Diversity and Distributions A Journal of Conservation Biogeography

BIODIVERSITY RESEARCH

Species–accumulation curves and taxonomic surrogates: an integrated approach for estimation of regional species richness

Antonio Terlizzi[1]*, Marti J. Anderson[2], Stanislao Bevilacqua[1] and Karl I. Ugland[3]

[1]*Laboratorio di Zoologia e Biologia Marina, Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, CoNiSMa, Lecce, 73100, Italy,* [2]*New Zealand Institute for Advanced Study (NZIAS), Massey University, Albany Campus, Private Bag 102 904, Auckland, 0745, New Zealand,* [3]*Marine Biology Research Group, Department of Biology, University of Oslo, Pb 1066 Blindern, Oslo, 0316, Norway*

*Correspondence: Antonio Terlizzi, Laboratory of Zoology and Marine Biology, Department of Biological and Environmental Science and Technologies (DiSTeBA), University of Salento, I-73100 Lecce, Italy.
E-mail: antonio.terlizzi@unisalento.it

ABSTRACT

Aim A species–accumulation curve may represent a direct expression of β-diversity, the rate at which diversity increases from local to regional scale. Patterns of variation in β-diversity tend to be consistent when measured across lower levels of the Linnaean taxonomic hierarchy (i.e. using species, genera or families). Our aim was to assess the relationships between species–accumulation curves and β-diversity at different taxonomic levels and to combine the logic of species–accumulation curves with taxonomic surrogacy to provide a new approach for cost-effective and reliable estimates of large-scale species richness (γ-diversity).

Location Mediterranean, N Atlantic and SW Pacific.

Methods We provide here a novel framework to extrapolate quantitative measures of species richness in large areas from accumulation curves based on extensive sampling at the family level coupled with estimation of species-to-family ratios from a subset of sampling units where organisms are identified to the species level. We demonstrated the effectiveness of the approach by analysing six datasets of diverse marine molluscan assemblages from different biogeographical regions and habitat types.

Results The approach proposed here can be used successfully to gain substantial efficiencies in sampling, potentially reducing the number of sampling units in which organisms have to be identified at species level between 50 and 75%, while still allowing reliable estimates of regional species richness.

Main conclusions Our results highlight the potential of this approach to improve the general exploration of biodiversity, especially for large-scale monitoring programs. The method we propose differs from previously described approaches by taking into account the spatial heterogeneity of species distributions within the sampled area and also by relying on estimates of species-to-family ratios obtained directly from the specific area of interest.

Keywords

Biodiversity monitoring and conservation, molluscs, species–accumulation curves, taxonomic surrogates, β-diversity, γ-diversity.

INTRODUCTION

Quantifying biodiversity in a given geographical region has long fascinated ecologists (Hutchinson, 1959; May, 1988). The original interest in advancing baseline knowledge of species diversity has also moved towards more practical intents, such as mitigation strategies, given the rapid world-wide increase in anthropogenic threats to biodiversity (Vitousek

DOI: 10.1111/ddi.12168
http://wileyonlinelibrary.com/journal/ddi

et al., 1997; Halpern *et al.*, 2008) and potential attendant depletion of ecosystem function (Worm & Duffy, 2003; Balmford & Bond, 2005).

Historically, biodiversity monitoring programmes have gathered large quantities of data, but often with an absence of clearly stated objectives (Yoccoz *et al.*, 2001). Moreover, information on species diversity is generally delivered at a rate that is not compatible with the urgent necessity to provide rapid cost-effective methods to assess human impacts (Wheeler *et al.*, 2004). Endeavours to collect detailed information on biodiversity may even be perceived as an outdated and unnecessary luxury (Beattie & Oliver, 1994) compared with pressing concerns arising under the current 'biodiversity crisis' (Worm & Duffy, 2003).

Beyond fragmented and limited information on the distributions of species and their roles within communities (Whittaker *et al.*, 2005), we still do not even know how many species exist on the planet. The total number of species currently described is approximately 1.5 million, which is paltry by comparison with recent estimates of a probable total of around 7–10 million existing species (Mora *et al.*, 2011). Most known species belong to well-studied groups, such as vascular plants, whereas the diversity of other groups such as marine invertebrates, remains largely unexplored (Cardoso *et al.*, 2011). An almost complete map of global biodiversity might be feasible (Wilson, 2004), but advancements in the exploration of biodiversity are hampered by the ongoing crisis of insufficient knowledge and expertise in systematics and taxonomy (Wheeler *et al.*, 2004).

Quantifying species richness at a regional or local scale can be problematic, as a consequence of the time and cost associated with taxonomic identification and a chronic lack of taxonomic expertise. An expedient means to overcome these problems is to use higher taxonomic ranks of the Linnaean hierarchy, rather than species (e.g. Williams & Gaston, 1994; Balmford *et al.*, 1996; Mazaris *et al.*, 2010). In this approach, a regression model of the number of species to the number of genera or families is fitted based on data from intensively sampled areas and then used to estimate the number of species in other areas where only genera or families have been recorded (Gaston & Williams, 1993; Andersen, 1995; Balmford *et al.*, 2000). This method assumes, however, that correlations between the number of species and the number of higher taxa from a given sampled area are consistent throughout the study area, which may not be realistic (Lewandowski *et al.*, 2010; Smale, 2010; Sutcliffe *et al.*, 2012). Predictable patterns in the taxonomic classification of species have also been used, more recently, to allow the total number of species within taxonomic groups to be estimated, allowing estimates of the total number of species globally (Mora *et al.*, 2011). Applying this latter approach to more limited spatial extents would be problematic, however, because the rate at which species and higher taxa are discovered within a limited area may not reflect broader-scale patterns in taxonomic structure and total species richness.

Another potential cost-saving approach for quantifying biodiversity across large areas is to use species richness estimators (Colwell & Coddington, 1994) based on a limited but representative collection of sampling units. This approach extrapolates species richness from species–accumulation curves, which are plots of the cumulative number of species versus increasing numbers of units. Although the accuracy of such tools is still under debate (e.g. He & Hubbell, 2011), some have been demonstrated to provide reliable estimates of total species number (Reichert *et al.*, 2010). For example, in the approach described by Ugland *et al.* (2003), an overall total–species (T–S) curve is obtained using only the terminal points of a set of species–accumulation curves from subareas within the total area of study (i.e. the points corresponding to the total species richness estimated at increasing number of subareas). In contrast to the traditional method, this procedure takes into account not only the variation in species richness at small scales (i.e. the scale of individual sampling units), but also the potential heterogeneity in species identities among subareas within the total area sampled (e.g. Matias *et al.*, 2011). Moreover, it has been shown that this approach provides the most accurate estimate of total richness out of a suite of classical estimation methods (Reichert *et al.*, 2010).

Importantly, the T–S curve of Ugland *et al.* (2003) estimates species richness, while also including and allowing for variation in the shapes of accumulation curves across the total area. The regression coefficient of the T–S curve measures how fast the number of species increases with the (log) number of sampling units; thus, variations in the coefficient are highly likely to be aligned with variations in β-diversity among subareas within the total area. Beta-diversity can be expressed in terms of non-directional changes in species composition (heterogeneity in identities of species) among sampling units within a given spatial, temporal or environmental extent (Anderson *et al.*, 2011). This concept of β-diversity is at the core of many of the main theoretical attempts to model spatial patterns of species distributions (MacArthur & Wilson, 1967; Nekola & White, 1999) having a central role in linking local and regional diversity (Witman *et al.*, 2004). As patterns of heterogeneity in identities of species (β-diversity) may be maintained at coarser levels of taxonomic resolution up to families (Terlizzi *et al.*, 2009), we hypothesize that there may be an intimate relationship between patterns of variation in the coefficient of the T–S curve and β-diversity at different levels of taxonomic resolution. We anticipate that estimated parameter coefficients of the T–S curve calculated at each level in the taxonomic hierarchy can be strongly correlated with the estimated β-diversity in a given area.

