Ecological Indicators 18 (2012) 485–492

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

How many dimensions of biodiversity do we need?

Olga Lyashevska, Keith D. Farnsworth*

School of Biological Sciences, Queen's University Belfast, Belfast BT9 7BL, UK

ARTICLE INFO

Article history:
Received 17 February 2011
Received in revised form
13 December 2011
Accepted 22 December 2011

Keywords:
Shannon diversity
Species richness
Taxonomic
Functional
Phylogenetic diversity
Community composition
Taxonomic modelling

ABSTRACT

Biodiversity is a measure of the total difference within a biological system. It is understood to arise at genetic, species and multiple levels of community organisation, hence is multidimensional in nature. Biodiversity indices have proliferated in attempts to capture this complexity but may now have confounded it. Here we attempt a reduction to the minimal set of metrics needed to describe biodiversity (often by default taken to be species richness). 1000 model communities with realistic taxonomic composition were synthesised using databases of marine benthic species. A battery of 19 biodiversity indices were calculated for every community and analysed by PCA to show inter-dependence and sensitivity to variation in taxonomic (a surrogate for genetic), functional (based on ecological roles) and structural (based on species abundance) diversity. We found the three major axes of biodiversity were (a) structural complexity, and (b) two different mixtures of taxonomic and functional diversity: it was well approximated by a three-dimensional space of these variables. A scalar distance from the origin of this space could serve as a single valued summary where needed, for example in economic valuations. The most widely used single biodiversity measure – species richness – missed 88.6% of the diversity, emphasising the importance of additional characters and the need for species databases to record functional traits, presence and abundance in communities, and phylogenetic information.

1. Introduction

Two questions motivated the work presented here: firstly, can an optimal measure of biodiversity be constructed, and secondly, how closely is it approximated by the most commonly used measure—species richness? By optimal we mean capturing the maximum information about diversity in a compact form: we seek a measure with the maximum information density.

There is already plenty to choose from. The rapidly growing biodiversity literature offers a substantial 'lexicon zoo' (Marcot, 2007) of indices, leading some commentators to refer to a confusion of meaning (Hamilton, 2005) and to the presence of ambiguities (Weesie and van Andel, 2003). Biodiversity is often taken as a constellation of meanings which can never be captured by a single number (Purvis and Hector, 2000; Mayer, 2006; Failing and Gregory, 2003). This diversity of meanings encompasses a diversity of measures, each of them intended to represent some facet of total biodiversity. Examples include genetic and phenotypic variance, species numbers, ecosystem structural properties and patterns of functional heterogeneity. This proliferation calls us to rationalisation and synthesis: to identify which features of biodiversity are mathematically independent and thereby to find the irreducible set of metrics which must be included to encompass total biodiversity. Implied in that goal is the identification of redundant metrics; those which are so mutually correlated that any one of them may be taken to approximate the others.

One possible guide to the 'lexicon zoo' comes from recognising the hierarchical structure of biological diversity, ranging from the variety of genes within and among organisms to the variety of community structures, e.g. foodwebs. This immediately implies a way to categorise existing biodiversity measures by the organisational 'level' to which they refer. At each level, we can identify metrics by the specific kind of biological difference that they measure: we term this the 'descriptor'. With level (**L**) and descriptor (**D**), existing and hypothetical biodiversity measures can be classified in a (**D**|**L**) permutation matrix, each element of which is a different combination of the kind of biological diversity and the organisational level of its measurement. This constitutes a formalisation of the influential ideas presented by Noss (1990), based on primary attributes recognised by Franklin (1988), who incorporated the descriptor categories: composition, structure and function into a hierarchy of indices.

An obvious way to examine the multitude of biodiversity indices is by comparing candidate measures from a set of field-observed communities using multivariate statistics (e.g. Gallardo et al., 2011). A broad analysis would require a large set of studies following a consistent protocol, with matched sampling effort, spanning a wide variety of communities. In practice, there is only a little comparative data beyond species richness and abundance (Lamb et al., 2009; Geburek et al., 2010; Péru and Dolédec, 2010; Guo et al.,

* Corresponding author.
E-mail address: k.farnsworth@qub.ac.uk (K.D. Farnsworth).

1470-160X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2011.12.016

2010; Gallardo et al., 2011; Rubio et al., 2011), though an increasing recognition of functional (Mouchet et al., 2010), phylogenetic, and taxonomic (Lopez-Osorio and Miranda-Esquivel, 2010; Schweiger et al., 2008) diversity could change that.

Inconsistency among study designs (a few exceptions noted above) strongly limits any attempt to describe relationships among field measurements of biodiversity, so comparative meta-analysis has typically used simulated data instead. Studies using simulated data have sought to reveal relationships among indicators of taxonomic diversity (Clarke and Warwick, 1998, 2001), functional diversity (Villéger et al., 2008), species and genetic diversity (Vellend, 2005) and phylogenetic diversity (Schweiger et al., 2008; Nipperess et al., 2010). Here we unite these individual categories of biodiversity into one analysis examining the correlation amongst, and sensitivity of a set of taxonomic, structural and functional diversity indices simultaneously. This is not a trivial task as it requires realistic relations, among diversity categories, to be built into the test-data. Previous studies had examined indicators against unstructured random assemblies of 'species', rather than realistic simulations of communities. This severely limited the possibility of finding relations that cut across the descriptor categories of Franklin (1988) or those operating at different levels of biological organisation. To enable possible relationships among very different aspects of biodiversity to emerge, we built artificial communities with taxonomic structures and distributions of species traits that statistically matched an example of a near-shore temperate marine ecosystems.

