *et al.* 2005). To conduct



**Figure 1** (a) Study sites for Indo-Pacific corals (circles) and fishes (stars), along with observed and fitted species-abundance distributions for (b) corals and (c) fishes, for slope habitats. Bars represent the means of the observed number of species per site, averaged across all sites in each metacommunity. Black whiskers show standard errors of the mean. Red lines represent the best-fit unified model fitted to the field data; red whiskers represent expected standard errors, generated by parametric bootstrapping of the unified model. Species are plotted in log<sub>2</sub> classes of abundance, with the addition of a zero-class (species observed somewhere in the metacommunity, but not present at a given site) at the far left of each panel.

an out-of-sample test of the model to compare more directly with observations, we also re-fitted the unified model to each metacommunity multiple times, each time omitting one site from the fitting procedure in turn. We then used the parameter estimates from each fitted data set to predict the number of species observed at one or more of the fitted sites that would not be observed at the excluded site:

$$\hat{S}_{+0}(n) = \widehat{S}_{\text{pool}} \int_{\lambda} \{ [1 - \Pr(\mathbf{0}_{n-1}|\lambda)] \Pr(0|\lambda) \} f(\lambda | \mu, \sigma^2) d\lambda \quad (10)$$

where  $n$  is the number of sites sampled ( $n = 1 \dots N$ ), and  $\mathbf{0}_{n-1}$  is a vector of  $n-1$  zeroes (one for each site included in the fitting procedure).

### Alternative models

In addition to evaluating the absolute fit of the unified model to the data, we compared its fit to that of alternative characterisations of intraspecific aggregation, spatial mean–variance scaling, and the metacommunity species-abundance distribution.

For aggregation, we consider three additional possibilities: the Poisson distribution (i.e. no aggregation), the Poisson-lognormal distribution, and the generalised Poisson distribution of Consul and Jain (1973). The Poisson-lognormal has also been used to characterise aggregation (Engen *et al.* 2002). The generalised Poisson can approximate the binomial, Poisson and negative binomial distributions very closely, and it also

allows for under-dispersed data (where the variance is less than the mean: Consul & Jain 1973). We tested between these alternative models of aggregation independently of assumptions about functional or distributional relationships among species (i.e. mean–variance scaling or species-abundance distributions). (See Appendix S1 for details.)

For mean–variance scaling, a recent study found that a model assuming a single shape parameter  $k$  for all species in a metacommunity (and thus a quadratic mean–variance scaling relationship) performed better than Taylor's power law (Karlson *et al.* 2011; also see He & Gaston 2003). Therefore, we consider an alternative model of interspecific variation in the strength of aggregation by fitting a negative binomial distribution with a single shape parameter  $k$  for all species (i.e. eqn 2 with  $k_i = k$ ). We tested between the power-law and constant- $k$  models by taking the best-fitting aggregation distribution from the test above, and then comparing the fit of these two models for the scaling of mean and spatial variance of abundance.

For the metacommunity species-abundance distribution, we consider the gamma distribution as an alternative to the lognormal. Many species-abundance models are either limiting cases of, or are closely approximated by, samples from a gamma distribution. Such models include several neutral models, the log series, the negative binomial and the gamma-binomial ('GamBin') model (Fisher *et al.* 1943; Uglund *et al.* 2007; Connolly *et al.* 2009, 2014). We therefore compare our original unified model with an alternative unified model for which metacommunity mean abundances follow a gamma distribution (i.e. eqn 6, but using a gamma distribution for the function  $f$  instead of a lognormal).

In all cases, we use AIC to identify the best-fitting models, both for each metacommunity considered individually, and in an 'overall' test that combines evidence across all metacommunities for each taxon. We consider a model as strongly supported if it has an Akaike weight  $> 0.95$ .

### Test of additional data set

As a further test of the unified model, we repeated all of the goodness-of-fit tests, model selection, and out-of-sample prediction analyses, using data for reef fishes on the Great Barrier Reef collected by the Australian Institute of Marine Science Long-Term Monitoring Program (Sweatman *et al.* 2008). We used the same year of data as a previous study of species-abundance distributions (Connolly *et al.* 2014). We considered all sampled mid and outer shelf reefs within a given latitudinal sector as a metacommunity, and we considered each reef as a local community, since all transects conducted on each reef sample are undertaken as near to the northeast flank as possible, and are within a common, narrow depth range. This yielded 59 local communities distributed among seven metacommunities (Fig. S5).