Here, we propose that the estimation of species richness for a given region might be carried out effectively by combining the idea of taxonomic surrogacy with the use of species–accumulation curves. We explored the potential to derive reliable estimates of regional species richness by extending the logic of T–S curves to taxonomic surrogates. Our approach both takes into account the spatial heterogeneity of species distributions within the sampled area and also exploits estimates of species-to-family ratios that are appropriately representative and derived from the specific area of interest alone.

We developed a framework with three essential steps. First, a T–S curve was obtained for a region at the family level (which might be referred to as a 'T–F curve'), using the full set of available sampling units. Second, a quantitative relationship between species-level and family-level richness was modelled (namely, a species-to-family ratio) using only a random subset of sampling units from the region for which organisms have been identified at the species level. Note that this differs from earlier approaches (e.g. Gaston & Williams, 1993), in which a simple correlation between species richness and higher taxon richness is estimated from some intensively sampled reference areas and assumed to be representative of the total area of interest. In our approach, instead, estimates of the species-to-family ratio are based on samples from the whole area of interest. Finally, the species-level richness for the region was estimated by combining the information given in the first two steps; namely, the estimated number of species can be calculated as the estimated number of families times the estimated species-to-family ratio.

We assessed the effectiveness of our proposed method for datasets of marine molluscan assemblages from different biogeographical regions and habitat types, including the Mediterranean Sea, the North Atlantic and the South Pacific Ocean. Specifically, we compared results obtained using our new approach (i.e. T–F curves and estimated species-to-family ratios for a given area), calculated from random representative subsets of data of different sizes, with what would be obtained if all units were identified to species level and T–S curves had been used.

METHODS

Study areas and datasets

Six datasets from previous studies investigating spatial patterns of macrofaunal assemblages were analysed to explore the correlation between the T–S curve coefficient and β-diversity at decreasing taxonomic resolution, and the potential to derive reliable estimates of regional species richness based on family-level data. Datasets were gathered in each of six marine regions (Fig. 1), having different areal extents, sizes of individual sampling units and habitat types. These were mud flats of the Norwegian Continental Shelf (A), coastal sand/mud bottoms of the southern Irish Sea (B), coastal sandy detritic/mud bottoms (C) and rocky cliffs (D) of the north Ionian Sea, coastal sand/mud bottoms of the southern China Sea (E) and (F) kelp forests of the south-west Pacific (see Table S1 in Supporting Information). Although all datasets included several different invertebrate phyla, we focused on one of the most widespread and diverse, the Mollusca, for which all six datasets included complete taxonomic identification of all individuals down to the species level. The effectiveness of families in depicting patterns of α- and β-diversity has already been assessed for most of these datasets (see for details Bevilacqua *et al.*, 2009; Terlizzi *et al.*, 2009).

Estimation of species-level richness using T–S curves

Consider a set of $i = 1,\ldots, U$ units taken from a region and suppose there are $k = 1,\ldots, S$ species that occur variously in those sampling units across the whole collection of U units. Next consider a random sample of size u and let the number

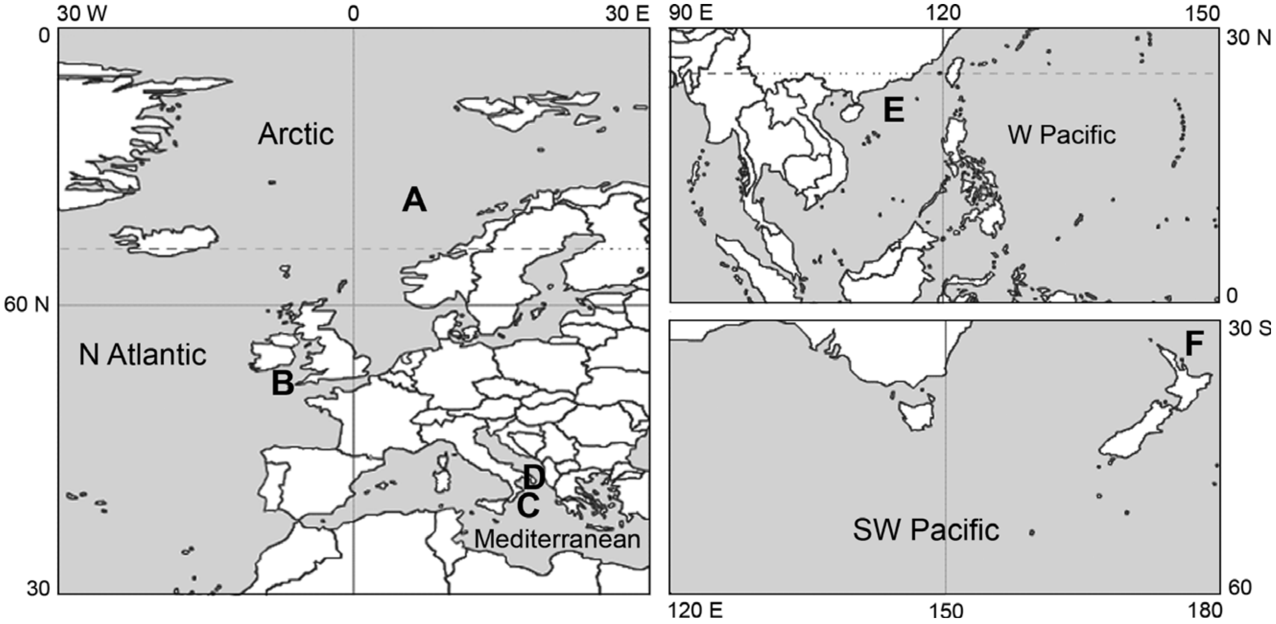

Figure 1 Map showing the regions from which datasets were obtained: (A) Norwegian Continental Shelf, (B) southern Irish Sea, (C) Ionian Sea soft bottoms, (D) Ionian Sea rocky cliffs, (E) southern China Sea and (F) New Zealand kelp forests.

of units in which the kth species occurs be denoted by u_k. Ugland *et al.* (2003) used the hypergeometric distribution to derive the expected number of species, $E[S_u]$ in a random collection of u units, as follows:

$$E[S_u] = \sum_{k=1}^{S} \left[1 - \frac{\binom{U-u_k}{u}}{\binom{U}{u}} \right] = S - \sum_{k=1}^{S} \frac{\binom{U-u_K}{u}}{\binom{U}{u}} \quad (1)$$

Expanding the binomial coefficients gives:

$$E[S_u]=S-\sum_{k=1}^{S}\left(1-\frac{u}{U}\right)\left(1-\frac{u}{U-1}\right)\dots\left(1-\frac{u}{U-(u_k-1)}\right) \quad (2)$$

We wish to consider the expected number of species as a function of an increasing sequence of random units from a collection to estimate the rate of increase of species number with increasing sample size. This approach involves an extrapolation method for the common situation where the sampling units are only a small fraction of the total area. In most natural assemblages, only a small fraction of the species are widely distributed, while the majority of species occur in a small subset of the available space. From extensive empirical investigations, Ugland *et al.* (2003) suggested that an adequate mathematical approximation to the exact analytical expression in equation (2) could be achieved by fitting a species–accumulation curve according to the semi-log relationship:

$$E[S_u] = \mu + \beta(\log u) \quad (3)$$

(here and throughout, 'log' indicates the natural logarithm).