Given the multidimensional nature of biodiversity, our search for the maximum information density involves defining a necessary and sufficient (irreducible) set of metrics which best approximate total biodiversity. This practically amounts to an ordination among measures, constructing principle axes of variation and interpreting them in biological terms. A single measure estimate may then be calculated as a distance metric in the reduced space of principle axes. Comparison between species richness and this composite measure will give an indication of how much biodiversity is missed by species richness.

2. Theory

Behind the myriad ways of measuring and describing biodiversity there is a simple unifying quantity. The word 'biodiversity' literally means the diversity within a biological system, where diversity quantifies the total difference among the system's parts. This definition coincides with the 'diaphoric definition of data' (Floridi, 2005) as the foundation of information, demonstrating the equivalence of biodiversity to information: not information *about* the system, but the natural information contained *within* it. Biodiversity, thus defined, is too much to be directly measurable; only aspects of it may be estimated by empirical indices, which may be used for comparing diversity among systems.

Biodiversity as 'difference' measures the total difference among the components of a biological system, where differences are defined on a set of axes, coinciding with the set of 'descriptors' **D**. We can envisage nine 'levels' **L** over which these descriptors may measure (see Table 1). Though there is no reason why in principal we should not measure, e.g. cell type diversity within an ecological community, in practice only the top three levels appear explicitly in the biodiversity literature, the lower levels being implicitly captured by genetic and functional differences among organisms.

We refer to elements of the matrix **D|L** as 'measures' of biodiversity. Measures may be combined in arbitrarily complex formulae to generate an unlimited set of biodiversity indices $\mathcal{I}$. For example, combining L = species and D = abundance produces indices such as: Simpson ($\mathcal{I}_1$), Shannon ($\mathcal{I}_2$) or Gini–Simpson ($\mathcal{I}_3$) indices:

$$\begin{aligned}\mathcal{I}_1 &= \sum_{i=1}^{S} p_i^2 \\ \mathcal{I}_2 &= -\sum_{i=1}^{S} p_i \log p_i \\ \mathcal{I}_3 &= 1 - \sum_{i=1}^{S} p_i^2,\end{aligned} \tag{1}$$

where $p_i = abundance(i)/\sum^S abundance$ for the ith species and S is the total number of species.

Table 1
A nine-level hierarchy of biocomplexity. Left column names the level of organisation and right column gives examples of the components from which diversity is created. Diversity at the community level arises from aggregating over all lower levels to form measures of temporal and spatial variation; genetic diversity generates all higher levels, notably by directing functional diversity.

Organization level	Diversity components
Ecological communities	Aggregate measures: α, β, γ-diversity
Populations	Abundances
Multi-cellular organisms	Species and their phylogenies
Tissues, organs and organ systems	Cell-interactions and organ function
Cells	Cell types and functions
Sub-cellular structures	Organelle specialisations
Molecular networks	Biochemical functions
DNA sequences: codons to genes	Genes and genetic networks
Molecular surfaces	Lock and key (e.g. enzyme) motifs

Since these indices are all represented by the same measure in the (**D|L**) matrix, they are constructed from the same data, so we would expect high correlation among them. Axiomatically, in general 'measures' are the empirical metrics from which all biodiversity indices are composed. Thus all indices found in the literature can be decomposed into their more fundamental 'measures' and these can be rearranged in arbitrary ways to construct new indices. New indices can then be designed, optimising them for information content. In practice, given a set of synthetic communities of known total difference in biodiversity, the task of defining an optimal index amounts to finding the one reporting the largest empirical diversity with the least number of measures. If diversity exists among d biological characteristics, then the index conveying diversity with greatest efficiency will be formed from a set of d orthogonal measures. Rank-ordering orthogonal axes of variation enables information density to be maximised by removing those axes with less than a statistically justified information content. We use principle components analysis (PCA) to do this, applying it to a population of artificial communities generated by (bootstrap-like) resampling of marine benthic-community data. In other words, we assess the sensitivity of indices to diversity within communities by measuring their sensitivity among communities.

3. Methods

3.1. Source data

Empirical data was obtained from the following sources:

- ITIS—the taxonomically structured species database (Bisby et al., 2009);
- BioMar—Irish benthic marine database (Picton et al., 1992);
- BIOTIC—Benthic biological traits information catalogue (BIOTIC, 2010)

Records from these datasets were merged into one database to create a source for artificial community generation. ITIS provided a species list from which to draw community members, BioMar provided taxonomic-distribution constraints and BIOTIC provided functional structure; all as detailed next.

3.2. Constructing simulated communities

First, following Storch and Sizling (2008) we relaxed the taxonomic unit invariance assumption, so that the distribution of number of daughter taxons within a parent was allowed to differ among taxonomic levels. Accordingly, we calculated the set of empirical probability distributions, denoted as: OiC, FiO, GiF, SiG, as they generated, respectively, the number of orders in a class, number of families in an order, number of genera in a family and number of species in a genus. (Note: the use of these distributions across all taxonomic units of equivalent level, tacitly assumes weak taxonomic invariance within taxonomic units (Storch and Sizling, 2008).)

To make taxonomically realistic communities, we followed the taxonomic sampling scheme described by Hillis (1998): First a taxonomic-tree topology was made for each community (using the empirical probability distributions: OiC, etc., to determine numbers of ramifications). Then species were selected by resampling with replacement from the ITIS database, until the topology was instantiated as a species list. Finally, species abundances were assigned by further resampling of ITIS, following a log-normal distribution, to give realistic abundance distributions. The following describes the community-simulation algorithm in more detail (see mathematical symbols in Appendix A).

3.2.1. Establishing taxonomic topology

1. The complete set of taxonomic classes in the BioMar database (total number 115) was identified as $\mathbf{C}_T$.
2. 1000 artificial communities were prepared as empty sets $\mathbf{W}_i$, each then being assigned a number, $N_c(i)$, of members of the class list $\mathbf{C}_T$, where $N_c = 1, \ldots, 115$. $N_c(i)$ was randomly generated for each community following a uniform distribution to give a spread of species richness among the synthetic communities.