## RESULTS

The unified model successfully and simultaneously explains patterns of commonness, rarity and species richness within and among local communities. Local species-abundance

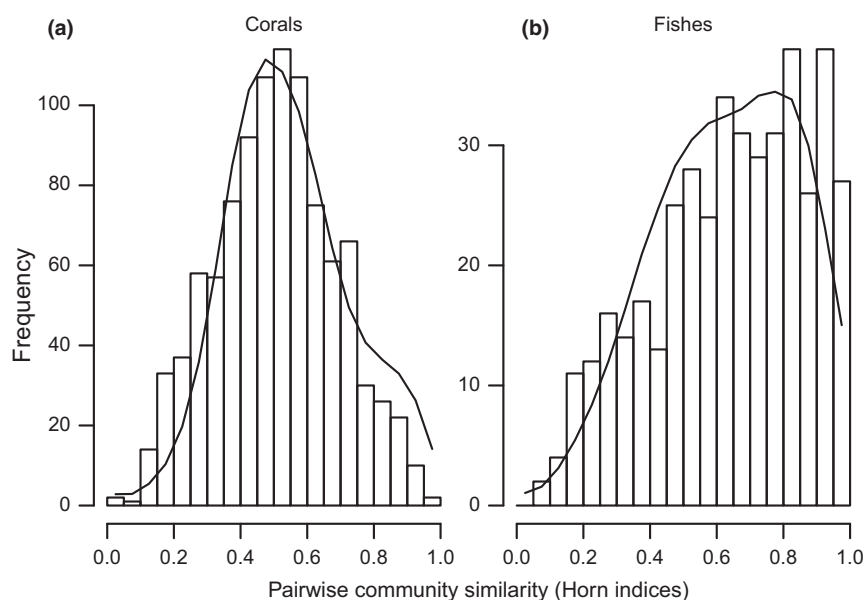
distributions are extremely well-characterised by the unified model (Fig. 1b and c, Fig. S1). The unified model also explains patterns of spatial variation in abundance. Intraspecific mean–variance scaling relationships predicted by the model closely match the data: both the average mean–variance scaling relationship and the amount of variation around that relationship are consistent between the model and the data (Fig. S2). A likelihood-based goodness-of-fit statistic confirms that the model fits the data as well as if the model were true, regardless of whether the testing is done separately at the metacommunity level, or for each taxonomic group in the aggregate (Table S1).

The unified model also captures patterns of species association exceptionally well (Fig. 2): observed distributions of Horn similarity do not differ significantly from unified model predictions. Again, this conclusion holds regardless of whether one tests each metacommunity separately, or tests corals and fishes in the aggregate (Table S2).

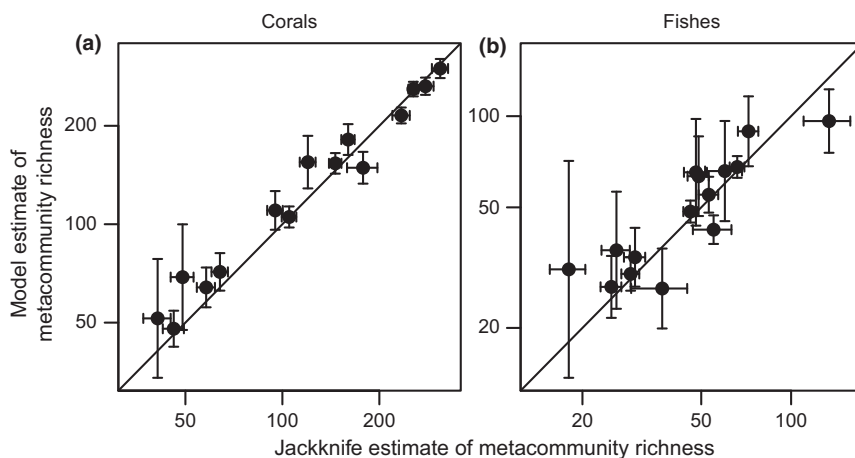
The model also produces robust estimates of species richness at both local and metacommunity scales. Specifically, the number of species observed in the metacommunity, but missing from any given site, is accurately estimated by the model (the leftmost, zero-class of abundances in Fig. 1, Fig. S1). Also, the model produces estimates of the total metacommunity richness (including unobserved species) that are highly consistent with non-parametric estimates (Fig. 3), and it reproduces the well-documented regional and between-habitat differences in species richness (Fig. S3). Moreover, the out-of-sample validation indicates that the model accurately predicted numbers of unshared species: residuals from the unity line did not differ significantly from zero for corals or fishes (Wilcoxon signed-ranks test,  $P = 0.27$  and  $0.57$  respectively; Fig. 4).