It is important to realize that the fundamental underlying assumption for this approximation is that species be randomly and independently distributed across the sampled area. Such an assumption might well be violated for communities with a large species number where the spatial distribution is not random and there are biological dependencies between the species. However, the robustness of this approximation to heterogeneous distributions of species within a sampled area has been confirmed by Reichert *et al.* (2010) in analyses of datasets from areas where the total species richness was actually known. Given a set of data, such a relationship can be fitted using ordinary least squares regression of calculated values of $E[S_u]$ for random samples of size u versus log (u). Ugland *et al.* (2003) further developed this idea to incorporate potential variation in the curves that might occur across several subareas within a region. For example, in a given region, sampling units are distributed within a total area A, which includes a number of subareas j = 1,…, a. Then, one species–accumulation curve is obtained for randomized units of all combinations of 1, 2, 3,…, a subareas, with (say) 100 random draws of sampling units for each combination. Let $Y^{(S)} = \{y_j^{(S)}\}$ (j = 1,…, a) be a random variable corresponding to only the terminal points of the species–accumulation curves obtained from different combinations of subareas (see Ugland *et al.*, 2003 for further details). The overall T–S curve is then obtained by fitting a semi-log curve:

$$y_j^{(S)} = \mu_S + \beta_S(\log u_j) + \varepsilon_j \quad (4)$$

where here u_j corresponds to the sample size at each of the endpoints of the j = 1,…, a subarea curves. Ordinary least squares regression (assuming ε_j are i.i.d. normal) gives an estimate for the intercept, $\hat{\mu}_S$, and also for the slope coefficient, $\hat{\beta}_S$, the latter of which is directly interpretable as an estimated rate of increase in the number of species with increasing (log) sample size. Given the relationship in equation (4), an estimate of the total number of species for the region ($\hat{S}_{\text{Tot}}$) is then obtained simply by replacing u with N, the total number of sampling units (given the area comprising each unit and the total area of the region) across the region's full spatial extent, as follows:

$$\hat{S}_{\text{Tot}} = \hat{\mu}_S + \hat{\beta}_S(\log N) \quad (5)$$

In practice, the T–S curve integrates species richness estimates from species–accumulation curves calculated for increasing numbers of subareas to take into account the heterogeneity in species distributions among subareas within the total area sampled.

Note that if from the outset we take k = 1,…, F families, rather than species, through all of the above steps in precisely the same procedure (i.e. equations 1 through 5 and replacing S with F throughout), then we can obtain a 'total families' (T–F) curve. More specifically for what follows, $E[F_u] = \mu_F + \beta_F(\log u)$, and we can obtain an estimate of the intercept $\hat{\mu}_F$ and slope coefficient based on families, $\hat{\beta}_F$, as well as estimating the total number of families in the region, $\hat{F}_{\text{Tot}}$. Indeed, similar such models and associated estimates can naturally be obtained for whatever taxonomic units (genera, orders, etc.) are of interest.

Correlation between T–S curve coefficient and β-diversity

The coefficient of a T–S curve is a measure of how fast the number of species increases with increasing (log) number of sampling units. It is therefore reasonable to propose that this coefficient may be used as a measure of β-diversity. We tested this prediction by correlating, at different levels of taxonomic resolution (i.e. at levels of species, genus, family, order and class), the estimated values of this coefficient with either a classical or distance-based multivariate measure of β-diversity (see Anderson *et al.*, 2011 for review).

For each dataset, matrices of molluscan assemblages were aggregated at different taxonomic levels following the Linnaean classification. For each aggregated matrix in each region, the T–S curve coefficient (or T–F coefficient, etc.)

was estimated following the procedure proposed by Ugland *et al.* (2003), as described above. Each aggregated matrix was used to obtain a taxon–accumulation curve for random samples of all possible combinations of subareas within the total area sampled. Subareas within regions were identified based on general geographical and environmental distinctiveness.

Matrices aggregated at different levels of taxonomic resolution were also used to calculate β-diversity. We used two measures for this: (1) the classical measure of Whittaker (1960), $\beta_W = \gamma/\bar{\alpha}$, where γ is the total richness (number of taxa) within a region ($\gamma = S$ in the case of species-level data) and $\bar{\alpha}$ is the average richness (number of taxa) per sampling unit and (2) the multivariate dispersion among sampling units, as $\bar{d}_{\mathrm{cen}}$ (see Anderson *et al.*, 2006 for further details). More specifically, $\bar{d}_{\mathrm{cen}}$ is the average distance to the group centroid of sampling units in the multivariate space defined by a given resemblance measure. We calculated $\bar{d}_{\mathrm{cen}}$ using the Jaccard dissimilarity measure in order to restrict our attention here only to the differences in the identities of species (or taxa) among sampling units. Note that centroids in the Jaccard space are not equivalent to the arithmetic averages in the space of the original variables. Thus, distances to centroids in Jaccard space must be calculated using the full set of principal coordinate axes (PCO axes) derived from the Jaccard distance matrix for the full set of sampling units, as described in detail in Anderson *et al.* (2006).

For each region, we examined the relationship between the T–S curve coefficients at decreasing taxonomic resolution and the corresponding values of β-diversity. Correlations were calculated for these relationships using Spearman's correlation (ρ) and analyses were performed separately for each of the two types of β-diversity measures.

Estimating regional species richness from regional family richness

The number of species within a given region S can be expressed as the product between the number of families F in the region and the ratio of the number of species to the number of families r:

$$S = F \times r \qquad (6)$$

Note that this does not assume any particular relationship, linear or otherwise, between S and F across subareas within the region. It assumes only that there is one ratio value that may be used for one given single region of interest. The terms in equation (6) can be substituted by their expected values for a given number of units to give:

$$E[S_u] = E[F_u] \times E[r_u] \qquad (7)$$

where $E[S_u]$ and $E[F_u]$ are as previously defined, and $E[r_u]$ is the expected value of the ratio of the number of species to the number of families in u units.

Following (3), equation (7) can be rearranged to give an expression for $E[r_u]$ as:

$$E[r_u] = \frac{\mu_s + \beta_S(\log u)}{\mu_F + \beta_F(\log u)} \qquad (8)$$

As the aim is to estimate species richness in very large areas, the extrapolation of regional species richness using T–S curves is carried out by considering a very large number of sampling units, covering the whole area of interest. We therefore consider equation (8) in the limit as $u \to \infty$, thus:

$$E[r_u] = \lim_{u\to\infty}\left[\frac{\mu_S + \beta_S(\log u)}{\mu_F + \beta_F(\log u)}\right] \qquad (9)$$

Dividing both numerator and denominator by the common value of log (u) gives:

$$E[r_u] = \frac{\lim_{u\to\infty}\left[\frac{\mu_S}{\log u} + \beta_S\right]}{\lim_{u\to\infty}\left[\frac{\mu_F}{\log u} + \beta_F\right]} = \frac{\beta_S}{\beta_F} \qquad (10)$$

Consequently,

$$E[S_u] = E[F_u] \times \frac{\beta_S}{\beta_F} \qquad (11)$$

Thus, for very large values of u, the total number of species in a region can be approximated by:

$$\hat{S}^{(F)}_{\mathrm{Tot}} = \hat{F}_{\mathrm{Tot}} \times \frac{\hat{\beta}_S}{\hat{\beta}_F} \qquad (12)$$

To reduce efforts of taxonomic identifications when estimating regional species richness on the basis of a given number of sampling units, we propose to estimate $\hat{F}_{\mathrm{Tot}}$ using a T–F curve based on all available units U at family level, whereas the ratio $\hat{\beta}_S/\hat{\beta}_F$ will be estimated using only a random subset of such units, analysed at species level. In practice, this will require identification of specimens at species level only for a limited number of sampling units, yielding considerable savings in time and costs, while at the same time providing a calibration of the use of taxonomic surrogates.