The following steps were repeated for each community $\mathbf{W}_i$ in turn ($i = 1, \ldots, 1000$).

3. For each taxonomic class $C_j(i)$, ($j = 1, \ldots, N_C(i)$), of the community set of classes $\mathbf{C}_W(\mathbf{i})$, the number of taxonomic orders $N_O(i, j)$ in $C_j(i)$ was assigned by random sampling following the OiC (orders in class) distribution.
4. For each taxonomic order $O_k(i, j)$, ($k = 1, \ldots, N_O(i, j)$), in each class $C_j(i)$, of the community set of orders $\mathbf{O}_W(\mathbf{i})$, the number of families $N_F(i, j, k)$ in $O_k(i, j)$ was assigned by random sampling following the FiO (families in order) distribution.
5. For each family $F_m(i, j, k)$ ($m = 1, \ldots, N_F(i, j, k)$) in each order $O_k(i, j)$, of the community set of families $\mathbf{F}_W(\mathbf{i})$, the number of genera $N_G(i, j, k, m)$ in $F_m(i, j, k)$ was assigned by random sampling following the GiF (genera in family) distribution.
6. For each genus $G_g(i, j, k, m)$ ($g = 1, \ldots, N_G(i, j, k, m)$) in each family $F_m(i, j, k)$, of the community set of genera $\mathbf{G}_W(\mathbf{i})$, the number of species $N_S(i, j, k, m, g)$ in $G_g(i, j, k, m)$ was assigned by random sampling following the SiG (species in genera) distribution.

This resulted in 1000 community taxonomic tree topologies, each described by a set of numbers of ramifications at each taxonomic level: $N_x(i)$ where $x = C, O, F, G, S$.

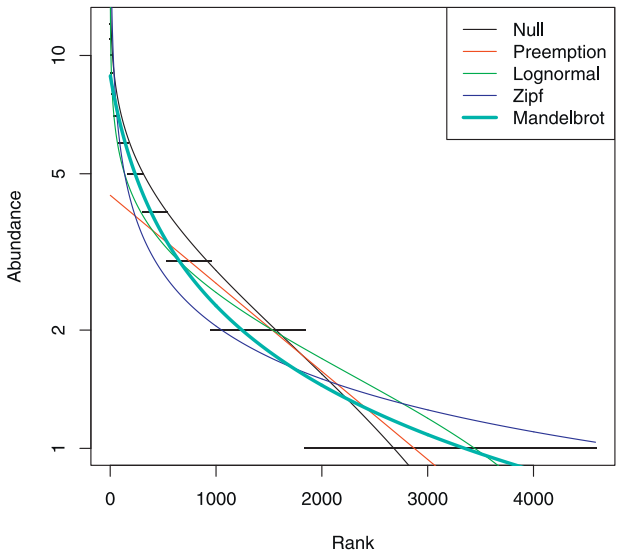

Fig. 1. Rank abundance distribution plot (Whittaker plot) showing logarithmic species abundances against species rank orders constructed for BioMar data. To analyse type of abundance distribution in communities, following Wilson (1991) several models were fitted. The horizontal bars are showing actual values. The simplest case is the null curve, where the individuals are randomly distributed among observed species, and there are no fitted parameters.

3.2.2. Establishing community from topology

Each community-tree was then instantiated by selecting species from the ITIS database, such that they fit into the taxonomic tree to give the correct number of each taxon in the community, using the following algorithm:

For each class $C_j(i)$ of community $W(i)$, select from ITIS a set of $N_O(i, j)$ orders which are members of that class. For each of these orders $O_k(i, j)$, select from ITIS a set of $N_F(i, j, k)$ families which are members of the order. For each of these families $F_m(i, j, k)$, select from ITIS a set of $N_G(i, j, k, m)$ genus which are members of the family. For each of these genus $G_g(i, j, k, m)$, select from ITIS a set of $N_S(i, j, k, m, g)$ species which are members of the genus.

In the above, all sampling followed a random uniform distribution, selecting by the database items index (ensuring unbiased and independent sample distributions). This resulted in a set of 1000 communities, each comprising a species list, differing in species richness and composition, but having distributions among taxa that match the BioMar dataset. The taxonomic richness-distribution (number of branches at each level) of BioMar data significantly differed from ITIS ($\chi^2 = 567$, $df = 389$), but the magnitude of this difference (never more than 60%) was small compared to the ratio in scale between the two data-sets, ITIS being approximately 50 times larger than BioMar. Since BioMar was consistently less rich than ITIS, the latter was always able to provide samples with a taxonomic richness-distribution representing that of BioMar. Each species carried 15 functional traits with it from the BIOTIC database (containing a total of 40 traits), so that the simulated communities also had a representative distribution of functional traits. Traits, combined into species-by-traits matrix, were used to calculate the functional diversity indices. Selected traits are listed in Appendix B.1 and include food type, maximum size, habitat, biogeographic range or biozone, coded as continuous, ordinal, nominal, or binary traits.

3.2.3. Assigning abundances

Finally each species in each community was assigned a population abundance, following the log-normal distribution Fig. 1. The original intention was to match species abundance to trophic level through the BIOTIC database traits, but we found only 2%

Table 2
Models parameters for rank abundance distribution plot.

	par1	par2	par3	Deviance	AIC	BIC
Null				2504.66	13428.37	13428.37
Preemption	0.00051608			842.30	11768.01	11774.44
Lognormal	0.42499	0.62712		517.93	11445.64	11458.50
Zipf	0.0052543	−0.44689		540.94	11468.64	11481.50
Mandelbrot	0.066681	−0.77576	212.56	182.16	11111.87	11131.16

overlap in trophic traits between BioMar and BIOTIC species, which was insufficient to create realistic correlation. Thus, in this instance we randomly allocated lognormal abundances to species, since this commonly fits empirical distribution of species abundance Wilson (1991) and fits the BioMar data best with two parameters (see Table 2).