The individual components of the unified model also outperform several alternatives. Model selection strongly favours a negative binomial distribution for within species aggregation, compared with the alternatives we considered. There was overwhelming support for this distribution in most metacommunities, and there were no metacommunities where an alternative model was strongly supported. Thus, in the aggregate, support for the negative binomial distribution was unequivocal (Table S3). Similarly, the power-law model was much more strongly supported than the constant- $k$  model. The power law was very strongly favoured in the aggregate, and individually in 18 of the 30 metacommunities; conversely, the constant- $k$  model was strongly favoured in only three metacommunities (Table S4).

At the metacommunity scale, the unified model's lognormal distribution is more plausible than a gamma distribution. Considered individually, the lognormal has over 95% support in only three of the 30 metacommunities, with the rest giving very similar levels of support to both models. However, in the aggregate, the lognormal is strongly favoured (Table S5). In addition, a gamma metacommunity distribution predicts species richness values that are implausibly large, exceeding, in several cases, the global richness of reef corals or labrid fishes by several orders of magnitude (Fig. S4a and b). The gamma metacommunity model also fails to reproduce well-documented among-habitat and among-region biodiversity differences (Fig. S4c and d).



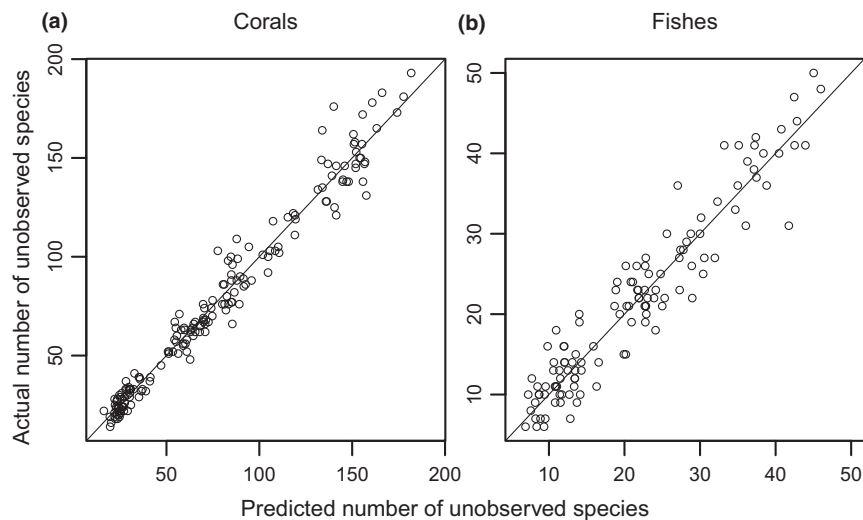
**Figure 2** Frequency distribution of Horn index community similarity values at the metacommunity scale for Indo-Pacific (a) corals, and (b) fishes. Similarity values are calculated for all pairs of sites from the same metacommunity. Bars represent the observed frequency distribution, and lines represent the mean over 1000 parametric bootstrap simulations of the unified model. Both observed and expected values are here pooled across metacommunities and habitats for plotting, but goodness-of-fit tests (Table S2) also were done separately for each metacommunity.



**Figure 3** Estimates of the metacommunity species pool size produced by the non-parametric jackknife estimate, compared with those produced by the unified model, for each metacommunity, for Indo-Pacific (a) corals, and (b) fishes. The solid line is the unity line, where the two estimates are identical. For the jackknife estimates, standard errors are produced using the method in ref. (Burnham & Overton 1979). For the unified model, standard errors are produced by parametric bootstrap.

The unified model also performed well when tested against the Great Barrier Reef fish data. The model successfully characterised local species-abundance patterns and mean–variance scaling relationships (Fig. S5, Fig. S6). Parametric bootstrap goodness-of-fit testing indicated that the model fits the data as well as if the model were true (Fig. S7; Table S6, S7). The model's estimates of metacommunity richness are consistent with non-parametric estimates (Fig. S8). The out-of-sample validation indicated good agreement between the model and the data: residuals from the unity line (observed = predicted) did not differ significantly from zero (Wilcoxon signed-ranks test,  $P > 0.67$ ; Fig. S9).