We used our datasets to test the effectiveness of the approach and to determine to what extent $\hat{S}^{(F)}_{\mathrm{Tot}}$ (equation 12) may be representative of actual total species richness for the region S_{Tot}, by comparing it with the estimate obtained using all sampling units at species level, namely S_{Tot} (equation 5). For each dataset, we calculated $\hat{F}_{\mathrm{Tot}}$ from the T–F curve based on all available units, and the ratio $\hat{\beta}_S/\hat{\beta}_F$ using random subsets of 25, 35 and 50% of the total number of available units. For each subset and each dataset, 1000 random such draws were done, yielding an empirical distribution of values of $\hat{S}^{(F)}_{\mathrm{Tot}}$, which allowed means and 95% confidence intervals (using the 0.025 and 0.975 quantiles of the empirical distribution) to be obtained. For each region, the corresponding value of $\hat{S}_{\mathrm{Tot}}$ was directly compared with these distributions of $\hat{S}^{(F)}_{\mathrm{Tot}}$ values.

All analyses were performed using R (R Development Core Team, 2010). The R code for calculations is also provided (see Appendix S1 in Supporting Information) along with example data (see Appendices S2 and S3 in Supporting Information).

RESULTS

Independently of the measure employed, β-diversity and T–S curve coefficients decreased with decreasing taxonomic resolution in all regions (Table S2 in Supporting Information). The decrease in the T–S curve coefficient was highly correlated with decreasing β-diversity (whether measured as β_W or $\bar{d}_{\text{cen}}$) in all cases ($\rho = 1$, $P < 0.05$), indicating a strong significant relationship between β-diversity and T–S curve coefficients (Fig. 2). T–S curves for each region based on all sampling units at species level are given in Fig. 3. The corresponding T–F curves based on all sampling units at family level also followed a logarithmic model (Fig. 4).

The estimated number of species for New Zealand kelp forests and Ionian rocky cliffs was lower than for the remaining regions (i.e. Norwegian Continental Shelf, southern Irish Sea, Ionian soft bottoms and southern China Sea; Table 1), which, however, had larger spatial extents (Table S1). Among these larger regions, the southern Irish Sea showed the highest and the southern China Sea showed the lowest estimated species richness, respectively (Table 1). The estimated species richness of Ionian soft bottoms was not much lower than that estimated for the other large regions (it was even higher in some case, i.e. versus the southern China Sea), although it was three to four orders of magnitude smaller in areal extent (Table S1). These relative patterns among regions were consistent for estimated family-level richness (Table 1).

$\hat{S}_{\text{Tot}}$ (i.e. species richness estimated from T–S curves based on all sampling units at species level) fell within the 95% empirical CI for the distribution of $\hat{S}^{(F)}_{\text{Tot}}$ (i.e. species richness obtained from the estimated regional family richness and species/family ratios) using a subset of 25% of the sample for datasets from Ionian rocky cliffs, southern China Sea soft bottoms and New Zealand kelp forests (Fig. 5a). Using a subset of 35% of the sample, the estimated species richness for the Norwegian Continental Shelf and Ionian soft bottoms also fell within the 95% CI (Fig. 5b), whereas $\hat{S}_{\text{Tot}}$ for the southern Irish Sea only fell within the 95% CI of $\hat{S}^{(F)}_{\text{Tot}}$ when a subset of 50% of the sampling units was analysed to species level and used to calculate the species-to-family ratio (Fig. 5c).

The calculation of $\hat{S}^{(F)}_{\text{Tot}}$ led to estimates that were, on average, higher than 85% of $\hat{S}_{\text{Tot}}$ (i.e. the number of species estimated when using all units) for all regions; these estimates were obtained using a subset of 25% of the sampling units at species level for Ionian rocky cliffs and New Zealand kelp forests, a subset of 35% of the sampling units for the Norwegian Continental Shelf and Ionian soft bottoms, and a subset of 50% of the sampling units for the southern Irish Sea and the southern China Sea (Table 2). In all cases, random selection of sample subsets to be identified at species level led to lower and upper limits for the 95% CI of $\hat{S}^{(F)}_{\text{Tot}}$ that were within 70 and 114% of $\hat{S}_{\text{Tot}}$, respectively (Table 2).

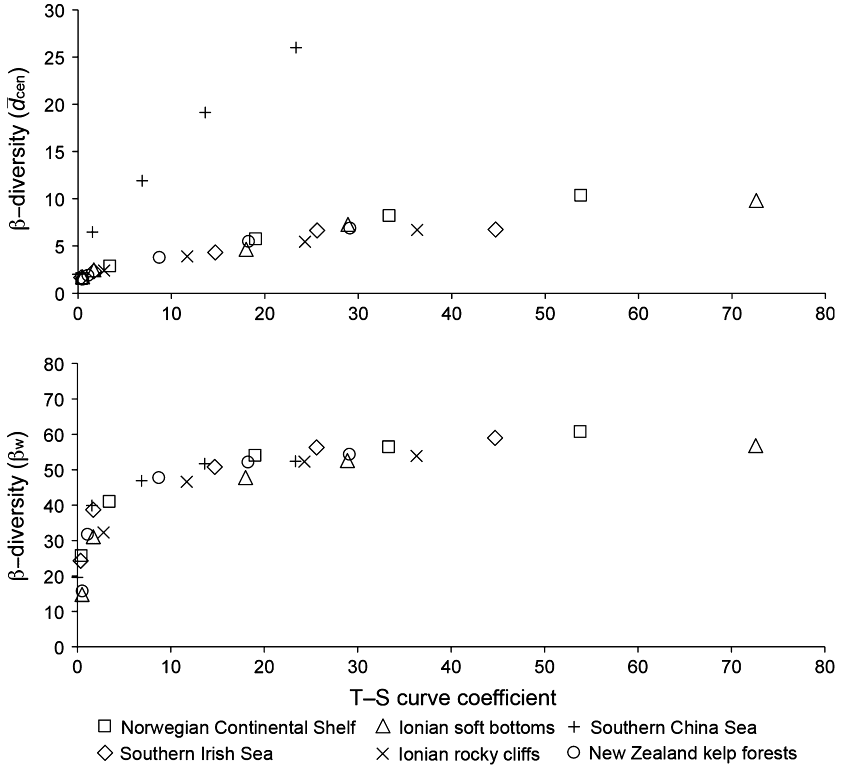

Figure 2 Relationships between β-diversity and T–S curve coefficients at increasing (i.e., from left to right: Class, Order, Family, Genus, Species) levels of taxonomic resolution for each region. Spearman's correlation $\rho = 1$ ($P < 0.05$) in all cases. β-diversity is expressed as β_W (Whittaker's β-diversity) and $\bar{d}_{\text{cen}}$ (multivariate dispersion as the average Jaccard distance to centroid, see Methods for details).