3.3. Indices and measures of biodiversity

Biodiversity indices $\mathcal{I}$ were then calculated using the published algorithms `dbFD{FD}`(Laliberté and Shipley, 2010) and `taxondive{vegan}`(Oksanen et al., 2010) for functional and taxonomic diversity, respectively. Structural diversity indices were calculated directly from species abundance distributions. Definitions of indices are shown in Appendix C.1.

3.4. Biodiversity measures

Principal components were calculated from a standardised correlation matrix using a singular value decomposition of the centered data matrix (*prcomp*{*stats*} R Development Core Team, 2010). The dendogram was constructed from hierarchical clustering of biodiversity indices applied to the columns of the rotation matrix. Scalar distances were computed using Manhattan metrics to preserve orthogonal additivity (though Euclidean would be an acceptable alternative).

4. Results

4.1. Simulation

Spearman correlation coefficients for pairs of biodiversity indices (Fig. 2) revealed the near-perfect relationship between Simpson (SIMP) and Pielou (PIEL); Jaccard (JACC) and Sorensen (SORE) and Chao–Jaccard (CHJA) and Chao–Sorensen (CHSO) pairs, demonstrating mathematical redundancy among these indices. Since Shannon (SHAN) entropy showed strong negative (correlation with the Simpson–Pielou pair, it too enters a cluster of composition-based indices, along with Turnover, which was correlated with the Jaccard–Sorrensen pair. Perhaps the most surprising and important result shown in Fig. 2 is the very low correlation species richness (RICH) makes with the other indicators.

Taxonomic diversity indices were not correlated with either composition or function indices (ρ lies in the range $(-0.054; 0.072)$ among pairings which include taxonomic indices. Particular pairs of taxonomic indices were correlated with one-another (in particular taxonomic distinctness (DSTR) is correlated with most other taxonomic indices). Most taxonomic diversity indices showed some correlation with species richness (RICH), the largest, being for taxonomic diversity accounting for species richness (SPLU), where $\rho = 0.995$. Similarly, we found quite strong correlations among functional diversity indices, but not among pairs of function and any other sort of index.

The first three axis of the PCA (of 19 indices) accounted for 61.7% of total variation, the first five axes accounted for 82.3%. The community composition cluster of indices (see Fig. 4) contributed

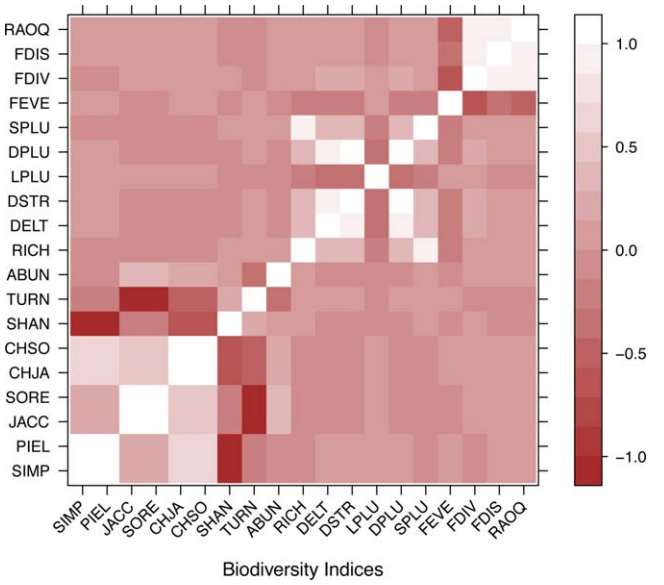

Fig. 2. The correlation between different biodiversity indices in a two-dimensional space represented as a heatmap. Dark shade gives negative correlation and light shade gives positive correlation.

almost exclusively to PC1, while taxonomic and functional diversity contributed approximately equally to PC2 and PC3 (Fig. 3). The largest single proportion of variance among the top three PCs was explained by the functional diversity cluster (Fig. 4) of indices.

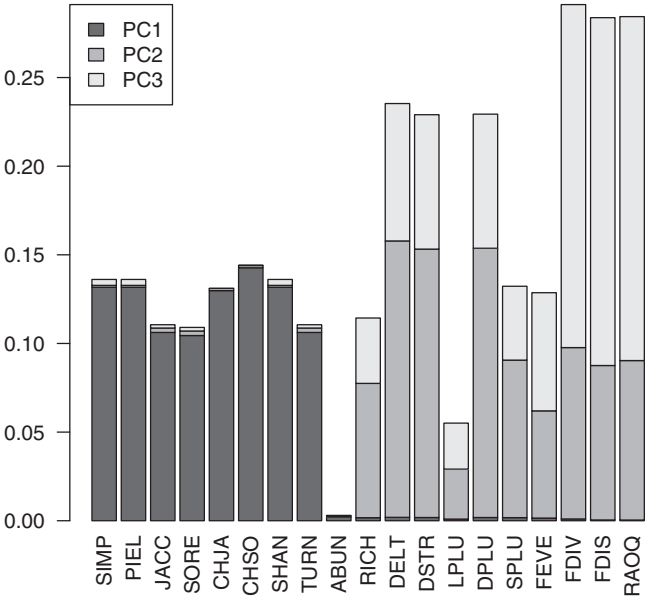

Fig. 3. Variance explained by each variable in three-dimensional space (61.7%). Different segments on the bar show contribution of the variables to corresponding principal components (see Appendix Table C.1 for key to index labels).