As with the Indo-Pacific data, model selection for Great Barrier Reef fishes supported the unified model. It favoured negative binomial aggregation over the alternatives in all metacommunities individually, and in the aggregate (Table S8). Power-law mean–variance scaling was favoured over the constant- $k$  model in the aggregate, but only slightly, with most metacommunities giving similar support to both models (Table S9). At the metacommunity scale, the lognormal is strongly favoured over the gamma distribution in four metacommunities, and in the aggregate (Table S10). Moreover, predictions of metacommunity richness from the gamma metacommunity model are as unrealistic as they are for the



**Figure 4** Out-of-sample test of unified model predictive ability, for Indo-Pacific (a) corals and (b) fishes. Predictions were made by fitting the unified model to all sites in a metacommunity except one, and then predicting the number of species in the data set that would be absent from the excluded site, using eqn 10. The solid line is the unity line.

Indo-Pacific data: they are more than an order of magnitude greater than those of the lognormal model and the non-parametric jackknife (Fig. S10).

## DISCUSSION

Recent tests of proposed unified theories of biodiversity (McGill 2010) have shown that such theories often fit individual patterns well, but fail when tested simultaneously against multiple patterns (Adler 2004; Dornelas *et al.* 2006; Xiao *et al.* 2015). Our model successfully captures several aspects of coral reef metacommunity structure, even though it, like unified theories, predicts a broad range of ecological patterns with few free parameters (two for the regional species-abundance distribution, and two power-law scaling parameters for aggregation). These patterns include classical local species-abundance distributions and mean-variance scaling relationships, but also the distribution of pairwise community similarity values, the proportion of species in the community matrix that are absent from any given site (the zero-class of abundances), metacommunity richness levels, and numbers of unshared species between sites.

The unified model explains more aspects of community structure than its four component elements (a lognormal metacommunity species-abundance distribution, power-law scaling of mean and spatial variance in abundance, negative binomial patterns of intraspecific aggregation and statistical independence of species spatial abundance patterns from one another) considered separately. For instance, none of those elements could independently fit a full sites-by-species matrix of abundances, predict species pool size at the metacommunity scale, or predict numbers of unshared species. Moreover, the fact that, in our model, interspecific variation in abundance follows a lognormal distribution, whereas spatial variation in abundance follows a negative binomial distribution (i.e. a poisson-gamma mixture distribution), implies strong differences in intraspecific and interspecific abundance

patterns, with interspecific variation exhibiting greater heterogeneity (i.e. a thicker tail of very highly abundant species) than intraspecific variation. Similarly, between species variance in mean abundance substantially exceeds spatial variance in abundance within species (by a median factor of 8).

Our unified model's predicted distributions of community similarity are much more consistent with the data than previous multi-scale tests of neutral theory on coral reefs, using three different models. Hubbell's original formulation of neutral theory could not reproduce the high variance observed in community similarity distributions (i.e. the spread of the histograms in Fig. 2), using the same reef coral data analysed here (Dornelas *et al.* 2006). A similar inability to reproduce high variance in community similarity distributions was found for fossil Caribbean reefs, using a spatially explicit neutral model (Bode *et al.* 2012). Volkov *et al.* (2007) attempted to explain the community similarity distributions in our data with an independent-species neutral model, where all species have the same density-independent per-capita birth and death rates, and the same immigration rates. However, that model implies community similarity values that are far too low to be realistic (Appendix S2).

A potential limitation of our approach is that the unified model is not explicitly process-based (i.e. derived explicitly from a metacommunity model whose parameters correspond to demographic rates of the constituent species). It is common in the species-abundance literature to classify species-abundance models as either process-based or mechanistic on one hand, or purely statistical or descriptive on the other (McGill *et al.* 2007; Harte 2011). One risk of conceptualising species-abundance models in this way is that it invites one to infer process parameter values (such as dispersal and speciation rates) from fits of process-based models to data. However, many such models produce very similar species-abundance distributions. For instance, several neutral models generate species-abundance distributions that are very well-approximated by a gamma distribution (Pueyo 2006; Connolly *et al.* 2014), but so does a process-based model in which species

respond differently to environmental fluctuations (Engen & Lande 1996a).