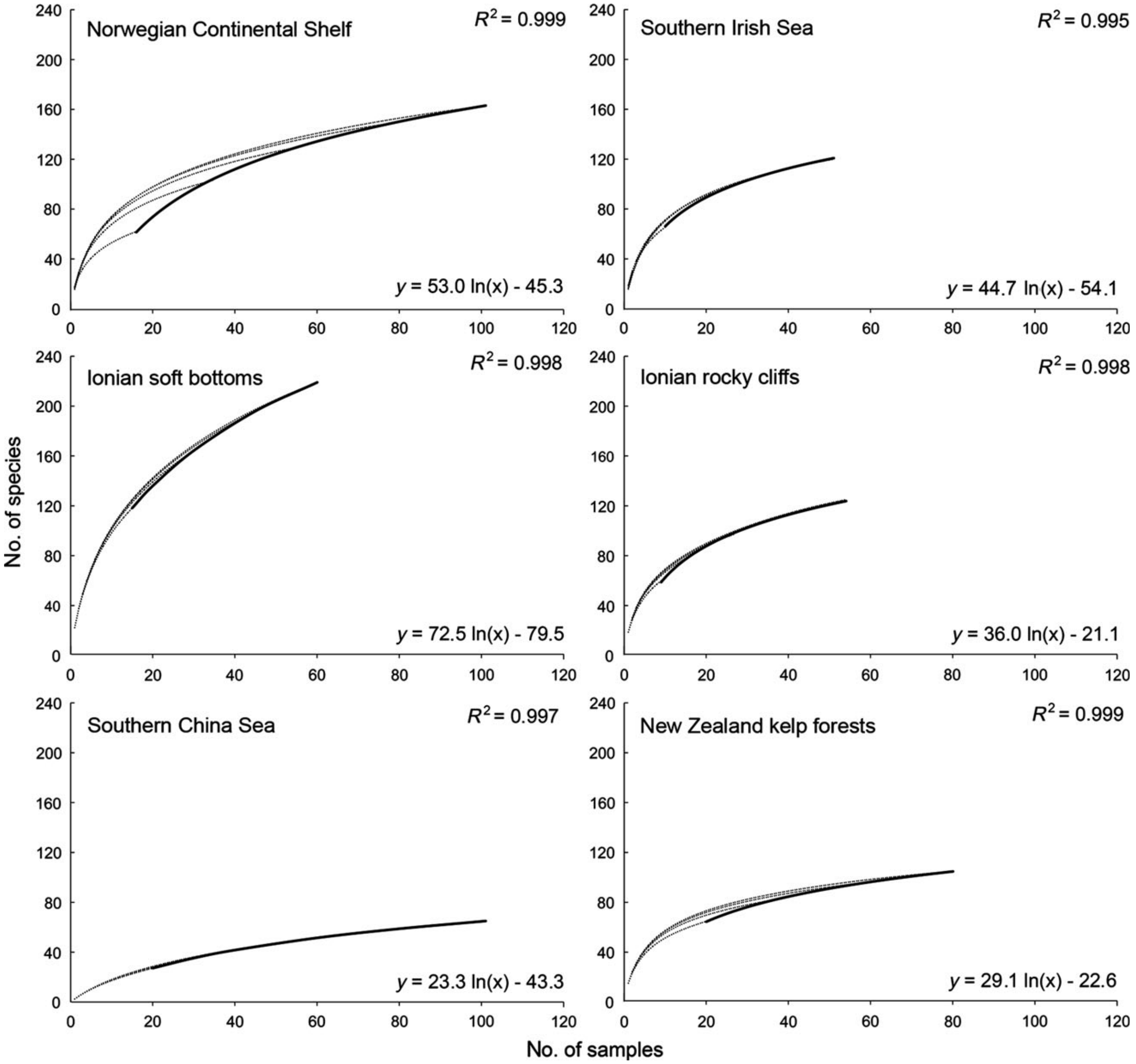

Figure 3 Total–species (T–S) curves (solid lines) based on all sampling units at species level for each region. Dotted lines indicate species–accumulation curves for increasing numbers of subareas within the region (see text for further details). The equation and R^2 statistic for the T–S curve are also reported for each region.

DISCUSSION

Mitigating human impacts on ecosystems and adequately identifying conservation priorities requires a deeper advance in our knowledge of biodiversity. This is not an easy task, involving huge endeavours to describe species and their distributional patterns (Lomolino, 2004; Whittaker *et al.*, 2005), and designing studies to understand underlying causal processes. These efforts require the development of reliable methodological tools to quantify components of biodiversity and to assess the effects of anthropogenic stressors on them (Underwood *et al.*, 2000). Our study combines taxonomic surrogacy with species–accumulation curves to achieve efficient estimation of species richness over large areas. Applying this method will help biodiversity assessment and monitoring programmes to cope with the current gaps in biodiversity knowledge.

The idea that the slope of accumulation curves might represent a measure of β-diversity dates back to the original efforts to model spatial patterns of species distributions, although theoretical and empirical research to understand such relationships has been rarely attempted (Scheiner, 2004). We have shown here that the relationship between the coefficient of the T–S curve and the two diversity measures (i.e. β_W and $\bar{d}_{\text{cen}}$) demonstrates that the coefficient of the T–S curve contains the same essential information about

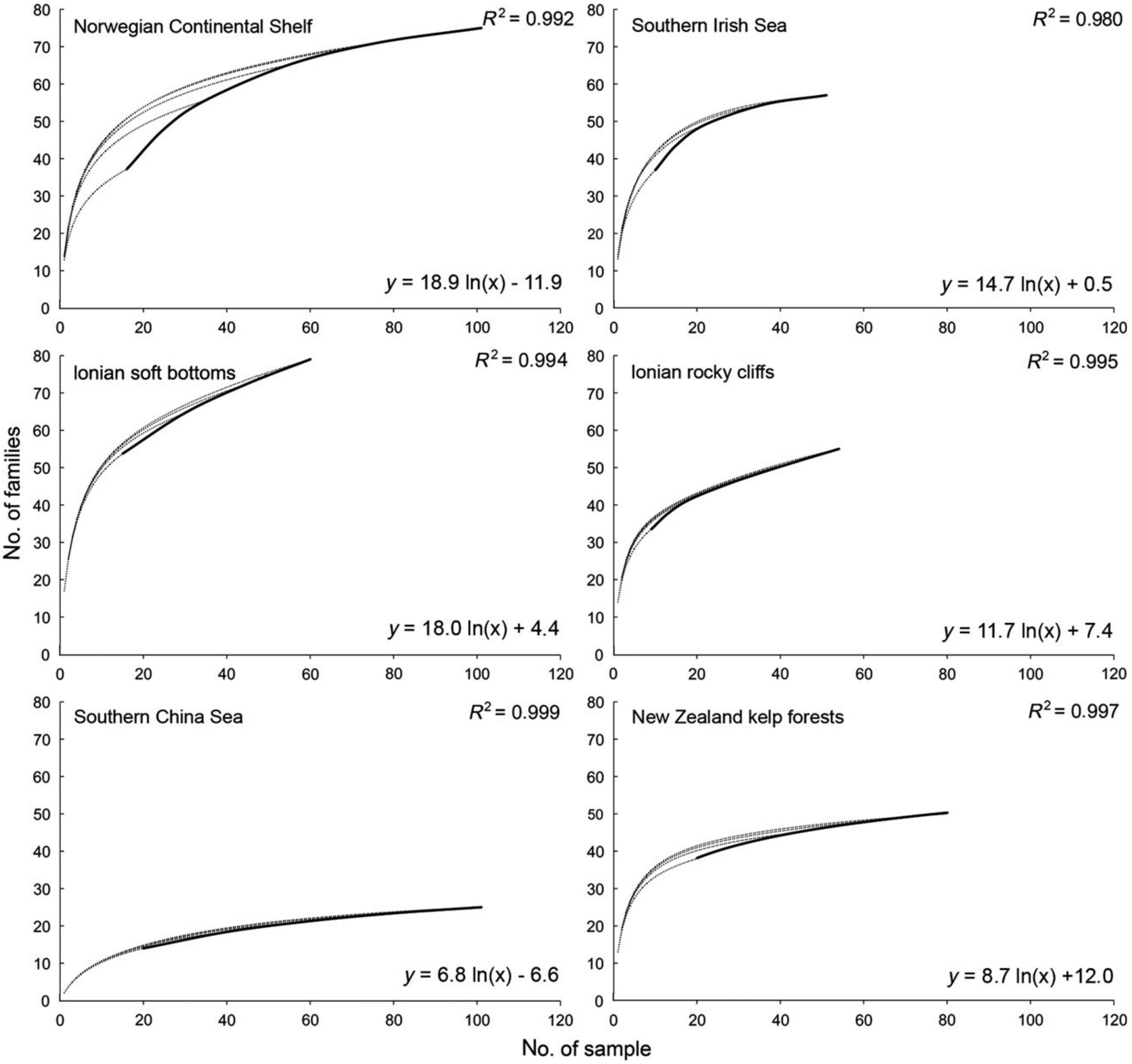

Figure 4 Total–family (T–F) curves (solid lines) based on all sampling units at family level for each region. Dotted lines indicate family accumulation curves for increasing numbers of subareas within the region (see text for further details). The equation and R^2 statistic for the T–F curve are also reported for each region. Note that the scale of the *y*-axis differs from that reported in Fig. 3.