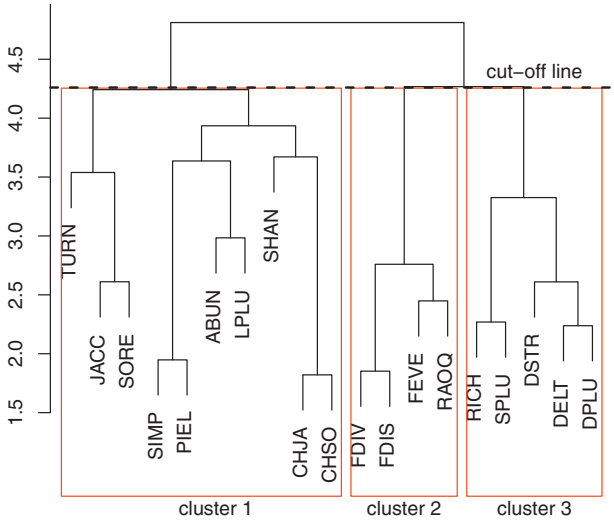

Fig. 4. Hierarchical clustering of biodiversity indices using Manhattan distance. This dendrogram is calculated using the rotation matrix of variables, the columns of this matrix contain eigenvectors (principal components). Depending on a cut off point several clusters can be observed: cluster 1, community structure and composition; cluster 2, functional diversity; cluster 3, taxonomic diversity.

This means that functional diversity was the most variable kind of biodiversity among simulated communities.

The proportion of total diversity estimated by species richness alone was very variable and typically low (Fig. 5). Inevitably, this single measure – species richness – gives a minimum estimate of total biodiversity, but the extent to which information is lost when this is the sole measure of diversity is striking.

5. Discussion

As conservation priorities move from single charismatic species to whole ecological communities and economists demand quantitative justifications for conservation, the need for a unifying measure of biodiversity mounts. At the start of this paper we asked how well the simplest, most commonly used biodiversity index – species richness – meets this demand. If biodiversity is truly

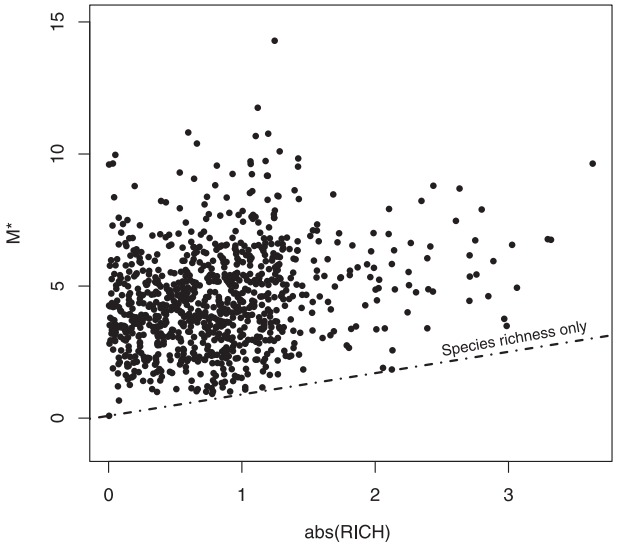

Fig. 5. The y-axis is scalar distance from the origin in the largest three PC's (M^*), plotted against species richness (RICH) for 1000 synthetic communities shown here rank-ordered by species richness from left to right. Dashed line shows the contribution of species richness alone on the y-axis.

the aggregate of functional, structural and taxonomic diversity, then Fig. 5 shows species richness to be missing a significant portion of the information. This happened because in our simulations, functional, structural and taxonomic varieties were not simple correlates of species richness, as indeed they are not in real life.

To explain this, we start with a simple 'null hypothesis' of random trait assignment, in which each species carries its own traits of taxonomic identity and functional role. In this model, as more species are randomly placed into a community, aggregate diversity increases. However, given a limited set of functional roles and using random selection from a species pool, functional diversity may be expected to rise asymptotically with species richness, saturating when the functional-set is complete. Only if the set-size is very large, will a near-linear increase with species richness be observed. Further, because taxonomic diversity is measured through distances on the taxonomic tree, random selection of species would confer highly unpredictable taxonomic diversity at low species richness, but would gradually converge onto the average taxonomic distance of the species-pool as richness (and therefore sample size) is increased. Finally, if relative species abundances were allocated at random, then structural diversity indicators would converge from great uncertainty at low abundance (small sample size) to some steady value (governed by the statistical distribution used to generate species abundances), for large communities. In short, even with random selection of species to construct communities, we would not expect species richness to make a good substitute for taxonomic, compositional or functional diversity (or all three).

Departing from earlier simulation studies where species had been selected at random (e.g. Clarke and Warwick, 1998, 2001; Villéger et al., 2008; Vellend, 2005; Nipperess et al., 2010), we composed model communities such that they reflected the higher-taxon composition of real communities, identified in the BioMar database. The resulting correlation among species taxa causes a narrowing of taxonomic diversity both within and among communities. Further, since real species were sampled, with their real functional traits attached, the functional diversity was also constrained and potentially correlated with taxonomic diversity. This is why it is significant that, as Fig. 2 shows, we found no evidence for correlation between functional and taxonomic indices. Although abundances generally reflect at least trophic structure, we were unable to obtain sufficient data to incorporate this into the synthetic communities. Instead, community composition indices were based on a lognormal abundance distribution that statistically matched the data (see Fig. 1), but with no correlation to functional traits.