Conversely, many abundance distributions that are considered statistical or descriptive nonetheless have ecological explanations, and in some cases can be generated by explicitly process-based models, including the components of our unified model. For instance, the lognormal species-abundance distribution can be produced by stochastic community dynamics models that incorporate environmental stochasticity and interspecific variation in demographic rates (Saether *et al.* 2013). This class of models can also produce power-law spatial mean–variance scaling (Engen *et al.* 2008). Power-law mean–variance scaling can also be produced by other population-dynamic models with dispersal and demographic or environmental stochasticity; moreover, such models can simultaneously produce patterns of aggregation that are identical or similar to a negative binomial distribution (e.g. Anderson *et al.* 1982; Keeling 2000). Finally, an explanation need not be derived explicitly from a process-based model to be ecological. For instance, lognormal distributions of species-abundance can arise as a consequence of interspecific variation in niche hypervolumes (Sugihara *et al.* 2003; Connolly *et al.* 2009), or as a second-order mathematical approximation to arbitrary departures from pure demographic stochasticity (Pueyo 2006). Explanations of this kind have also been offered for power-law scaling (e.g. Eisler *et al.* 2008) and negative binomial patterns of aggregation (e.g. Boswell & Patil 1970).

The fact that different process-based models can generate the same distributions of species-abundance, and that some statistical models of abundance patterns can be generated by one or more process-based models, indicates that the kind of inferences that can be drawn from fits to these two classes of models are broadly similar. Good or bad fit of either type of model to species-abundance data provides information about the processes that could plausibly generate the observed patterns, and, conversely, can potentially falsify process-based explanations if models that include those processes cannot generate a distribution that fits the data well. For instance, neutral models consistently produce species-abundance distributions that are better-characterised by gamma than lognormal distributions (Pueyo 2006; Connolly *et al.* 2014). Consequently, our finding that lognormal abundance distributions remain superior to gamma distributions, even after accounting for the contribution of local aggregation to skew in species-abundance distributions at the local scale, indicates that neutral models omit key processes determining commonness and rarity of species on coral reefs. An additional question, given that our model unifies four well-known macroecological patterns, is whether unified model parameters covary across different taxa or assemblages from different types of ecosystems: for instance, do communities with stronger mean–variance scaling relationships tend to have more even or uneven abundance distributions? If so, do particular process-based models reproduce those patterns more readily than others?

Despite strong interest in applying macroecological theory to conservation problems (Hubbell 2001; Harte *et al.* 2009; McGill 2010; Economo 2011), there has been little uptake of unified theories in applied ecology, due in part to reservations about the restrictive suite of processes incorporated by such

theories, or their inability to make successfully predict multiple patterns simultaneously (Clark 2012; Driscoll & Lindenmayer 2012). However, applications of the unified model proposed here do not hinge on restrictive assumptions about which processes generate macroecological patterns, but only on the nature of the patterns themselves. This makes the unified model potentially useful in conservation applications. Existing approaches to predicting changes in abundance from occupancy, for instance, require substantial species-level baseline occupancy data for each species of interest (Sileshi *et al.* 2006; Hui *et al.* 2012). The unified model provides a way to overcome this limitation, by leveraging assemblage-level abundance information to infer abundance patterns of rare species, including species missing from field samples. Similarly, whole-assemblage approaches to conservation often rely heavily on patterns of biotic turnover, but estimates of beta diversity are notoriously biased by incomplete sampling (Chao *et al.* 2006). Our approach, however, produces estimates of biotic turnover (change in community composition and abundance) that explicitly account for incomplete sampling. Consequently, they can provide a foundation, for instance, for the more rigorous delineation of habitat and regional faunal boundaries in conservation priority-setting. Thus, the unified model offers a robust way to apply our understanding of macroecological regularities in nature to the regional-scale conservation and biodiversity management problems that are so pressing in the contemporary world.

#### ACKNOWLEDGEMENTS

For statistical programming assistance, we thank M. Hisano and L. Thibaut. For field assistance, we thank A. Baird, M. Boyle, H. Cornell, E. Dinsdale, A. Hoey, R. Karlson, M. Kasparov, P. Osmond, and J. Wolstenholme. We thank H. Sweatman and the AIMS Long-Term Monitoring Team for permission to use the Great Barrier Reef fish data. We also thank F. He, T. Northfield, and three anonymous reviewers for helpful comments and suggestions on an earlier version of this manuscript. This contribution was funded by the Australian Research Council (DP0880544 and CE140100020).