Table 1 Estimates of regional species ($\hat{S}_{\mathrm{Tot}}$) and family ($\hat{F}_{\mathrm{Tot}}$) richness from the total–species (T–S) and total–family (T–F) accumulation curves based on all sampling units in each region. The associated estimated slope ($\hat{\beta}$) and intercept ($\hat{\mu}$) of the T–S (for species) or T–F (for families) curves are also provided

	Species				Family			
Region	$\hat{\beta}_S$	$\hat{\mu}_S$	R^2	$\hat{S}_{\mathrm{Tot}}$	$\hat{\beta}_F$	$\hat{\mu}_F$	R^2	$\hat{F}_{\mathrm{Tot}}$
Southern Irish Sea	44.7	−54.1	0.995	1094	14.7	0.5	0.980	378
Norwegian Continental Shelf	53.0	−45.3	0.999	1002	18.9	−11.9	0.992	376
Ionian Sea (soft bottoms)	72.5	−79.5	0.998	837	18.0	4.4	0.994	232
Southern China Sea	23.3	−43.3	0.997	457	6.8	−6.6	0.999	140
Ionian Sea (rocky cliffs)	36.3	−21.1	0.998	386	11.7	7.4	0.995	139
New Zealand kelp forests	29.1	−22.6	0.999	306	8.7	12.0	0.997	110

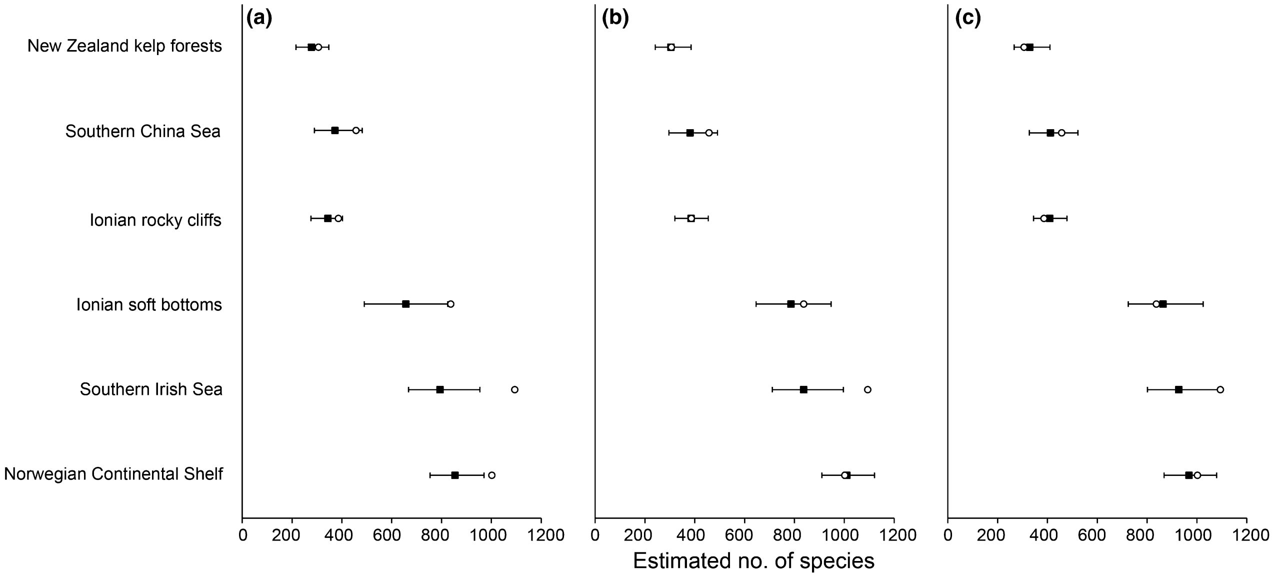

Figure 5 Average estimated number of species in each region ($\hat{S}^{(F)}_{\mathrm{Tot}} \pm 95\%\mathrm{CI}$) on the basis of (a) 25%, (b) 35% or (c) 50% of the sample identified at the species level being used to estimate the species-to-family ratio $\hat{\beta}_S/\hat{\beta}_F$ (see text for details). White circles indicate the number of species estimated from T–S curves based on all units being identified at the species level ($\hat{S}_{\mathrm{Tot}}$).

Table 2 Estimated regional species richness $\hat{S}^{(F)}_{\mathrm{Tot}}$ obtained from the estimated family richness ($\hat{F}_{\mathrm{Tot}}$) from the T–F curve based on all sampling units and the estimated species-to-family ratio ($\hat{\beta}_S/\hat{\beta}_F$, see text for details), as obtained using only a subset of the sampling units identified down to the species level. The estimated values and 95% CI for $\hat{S}^{(F)}_{\mathrm{Tot}}$ are also shown expressed as a percentage of the regional species richness estimated from the T–S curve based on all sampling units being identified at species level ($\hat{S}_{\mathrm{Tot}}$). For each region, only results for the smallest subset required to obtain representative estimates (i.e. ≥ 85% of $\hat{S}_{\mathrm{Tot}}$) are reported here

Region	Southern Irish Sea	Norwegian Continental Shelf	Ionian Sea (soft bottoms)	Southern China Sea	Ionian Sea (rocky cliffs)	New Zealand kelp forests
Estimate of regional family richness ($\hat{F}_{\mathrm{Tot}}$) from the T–F curve based on all sampling units	378	376	232	140	139	111
Estimate of the species-to-family ratio r (= $\hat{\beta}_S/\hat{\beta}_F$)	2.5	2.4	3.4	2.9	2.5	2.5
Percentage of sampling units at species level used to estimate r (%)	50	35	35	50	25	25
Estimate of regional species richness as $\hat{S}^{(F)}_{\mathrm{Tot}} = \hat{F}_{\mathrm{Tot}} \times \hat{\beta}_S/\hat{\beta}_F$	945	902	789	406	348	278
Estimate of regional species richness ($\hat{S}_{\mathrm{Tot}}$) from the T–S curve based on all sampling units	1094	1002	837	457	386	306
$\hat{S}^{(F)}_{\mathrm{Tot}}$ as a percentage of $\hat{S}_{\mathrm{Tot}}$ (%)	86	90	94	89	90	91
95% CI of $\hat{S}^{(F)}_{\mathrm{Tot}}$ (%)	73–100	79–101	78–114	74–113	76–108	70–112

β-diversity as the two other measures across taxonomic levels for a given dataset (Fig. 2).

For four of the datasets considered here, a subsample of about 1/3 of the total number of sampling units at species level was sufficient for the resulting estimate obtained using the species-to-family ratio to be within 20% of the estimate obtained from T–S curves based on all units. For the remaining two regions, it was necessary to increase the fraction of subsamples up to 50% to achieve the same precision. At a subsampling intensity of 35–50%, the confidence interval incorporated the estimate obtained from T–S curves based on all units (Fig. 5). Thus, our simulation indicates, at least for molluscs, that it is possible to reduce the efforts in terms of sampling units analysed at species level by 50% and still obtain reliable species richness estimates.

The implications of this outcome are many. Following a pilot investigation of biodiversity characterizing a given region at species level, the calibration of identification efforts based on the proposed framework could be conducive to cost-effective subsequent investigations. This is of crucial importance to large-scale monitoring of biodiversity, which is often deemed impractical. For instance, following a first extensive sampling exercise, with identifications being carried out at species level, the sample subset required to obtain reliable estimates of regional species richness could be identified. Then, subsequent samplings could be carried out analysing all units at family level and only a limited number of them at species level. The ensuing estimate of regional species richness, if falling below or above the 95% CI of the pilot estimate, could be directly interpreted then as a significant decrease or increase in regional species richness.