Strikingly, Fig. 3 shows that species richness contributes almost nothing to the first principal axis, which is dominated by the community composition indices. Variation in all, but most notably compositional diversity indices, was independent of species richness. The lack of correlation between compositional and functional indices does not weaken that finding. Community ecologists have long known that species richness alone misses much diversity (e.g. Magurran, 2004; Wilsey et al., 2005), which is why compositional indices such as Simpson's are well supported. Also Cadotte et al. (2010) recognised that composition and phylogenetic diversity were largely independent and developed unifying indices for plant communities and whilst Gallardo et al. (2011) found correlation among some functional and compositional indices, they concluded that taxonomic, functional and compositional diversity each provided independent and useful information. In corroboration, when we aggregated these three categories of diversity (Fig. 5), we found that species richness no more than sets the lower limit of biodiversity, which varies above this limit in ways unrelated to species richness. So our first message must be that species richness, quick and simple though it is, turns out to be a rather poor estimate of biodiversity as we have defined it.

The intuition of Franklin (1988) and Noss (1990) that biodiversity is essentially three-dimensional, the axes being: structural, taxonomic and functional diversity is partially confirmed by our analysis. However, we also see some built-in correlations among these traits of biodiversity, so the axes are unlikely to be strictly orthogonal. In simulated communities, compositional aspects of biodiversity were found to be least correlated with function and taxonomy, but this is not surprising since abundances were generated by a process that was independent of the taxonomy and function. Correlations among various types of biodiversity indices have often been reported elsewhere (e.g. Mérigot et al., 2007; Heino, 2008). Winter et al. (2009) showed how correlation between species richness and taxonomic or phylogenetic indices depends on the taxonomic distance of introduced (or lost) species. Gallardo et al. (2011) found significant correlations among some biodiversity indicators, especially with Shannon diversity, but these were likely co-variates of environmental variation, particularly in relation to human disturbance. The communities generated by resampling in our work are of course independent of environmental conditions, so correlations among indices must reflect underlying structure. These results raise important questions about priorities in biodiversity measurement. If function, taxonomy (or better still phylogeny) and community composition are substantially independent, then any observed community structure may be produced from a wide variety of species (implying species substitutability)—and vice versa. Further, many different communities, with different species compositions, could perform equivalent functions (also implying species substitutability). Thus the strength of inherent correlations among the three major categories of biodiversity sheds light on species substitutability.

By defining biodiversity as the total difference among a biological system's parts we were able to see it as information *embodied within* the system, rather than merely empirical information *about* the system. This gives a different perspective from the more usual view of biodiversity simply as an indicator of ecological change. In our more ontological view, biodiversity is the information required to fully describe or reproduce the community as a living complex system. As such, it is very much greater than we can measure with biodiversity indices—these are merely comparative estimators. Diversity, as a quantification of difference, is multidimensional: each dimension being an axis of variation in the system. Even species themselves differ in an uncounted diversity of ways, making for a very large and unknown dimensionality to total biodiversity. Existing indices provide transect projections and cross-sectional views to sample this multi-dimensional space. The most efficient sampling would be achieved by a set of orthogonal estimators, in particular those projecting along the major axes of variation: lines of greatest variance in diversity space. By deconstructing existing indices into their 'level' and 'descriptor' components, we have shown it is possible to identify the available projections as 'filled elements' in the permutation matrix of all possible level and descriptor pairs. Statistical ordination can then identify the desired orthogonal set of major axes, given a suitable data source.

The practical consequence of our analysis is a parsimonious one. Faced with the urgent need to describe the rapidly declining diversity of life on earth, as comprehensively as possible but with limited resources, we see that no more than three well chosen indices are necessary. In the extreme of emergency cataloguing, we find that the simplest of all indices – species richness – performs poorly as a single surrogate for the three aspects of biodiversity, but of course it still may be the only practical option. When species, their phylogeny and significant functional traits are catalogued together in accessible databases, then field-collected species lists will serve as a key to estimating biodiversity in its fuller meaning. The need for this development sets an urgent goal for future biodiversity action.

Acknowledgement

This work is supported by the Environmental Protection Agency of the Republic of Ireland under Science Technology Research and Innovation for the Environment programme (2007-PhD-SD-3).

Appendix A. List of symbols

See Table A.1.

Appendix B. Functional traits

See Table B.1.

Appendix C. Biodiversity indices

See Tables C.1–C.3.

Table A.1
Mathematical symbols used in the text.

$\mathbf{D}$	The set of biodiversity descriptors
$\mathbf{L}$	The set of biodiversity levels
D	A particular biodiversity descriptor
L	A particular biodiversity level
$\mathcal{I}$	A biodiversity index
M^*	Scalar distance in principal-axis space, summarising total biodiversity
$\mathbf{W}$	Set of communities
$\mathbf{W}_i$	The ith community
$\mathbf{C}_T$	Classes in Biomar
$N_c(i)$	Number of classes in ith community
$\mathbf{C}_W(i)$	The set of classes in community i
$C_j(i)$	The jth class of the ith community
$N_O(i,j)$	Number of orders in the jth class of ith community
$O_k(i,j)$	The kth order of the jth class of ith community
$\mathbf{O}_W(i)$	The set of orders in ith community
$N_F(i,j,k)$	Number of families in kth order of jth class in ith community
$F_m(i,j,k)$	mth family in kth order of jth class in ith community
$\mathbf{F}_W(i)$	The set of families in ith community
$N_G(i,j,k,m))$	Number of genus in mth family of the kth order of jth class in ith community
$G_g(i,j,k,m)$	gth genus in mth family of kth order of jth class in ith community
$\mathbf{G}_W(i)$	The set of genus in ith community
$N_S(i,j,k,m,g))$	Number of species in gth genus of mth family of the kth order of jth class in ith community

Table B.1
Functional traits and their values.