#### AUTHORSHIP

All authors conceived the study. T.H. collected the Indo-Pacific coral data, and D.B. collected the Indo-Pacific fish data. S.C. developed the unified model, and analysed the data. S.C. led the writing of the manuscript, with substantial input from T.H. and D.B.

#### DATA ACCESSIBILITY

Data available in the Dryad Digital Repository: doi:10.5061/dryad.2qp80

#### REFERENCES

- Adler, P.B. (2004). Neutral models fail to reproduce observed species-area and species-time relationships in Kansas grasslands. *Ecology*, 85, 1265–1272.



- Anderson, R.M., Gordon, D.M., Crawley, M.J. & Hassell, M.P. (1982). Variability in the abundance of animal and plant species. *Nature*, 296, 245–248.
- Bellwood, D., Hughes, T.P., Folke, C. & Nystrom, M. (2004). Confronting the coral reef crisis. *Nature*, 429, 827–833.
- Bode, M., Connolly, S.R. & Pandolfi, J.M. (2012). Species differences drive nonneutral structure in Pleistocene coral communities. *Am. Nat.*, 180, 577–588.
- Boswell, M.T. & Patil, G.P. (1970). Chance mechanisms generating the negative binomial distribution. In: *Random Counts in Scientific Work* (ed Patil, G.P.). Pennsylvania State University Press University, Park, pp. 3–22.
- Burnham, K.P. & Overton, W.S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60, 927–936.
- Callaway, R.M. (1997). Positive interactions in plant communities and the individualistic-continuum concept. *Oecologia*, 112, 143–149.
- Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics*, 62, 361–371.
- Clark, J.S. (2012). The coherence problem with the Unified Neutral Theory of Biodiversity. *Trends Ecol. Evol.*, 27, 198–202.
- Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005). Community structure of corals and reef fishes at multiple scales. *Science*, 309, 1363–1365.
- Connolly, S.R., Dornelas, M., Bellwood, D.R. & Hughes, T.P. (2009). Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology*, 90, 3138–3149.
- Connolly, S.R., MacNeil, M.A., Caley, J., Knowlton, N., Cripps, E., Hisano, M. *et al.* (2014). Commonness and rarity in the marine biosphere. *Proc. Natl Acad. Sci. USA*, 111, 8524–8529.
- Consul, P.C. & Jain, G.C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15, 791–799.
- Dornelas, M., Connolly, S. & Hughes, T. (2006). Coral reef diversity refutes the neutral theory of biodiversity. *Nature*, 440, 80–82.
- Driscoll, D.A. & Lindenmayer, D.B. (2012). Framework to improve the application of theory in ecology and conservation. *Ecol. Monogr.*, 82, 129–147.
- Economu, E.P. (2011). Biodiversity conservation in metacommunity networks: linking pattern and persistence. *Am. Nat.*, 177, E167–E180.
- Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC, Boca Raton, FL.
- Eisler, Z., Bartos, I. & Kertesz, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.*, 57, 89–142.
- Engen, S. & Lande, R. (1996a). Population dynamic models generating species abundance distributions of the gamma type. *J. Theor. Biol.*, 178, 325–331.
- Engen, S. & Lande, R. (1996b). Population dynamic models generating the lognormal species abundance distribution. *Math. Biosci.*, 132, 169–183.
- Engen, S., Lande, R., Walla, T. & DeVries, P.J. (2002). Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. *Am. Nat.*, 160, 60–73.
- Engen, S., Lande, R. & Saether, B.E. (2008). A general model for analyzing Taylor's spatial scaling laws. *Ecology*, 89, 2612–2622.
- Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943). 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.
- Harte, J. (2011). *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford University Press, Oxford, New York.
- Harte, J., Smith, A.B. & Storch, D. (2009). Biodiversity scales from plots to biomes with a universal species-area curve. *Ecol. Lett.*, 12, 789–797.
- He, F. & Gaston, K.J. (2003). Occupancy, spatial variance and the abundance of species. *Am. Nat.*, 162, 366–375.
- He, F.L. & Hubbell, S.P. (2011). Species-area relationships always overestimate extinction rates from habitat loss. *Nature*, 473, 368–371.