Estimates of species richness over large areas are inherently dependent on sample size, irrespective of the estimation method used. In most cases, a number of factors such as the spatial extent of the investigated area, and the time and costs required for analysing samples may constrain sampling intensity. Thus, estimates of species richness often rely on a subsample of the whole area of interest. In this respect, our approach is no exception. Estimates of regional family richness are likely to be affected by sample size as well. However, as previously outlined, an initial pilot assessment of regional species richness and an appropriate choice of sample size to be adopted should be achieved prior to implementation of the approach outlined here. Although sample size is often dictated by experimental and funding constraints and the statistical procedures to be used in order to define a sufficient sample size for reliable estimates of regional species richness are debated, considerable work on this topic has been done and is described elsewhere (e.g. Chao *et al.*, 2009). It could be argued that our approach might introduce an additional bias related to sample size as species richness estimates rely on numerical relationships between taxonomic ranks (i.e. r, the species-to-family ratio), which may vary depending on the number of samples (Gotelli & Colwell, 2001). However, our species-to-family ratio, despite being calculated from a limited number of samples, is not assumed to be invariant across the whole investigated area. In fact, this would not be realistic and would not be conducive to obtaining reliable estimates of regional species richness (see Table S3 in Supporting Information). Instead, this ratio is modelled in our approach against the number of samples, and its value is estimated as if the whole area of interest was sampled.

The method proposed in the present study introduces several improvements to traditional methods for estimation of regional species richness by taking into account the spatial heterogeneity of species distributions within the sampled area and relying on estimates of species-to-family ratios derived precisely from the specific area of interest alone. The approach does assume that suitable subareas of geographical and/or environmental distinctiveness within the region of interest have been identified *a priori*, but the increasing availability of broad-scale environmental data makes this task quite feasible for the majority of investigations. Note that our procedure does not necessitate complete baseline taxonomic information on reference areas, requiring only a fully representative pilot investigation at species level within the region of interest, thus providing a versatile tool for rapid assessment of species richness over large areas. Clearly, further efforts to apply this method across a larger array of organisms and geographical regions are needed before any further generalizations can be made regarding its utility. Nevertheless, our approach is likely to provide important insights and to yield helpful sampling efficiencies in the early investigation of biodiversity across large unexplored areas or environments.

Estimation of species richness by extrapolation of the T–S curve takes into account the effects of habitat heterogeneity (Ugland *et al.*, 2005). As families tend to incorporate several species in a given habitat, the accumulation curves of families will be less influenced by heterogeneity (Gotelli & Colwell, 2001). Our suggested procedure is therefore less influenced by heterogeneity than extrapolation based on presence–absence of species. We also expect that the total–family (T–F) curves stabilize more rapidly than T–S curves as the sample size increases, as suggested by the decreasing values of slope coefficients with decreasing taxonomic resolution (Table 2), although this, in itself, is certainly a topic worthy of future investigation.

The approach described here does rely, however, upon the validity of the semi-log approximation described by Ugland *et al.* (2003) (equation (3) above). Although this method was fairly recently identified by Reichert *et al.* (2010) as the best among a suite of potential estimation methods, construction of separate accumulation curves for rare, intermediate and common species revealed that the shape of the randomized empirical accumulation curve for all species is determined primarily by the occurrence of rare species. Furthermore, their simulations showed that the T–S approach of Ugland *et al.* (2003) will either under- or overestimate total species richness if the occupancy of rare species is unusually low or high, respectively. This suggests, therefore, that caution should be exercised in applying the proposed method in such cases.

The approach we have outlined here is purely instrumental, with the goal of sampling efficiency, and does not attempt to provide a theoretical underpinning for the ecological meaning of taxonomic surrogacy (e.g. Bevilacqua *et al.*, 2012). However, the basic ideas presented here might have potential for application with other types of measures of diversity, such as functional or phylogenetic diversity (e.g. Graham & Fine, 2008; Swenson *et al.*, 2010). Functional diversity may not be hierarchical, however (Laliberté & Legendre, 2010), and the lengths of branches in different parts of a hierarchical phylogeny may vary, meaning that comparable entities at higher levels could be difficult to delineate. The potential to combine analyses of functional, phylogenetic and taxonomic diversity to provide insights into broad-scale patterns of community assembly is, however, a fruitful area of current ecological research (e.g. Stegen &

Hurlbert, 2011; Anacker & Harrison, 2012), and cost-effective methods, such as that proposed here, for quantifying biodiversity, will certainly serve to enhance such endeavours.

Above all, the approach of combining a method that accounts for compositional heterogeneity in natural assemblages with the use of taxonomic surrogates is likely to represent a powerful way to help fill the gaps in our current knowledge of human-driven changes in macro-ecological patterns. The ongoing decline of global biological diversity has raised wide concern about expected deleterious effects on the functioning of ecosystems and the flow-on consequences in terms of our ability to derive ecosystem goods and services from such compromised systems (Balmford & Bond, 2005). Potential negative outcomes of biodiversity loss are particularly alarming for seas and oceans (Halpern *et al.*, 2008), for which wide gaps in biodiversity information on many marine organisms (Archambault *et al.*, 2010; Fraschetti *et al.*, 2011) further increase uncertainties on the magnitude of human-driven species extinctions and functional losses in marine ecosystems (Sala & Knowlton, 2006). We consider that the predictions made by our procedure have a wide range of potential applications, including conservation, large-scale biodiversity surveys and routine monitoring of biodiversity changes in relation to global changes or localized human impacts. Importantly, although the taxonomic efforts required are reduced by our approach, the concept at the base of the procedure does not disregard the importance of the identification of species and thus the role of taxonomy, a crucial discipline that lies at the heart of any knowledge or study of biodiversity.

ACKNOWLEDGEMENTS

This work was carried out in the frame of various projects on marine biodiversity funded by MIUR, Provincia di Brindisi, ENI S.p.A. and CoNISMa. The research leading to these results was instrumental for the preparation of the EU Community's Seventh Framework (FP7/2007–2013) Programs COCONET, Grant Agreement No. 287844) and PERSEUS (Grant Agreement No. 287600).

REFERENCES

Anacker, B.L. & Harrison, S.P. (2012) Historical and ecological controls on phylogenetic diversity in Californian plant communities. *The American Naturalist*, **180**, 257–269.

Andersen, A. (1995) Measuring more of biodiversity: genus richness as a surrogate for species richness in australian ant faunas. *Biological Conservation*, **73**, 39–43.

Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, **9**, 683–693.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C. & Swenson, N.G. (2011) Navigating the multiple meanings of β-diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.

Archambault, P., Snelgrove, P.V.R., Fisher, J.A.D., Gagnonm, J.-M., Garbary, D.J., Harvey, M., Kenchington, E.L., Lesage, V., Levesque, M., Lovejoy, C., Mackas, D.L., McKindsey, C.W., Nelson, J.R., Pepin, P., Piché, L. & Poulin, M. (2010) From sea to sea: Canada's three oceans of biodiversity. *PLoS ONE*, **5**, e12182.

Balmford, A. & Bond, W. (2005) Trends in the state of nature and implications for human well-being. *Ecology Letters*, **8**, 1218–1234.

Balmford, A., Green, M.J.B. & Murray, M.G. (1996) Using higher taxon richness as a surrogate for species richness: I. Regional tests. *Proceedings of the Royal Society of London, Series B*, **263**, 1267–1274.

Balmford, A., Lyon, A.J.E. & Lang, R.M. (2000) Testing the higher taxon approach to conservation planning in a megadiverse group: the macrofungi. *Biological Conservation*, **93**, 209–217.

Beattie, A.J. & Oliver, I. (1994) Taxonomic minimalism. *Trends in Ecology and Evolution*, **9**, 488–490.

Bevilacqua, S., Fraschetti, S., Musco, L. & Terlizzi, A. (2009) Taxonomic sufficiency in the detection of natural and human-induced changes in marine assemblages: a comparison of habitats and taxonomic groups. *Marine Pollution Bulletin*, **58**, 1850–1859.