Traits	Values
Food type	Zooplankton, phytoplankton, detritus, suspended particles
Size	1–50 cm
Habitat	Free living, attached, erect
Regeneration	Yes/no
Life span	1–100 years
Reproduction frequency	Annual/biannual, protracted/episodic
Fertilisation type	External/internal
Biogeographic range	Cold/temperate
Depth range	0–1765 m
Biozone	Littoral/pelagic
Environmental position	Epifaunal, epifloral, demersal, pelagic
Feeding method	Herbivore, predator, scavenger, suspension feeder
Growth form	Radial, stellate, turf
Mobility	Crawler, drifter, swimmer
Reproduction type	Vegetative, budding, self-fertilisation

Table C.1
Community composition indices.

Composition		
Simpson		SIMP
Pielou		PIEL
Jaccard		JACC
Sorensen		SORE
Chao–Jaccard		CHJA
Chao–Sorensen		CHSO
Shannon		SHAN
Turnover		TURN
Abundance		ABUN
Richness		RICH
Taxonomic diversity		
Taxonomic diversity	Clarke and Warwick (1998)	DELT
Taxonomic distinctness	Clarke and Warwick (1998)	DSTR
Variation in taxonomic distinctness	Clarke and Warwick (2001)	LPLU
Taxonomic diversity for presence/absence	Clarke and Warwick (1998)	DPLU
Taxonomic diversity accounting for species richness	Clarke and Warwick (2001)	SPLU
Functional diversity		
Functional evenness	Villéger et al. (2008)	FEVE
Functional divergence	Mason et al. (2003) and Villéger et al. (2008)	FDIV
Functional dispersion	Anderson (2006) and Laliberté and Legendre (2010)	FDIS
Rao quadratic entropy	Rao (1982) and Botta-Dukát and Wilson (2005)	RAOQ

Table C.2
Taxonomic diversity and distinctness indices. Where $x_i (i = 1, \ldots, s)$ is the abundance (presence/absence for Δ^+), n is the total number of individuals in the community and ω_{ij} is the distinctness weight of the path length linking any two entities.

Index	Formula
Taxonomic diversity	$\Delta = \frac{\sum\sum_{i<j} \omega_{ij} x_i x_j}{n(n-1)/2}$
Taxonomic distinctness	$\Delta^* = \frac{\sum\sum_{i<j} \omega_{ij} x_i x_j}{\sum\sum_{i<j} x_i x_j}$
Variation in taxonomic distinctness	$\Lambda^+ = \frac{\sum\sum_{i<j} \omega_{ij}^2}{n(n-1)/2} - (\Delta^+)^2$
Average taxonomic distinctness	$\Delta^+ = \frac{\sum\sum_{i<j} \omega_{ij} x_i x_j}{n(n-1)/2}$

Table C.3
Functional diversity indices.

Index	Formula
Functional dispertion	FDis $= \sum(a_j z_j)/\sum(a_j)$, where a_j abundance of species j, z_j is the distance of species j to the weighted centroid c
Functional evenness	FEve $= \frac{\sum_{LC=1}^{nLC-1} \min(PEW_{LC}, (1/(nLC-1))) - (1/(nLC-1))}{1-(1/(nLC-1))}$ PEW_{LC} is a partial weighted evenness $PWE_{LC} = \frac{EW_{LC}}{\sum_{LC=1}^{nLC-1} EW_{LC}}$ $EW_{LC} = (dist(i,j)/(\omega i + \omega j))$ with i and j being a pair of taxonomic units and ω their weight.
Functional divergence	FDiv $= (\Delta d + \overline{dG})/(\Delta\lvert d\rvert + \overline{dG})$; Δd is the sum of abundance-weighted deviances: $\Delta d = \sum_{i=1}^{S} \omega_i x(\Delta G_i - \overline{dG})$ $\Delta\lvert d\rvert$ is absolute abundance-weighted deviance; $\overline{dG}$ is the mean euclidean distance to the center of gravity: $\overline{dG} = (1/S)\sum_{i=1}^{S} \Delta G_i$
Rao quadratic entropy	$Q = \sum_{i=1}^{nLC-1} \sum_{j=i+1}^{nLC} d_{ij} p_i p_i$ the relative abundances of species p (or other taxonomic units) and a measure of the pairwise functional differences between them—d_{ij}

References

Anderson, M.J., 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62 (1), 245–253.

BIOTIC, 2010. The Biological Traits Information Catalogue (Digital resource at http://www.marlin.ac.uk).

Bisby, F., Roskov, Y., Orrell, T., Nicolson, D., Paglinawan, L., Bailly, N., Kirk, P., Bourgoin, T., Baillargeon, G. (Eds.), 2009. Species 2000 and ITIS Catalogue of Life: 2009 Annual Checklist. Species 2000, Reading, UK (Digital resource at http://www.catalogueoflife.org/annual-checklist/2009/).

Botta-Dukát, Z., Wilson, J.B., 2005. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. J. Veg. Sci. 16 (5), 533–540.

Cadotte, M.W., Davies, T.J., Regetz, J., Kembel, S.W., Cleland, E., Oakley, T.H., 2010. Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. Ecol. Lett. 13 (1), 96–105.

Clarke, K.R., Warwick, R.M., 1998. A taxonomic distinctness index and its statistical properties. J. Appl. Ecol. 35 (4), 523–531.

Clarke, K.R., Warwick, R.M., 2001. A further biodiversity index applicable to species lists: variation in taxonomic distinctness. Mar. Ecol. Prog. Ser. 216, 265–278.

Failing, L., Gregory, R., 2003. Ten common mistakes in designing biodiversity indicators for forest policy. J. Environ. Manage. 68 (2), 121–132.

Floridi, L., 2005. Is semantic information meaningful data? Philos. Phenom. Res. 70 (2), 351–370.