- Holt, A.R., Gaston, K.J. & He, F.L. (2002). Occupancy-abundance relationships and spatial distribution: a review. *Basic Appl. Ecol.*, 3, 1–13.
- Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Hughes, T.P., Bellwood, D.R., Connolly, S.R., Cornell, H.V. & Karlson, R.H. (2014). Double jeopardy and global extinction risk in corals and reef fishes. *Curr. Biol.*, 24, 1–6.
- Hui, C., Boonzaaier, C. & Boyero, L. (2012). Estimating changes in species abundance from occupancy and aggregation. *Basic Appl. Ecol.*, 13, 169–177.
- Jost, L., Chao, A. & Chazdon, R.L. (2011). Compositional similarity and beta diversity. In: *Biological Diversity: Frontiers in Measurement and Assessment* (eds Magurran, A.E., McGill, B.J.). Oxford University Press, Oxford, UK, pp. 66–84.
- Karlson, R.H., Connolly, S.R. & Hughes, T.P. (2011). Spatial variance in abundance and occupancy of corals across broad geographic scales. *Ecology*, 92, 1282–1291.
- Keeling, M.J. (2000). Simple stochastic models and their power-law type behaviour. *Theor. Popul. Biol.*, 58, 21–31.
- Kendall, D.G. (1949). Stochastic processes and population growth. *J. Roy. Stat. Soc.: Ser. B (Methodol.)*, 11, 230–282.
- Lawton, J.H. (1999). Are there general laws in ecology? *Oikos*, 84, 177–192.
- Loreau, M. (2010). *From populations to ecosystems: theoretical foundations for a new ecological synthesis*. Princeton University Press, Princeton.
- McGill, B.J. (2010). Towards a unification of unified theories of biodiversity. *Ecol. Lett.*, 13, 627–642.
- McGill, B.J., Hadly, E.A. & Maurer, B.A. (2005). Community inertia of quaternary small mammal assemblages in North America. *Proc. Natl Acad. Sci. USA*, 102, 16701–16706.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., *et al.* (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995–1015.
- Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M. *et al.* (2013). Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.*, 11, e1001569.
- Nachman, G. (2000). Effects of demographic parameters on metapopulation size and persistence: an analytical stochastic model. *Oikos*, 91, 51–65.
- Pueyo, S. (2006). Diversity: between neutrality and structure. *Oikos*, 112, 392–405.
- Saether, B.E., Engen, S. & Grotan, V. (2013). Species diversity and community similarity in fluctuating environments: parametric approaches using species abundance distributions. *J. Anim. Ecol.*, 82, 721–738.
- Sileshi, G., Hailu, G. & Mafongoya, P.L. (2006). Occupancy-abundance models for predicting densities of three leaf beetles damaging the multipurpose tree *Sesbania sesban* in eastern and southern Africa. *Bull. Entomol. Res.*, 96, 61–69.
- Smith, M.D. & Knapp, A.K. (2003). Dominant species maintain ecosystem function with non-random species loss. *Ecol. Lett.*, 6, 509–517.
- Sugihara, G., Bersier, L.F., Southwood, T.R.E., Pimm, S.L. & May, R.M. (2003). Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA*, 100, 5246–5251.
- Sweatman, H., Cheal, A., Coleman, G., Emslie, M., Johns, K., Jonker, M. *et al.* (2008). Long-term monitoring of the Great Barrier Reef, Status Report Number, 8, 369.
- Taylor, L.R. (1961). Aggregation, variance and mean. *Nature*, 189, 732–735.
- Ugland, K.I., Lamshead, F.J.D., McGill, B., Gray, J.S., O'Dea, N., Ladle, R.J. *et al.* (2007). Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evol. Ecol. Res.*, 9, 313–324.

- Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2007). Patterns of relative species abundance in rainforests and coral reefs. *Nature*, 450, 45–49.
- Xiao, X., McGlinn, D.J. & White, E.P. (2015). A strong test of the maximum entropy theory of ecology. *Am. Nat.*, 185, E70–E80.
- Zillio, T. & He, F.L. (2010). Modeling spatial aggregation of finite populations. *Ecology*, 91, 3698–3706.

Editor, Fangliang He

Manuscript received 14 October 2016

First decision made 23 November 2016

Second decision made 9 January 2017

Manuscript accepted 22 January 2017

#### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.