Bevilacqua, S., Terlizzi, A., Claudet, J., Fraschetti, S. & Boero, F. (2012) Taxonomic relatedness does not matter for species surrogacy in the assessment of community response to environmental drivers. *Journal of Applied Ecology*, **49**, 357–366.

Cardoso, P., Erwin, T., Borges, P.A.V. & New, T.R. (2011) The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, **144**, 2647–2655.

Chao, A., Colwell, R.K., Lin, C.-W. & Gotelli, N.J. (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology*, **90**, 1125–1133.

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **345**, 101–118.

Fraschetti, S., Guarnieri, G., Bevilacqua, S., Terlizzi, A., Claudet, J., Russo, G.F. & Boero, F. (2011) Conservation of Mediterranean seascapes and the biodiversity countdowns: what information do we really need? *Aquatic Conservation: Marine and Freshwater Ecosystems*, **21**, 299–306.

Gaston, K. & Williams, P.H. (1993) Mapping the world's species – the higher taxon approach. *Biodiversity Letters*, **1**, 2–8.

Gotelli, N. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R. & Watson, R. (2008) A global map of human impact on marine ecosystems. *Science*, **319**, 948–952.

He, F. & Hubbell, S.P. (2011) Species–area relationships always overestimate extinction rates from habitat loss. *Nature*, **473**, 368–371.

Hutchinson, G.E. (1959) Homage to Santa Rosalia or why are there so many kinds of animals? *The American Naturalist*, **93**, 145–159.

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

Lewandowski, A.S., Noss, R.F. & Parsons, D.R. (2010) The effectiveness of surrogate taxa for the representation of biodiversity. *Conservation Biology*, **24**, 1367–1377.

Lomolino, M.V. (2004) Conservation biogeography. *Frontiers of biogeography: new directions in the geography of nature* (ed. by M.V. Lomolino and L.R. Heaney), pp. 293–296. Sinauer Associates, Sunderland, MA.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ.

Matias, M.G., Underwood, A.J., Hochuli, D.F. & Coleman, R.A. (2011) Habitat identity influences species–area relationships in heterogeneous habitats. *Marine Ecology Progress Series*, **437**, 135–145.

May, R.M. (1988) How many species are there on the Earth? *Science*, **241**, 1441–1449.

Mazaris, A.D., Kallimanis, A.S., Tzanopoulos, J., Sgardelis, S.P. & Pantis, J.D. (2010) Can we predict the number of plant species from the richness of a few common genera, families or orders? *Journal of Applied Ecology*, **47**, 662–670.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B. & Worm, B. (2011) How many species are there on Earth and in the Ocean? *PLoS Biology*, **9**, e1001127.

Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

R Development Core Team. (2010) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org (accessed 26 December 2013).

Reichert, K., Ugland, K.I., Bartsch, I., Hortal, J., Bremner, J. & Kraberg, A. (2010) Species richness estimation: estimator performance and the influence of rare species. *Limnology and Oceanography –. Methods*, **8**, 294–303.

Sala, E. & Knowlton, N. (2006) Global marine biodiversity trends. *Annual Review of Environment and Resources*, **31**, 93–122.

Scheiner, S.M. (2004) A mélange of curves – further dialogue about species–area relationships. *Global Ecology and Biogeography*, **13**, 479–484.

Smale, D.A. (2010) Monitoring marine macroalgae: the influence of spatial scale on the usefulness of biodiversity surrogates. *Diversity and Distributions*, **16**, 985–995.

Stegen, J.C. & Hurlbert, A.H. (2011) Inferring ecological processes from taxonomic, phylogenetic and functional trait beta diversity. *PLoS ONE*, **6**, e20906.

Sutcliffe, P.R., Pitcher, C.R., Caley, M.J. & Possingham, H.P. (2012) Biological surrogacy in tropical seabed assemblages fails. *Ecological Applications*, **22**, 1762–1771.

Swenson, N.G., Anglada-Cordero, P. & Barone, J.A. (2010) Deterministic tropical tree community turnover: evidence from patterns of functional beta diversity along an elevational gradient. *Proceedings of the Royal Society of London, Series B*, **278**, 877–884.

Terlizzi, A., Anderson, M.J., Bevilacqua, S., Fraschetti, S., Wodarska-Kowalczuk, M. & Ellingsen, K.E. (2009) Beta diversity and taxonomic sufficiency: do higher-level taxa reflect heterogeneity in species composition? *Diversity and Distributions*, **15**, 450–458.

Ugland, K.I., Gray, J.S. & Ellingsen, K.E. (2003) The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888–897.

Ugland, K.I., Gray, J.S. & Lambshead, J. (2005) Species accumulation curves analysed by a class of null models discovered by Arrhenius. *Oikos*, **108**, 263–274.

Underwood, A.J., Chapman, M.G. & Connell, S.D. (2000) Observations in ecology: you can't make progress on process without understanding the patterns. *Journal of Experimental Marine Biology and Ecology*, **250**, 97–115.

Vitousek, P.M., Mooney, H.A., Lubchenco, J. & Melillo, J.M. (1997) Human domination of Earth's ecosystems. *Science*, **277**, 494–499.

Wheeler, Q.D., Raven, P.H. & Wilson, E.O. (2004) Taxonomy: impediment or expedient? *Science*, **303**, 285.

Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 279–338.

Whittaker, R.J., Araújo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity & Distributions*, **11**, 3–23.

Williams, P.H. & Gaston, K.J. (1994) Measuring more of biodiversity: can higher-taxon richness predict wholesale species richness? *Biological Conservation*, **67**, 211–217.

Wilson, E.O. (2004) Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society, Series B*, **359**, 739.

Witman, J.D., Etter, R.J. & Smith, F. (2004) The relationship between regional and local species diversity in marine benthic communities: a global perspective. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15664–15669.

Worm, B. & Duffy, J.E. (2003) Biodiversity, productivity, and stability in real food webs. *Trends in Ecology and Evolution*, **18**, 628–632.

Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, **16**, 446–453.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Dataset information.

Table S2 β-diversity values and T–S curve coefficients at decreasing taxonomic resolution.

Table S3 Estimated regional species richness $\hat{S}^{(F)}_{\mathrm{Tot\,sub}}$ obtained from the estimated family richness ($\hat{F}_{\mathrm{Tot}}$) from the T–F curve based on all sampling units and the mean species-to-family ratio r_{sub}, as obtained using sampling units from a single subarea.

Appendix S1 R code for the calculation of T–S curve, T–F curve, and estimate of species to families ratio (r) ± 95% empirical CI.

Appendix S2 Species-level data of mollusc assemblages from New Zealand kelp forest (NZK) as an example for the application of the R code in Appendix 1.

Appendix S3 Family-level data of mollusc assemblages from New Zealand kelp forest (NZK) as an example for the application of the R code in Appendix 1.

BIOSKETCH

Antonio Terlizzi is Associate Professor of Zoology at the LZMB of the University of Salento, together with **Stanislao Bevilacqua**, a postdoctoral researcher. They study human-driven changes to marine biodiversity and are currently focusing on links among taxonomic, phylogenetic and functional diversity and their potential application for both biodiversity conservation and impact assessment. Common research interests have led, since several years, to the collaboration with **Marti Jane Anderson**, Professor of Ecology at the NZIAS, whose contributions to multivariate statistical analyses are greatly influencing current ecological research. More recently, **Karl Inne Ugland**, a senior scientist at the University of Oslo, joined the group providing further insights for modelling spatial patterns of species distribution.

Author contributions: A.T. conceived the study, M.J.A. designed the R code for analyses; all authors provided the data; A.T., S.B and K.I.U. analysed the data; A.T. and S.B. wrote the paper with considerable improvements and additional statistical guidance provided by M.J.A. and K.I.U.

Editor: Omar Defeo