Franklin, J., 1988. Biodiversity. National Academy Press, Washington, DC (Ch. Structural and functional diversity in temperate forests., pp. 166–175).

Gallardo, B., Gascón, S., Quintana, X., Comín, F.A., 2011. How to choose a biodiversity indicator—redundancy and complementarity of biodiversity metrics in a freshwater ecosystem. Ecol. Indic. 11 (5), 1177–1184.

Geburek, T., Milasowszky, N., Frank, G., Konrad, H., Schadauer, K., 2010. The Austrian forest biodiversity index: all in one. Ecol. Indic. 10 (3), 753–761.

Guo, Q., Ma, K., Yang, L., Cai, Q., He, K., 2010. A comparative study of the impact of species composition on a freshwater phytoplankton community using two contrasting biotic indices. Ecol. Indic. 10 (2), 296–302.

Hamilton, A.J., 2005. Species diversity or biodiversity? J. Environ. Manage. 75 (1), 89–92.

Heino, J., 2008. Patterns of functional biodiversity and function–environment relationships in lake littoral macroinvertebrates. Limnol. Oceanogr. 53, 1446–1455.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47 (1), 3–8.

Laliberté, E., Legendre, P., 2010. A distance-based framework for measuring functional diversity from multiple traits. Ecology 91, 299–305.

Laliberté, E., Shipley, B., 2010. FD: Measuring Functional Diversity from Multiple Traits, and Other Tools for Functional Ecology, R Package version 1.0-9.

Lamb, E.G., Bayne, E., Holloway, G., Schieck, J., Boutin, S., Herbers, J., Haughland, D.L., 2009. Indices for monitoring biodiversity change: are some more effective than others? Ecol. Indic. 9 (3), 432–444.

Lopez-Osorio, F., Miranda-Esquivel, D.R., 2010. A phylogenetic approach to conserving Amazonian biodiversity. Conserv. Biol. 24 (5), 1359–1366.

Magurran, A., 2004. Measuring Biological Diversity. Blackwell Publishing.

Marcot, B.G., 2007. Biodiversity and the lexicon zoo. For. Ecol. Manage. 246 (1), 4–13.

Mason, N.W.H., MacGillivray, K., Steel, J.B., Wilson, J.B., 2003. An index of functional diversity. J. Veg. Sci. 14 (4), 571–578.

Mayer, P., 2006. Biodiversity—the appreciation of different thought styles and values helps to clarify the term. Restor. Ecol. 14 (1), 105–111.

Mérigot, B., Bertrand, J., Mazouni, N., Manté, C., Durbec, J.-P., Gaertner, J., 2007. A multi-component analysis of species diversity of groundfish assemblages on the continental shelf of the Gulf of Lions (north-western Mediterranean Sea). Estuar. Coast. Shelf Sci. 73, 123–126.

Mouchet, M.A., Villéger, S., Mason, N.W.H., Mouillot, D., 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Funct. Ecol. 24 (4), 867–876.

Nipperess, D.A., Faith, D.P., Barton, K., 2010. Resemblance in phylogenetic diversity among ecological assemblages. J. Veg. Sci. 21 (5.).

Noss, R.F., 1990. Indicators for monitoring biodiversity—a hierarchical approach. Conserv. Biol. 4 (4), 355–364.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2010. vegan: Community Ecology Package, R Package Version 1.17-2.

Péru, N., Dolédec, S., 2010. From compositional to functional biodiversity metrics in bioassessment: a case study using stream macroinvertebrate communities. Ecol. Indic. 10 (5), 1025–1036.

Picton, B., Emblow, C., Morrow, C., Sides, E., Tierney, P., McGrath, D., McGeough, G., McCrea, M., Dinneen, P., Falvey, J., Dempsey, S., Dowse, J., Costello, M.J., 1992. Marine Sites, Habitats and Species Data Collected during the BioMar Survey of Ireland (Digital Resource).

Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. Nature 405 (6783), 212–219.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theor. Popul. Biol. 21 (February (1)), 24–43.

Rubio, A., Gavilán, R.G., Montes, F., Gutiérrez-Girón, A., Díaz-Pines, E., Mezquida, E.T., 2011. Biodiversity measures applied to stand-level management: can they really be useful? Ecol. Indic. 11 (2), 545–556.

Schweiger, O., Klotz, S., Durka, W., Kühn, I., 2008. A comparative test of phylogenetic diversity indices. Oecologia 157 (3), 485–495.
Storch, D., Sizling, A.L., 2008. The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? Folia Geobot. 43 (3), 329–344.
Vellend, M., 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. Am. Nat. 166 (2), 199–215.
Villéger, S., Mason, N.W.H., Mouillot, D., 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology 89 (8), 2290–2301.
Weesie, P., van Andel, J., 2003. On biodiversity and its valuation. In: The CDS Research Report Series. Centre of Development Studies, The University of Groningen, the Netherlands.
Wilsey, B.J., Chalcraft, D.R., Bowles, C.M., Willig, M.R., 2005. Relationships among indices suggest that richness is an incomplete surrogate for grassland biodiversity. Ecology 86 (5), 1178–1184.
Wilson, J.B., 1991. Methods for fitting dominance/diversity curves. J. Veg. Sci. 2 (1), 35–46.
Winter, M., Schweiger, O., Klotz, S., Nentwig, W., Andriopoulos, P., Arianoutsou, M., Basnou, C., Delipetrou, P., Didžiulis, V., Hejda, M., Hulme, P.E., Lambdon, P.W., Pergl, J., Pyšek, P., Roy, D.B., Kühn, I., 2009. Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. Proc. Natl. Acad. Sci. U. S. A. 106 (51), 21721–21725.