

REVIEW

Measure and mis-measure of species diversity in deep-sea chemosynthetic communities

Olivier Gauthier*, Jozée Sarrazin, Daniel Desbruyères

Ifremer, Centre de Brest, DEEP/LEP, BP 70, 29280 Plouzané, France

ABSTRACT: Our knowledge of species diversity in deep-sea chemosynthetic communities has come a long way since their discovery in the late 1970s. However, their study poses serious challenges that are linked to their remoteness, the variable selectivity and effectiveness of sampling tools in different conditions, a lack of consensus on the size compartments and sieve size used in practice, and the persistent discovery of species that are previously unknown to science. Still, with increasing accessibility and new sampling tools, quantitative data are becoming both more frequent and extensive, reducing limitations in the estimation and comparison of species diversity. Here, we review the literature for common difficulties encountered in the measurement and comparison of species diversity within and among deep-sea chemosynthetic communities. We lay out recommendations for future and ongoing work that should lead to more accurate evaluations and more robust comparisons. We advocate: (1) complete transparency in the taxonomic levels used to compute reported diversity measures and public access to data tables after publication; (2) an effort to standardize size compartments and sieves used within the research community; (3) the use of families of species diversity measures that allow for differential weighting of abundant and rare species; (4) the evaluation of the completeness of inventories and the stability of all diversity measures for all groups of samples under comparison; and (5) caution in the use and interpretation of statistical tests on species diversity. Only through appropriate tools can we hope to gain the knowledge and understanding necessary for the management and protection of the often endemic fauna found in deep-sea chemosynthetic communities.

KEY WORDS: Deep-sea ecosystems · Chemosynthetic community · Species diversity · Richness · Evenness · Rarefaction · Diversity profiles · Diversity ordering · Alpha diversity · Statistical analysis

Resale or republication not permitted without written consent of the publisher

'Diversity measures are likely to be rather dependent on a complex combination of the sieve mesh size used, the area of the sample and the size distribution of the organisms present. This makes comparison of diversity measures from different studies difficult to interpret.'

(Warwick & Clarke 1996)

'Indices like Simpson's and Shannon's are products of the mechanical calculator age when characterising fauna samples with more than a single value was a lengthy process. [...] Scientifically, multiple measures should be used beginning with species richness, rarefacted values to allow for better comparisons, and then indices giving different weights to different fractions.'

(Carney 2007)

INTRODUCTION

Describing, explaining and modelling patterns of biological diversity is one of the major challenges facing ecologists today. Despite substantial advances in the sampling, processing, and analysis of data, researchers are still confronted with problems when engaging in this task. This is particularly true for deep-sea chemosynthetic habitats that are characterized by difficult access and scarcity of data, and for which unknown species are commonly found in surveys, especially in the smaller size-components (e.g. meiofauna, Vanreusel et al. 1997, Bright et al. 2003, Sarrazin et al.

2006, Van Gaeve et al. 2006, Copley et al. 2007, Limen et al. 2007). Even the seemingly 'simple' task of quantifying species diversity within or among groups of sites can turn out to be a difficult endeavour (Purvis & Hector 2000). Biological diversity can have many different meanings. It can refer to ecological, molecular, phylogenetic or even functional (processes) diversity, depending on the level of observation. Here, we are concerned with the 'traditional' ecological meaning of diversity, which is commonly called 'species diversity' or ' α -diversity'. When considering samples that are drawn from a community, it usually refers to both 'species richness' (the number of species) and 'species evenness' (the relative abundance of taxa) (Simpson 1949). However, it is also possible to refer to guild diversity, trophic diversity, and genus diversity (or any higher taxonomic groups); any subdivision of the community to which some biological meaning can be assigned could be worthy of such quantification. The 2 basic components of species diversity (richness and evenness) are usually summarized as 'simple' numbers in the form of 'diversity indices' or 'diversity measures'. They include, but are not limited to, species richness (S ; McIntosh 1967), Shannon's entropy (H ; Shannon 1948), or some form of Simpson's concentration (λ ; Simpson 1949).

Publication of quantitative ecological data, i.e. counts of individuals by taxon, from hydrothermal vents and other deep-sea chemosynthetic communities have only become common in the last decade (e.g. Sibuet & Olu 1998, Sarrazin & Juniper 1999, Van Dover & Trask 2000, Van Dover 2002, Baco & Smith 2003, Tsurumi 2003, Tsurumi & Tunnicliffe 2003, Turnipseed et al. 2003, 2004, Bergquist et al. 2003, 2005, Cordes et al. 2005, Flint et al. 2006, Zekely et al. 2006a,b, Copley et al. 2007). The advent of quantitative faunal data from deep-sea chemosynthetic communities calls for better evaluations of the comparability of samples and the use of robust diversity ordering methods. A lack of consensus in the methodology of collecting, describing and analysing data often makes it difficult to confidently compare results both within and between independent studies (Sutherland 1996, Warwick & Clarke 1996). Observed species frequencies depend on the underlying diversity, but also vary with the capture probability distribution. The latter is a function of the sampling tools and strategy, the structure of the habitats, and the density of organisms (Gotelli & Colwell 2001, Colwell et al. 2004). It is thus necessary to ensure that adequate and comparable sampling is realized in all habitats and communities. Indeed, in order to be reliable, species diversity comparisons have to consider sampling method, mesh size, the number of samples, the number of individuals, the survey area, and the area or volume covered. Since such discrepancies are often ignored or

not deemed sufficient to warrant concern, we consider it necessary to briefly review some of the methods, tools and requirements of species diversity research in their application to research in deep-sea chemosynthetic communities. We report on some problematic aspects of sampling and numerical analyses that are sometimes neglected in species diversity analysis in these communities. They include inconsistencies in sampling protocol, identification, or data analyses for which there are often, but not always, simple solutions. Some of these issues have been documented elsewhere for other benthic habitats (e.g. Sutherland 1996, Warwick & Clarke 1996).

It is understandable that the paucity of previously published work and the enormous difficulties and costs attached to the acquisition of data in extreme deep-sea environments have played a major role in this situation. However, it is now essential to adopt a more pragmatic and critical strategy to diversity estimation and comparison in deep-sea chemosynthetic ecosystems. The aim of this review is to emphasize the necessity of considering all factors that influence assessments of species diversity, and to properly consider what is being compared. First, a review of some of the submersible sampling tools that are used in deep-sea habitats is presented. Their selectivity with regards to hard and soft substratum is emphasized, but organism size and mobility are also considered. We then revisit commonly used measures and indices of diversity, and discuss some alternatives. Problems encountered in some diversity characterizations are illustrated with examples from the literature. One crucial aspect of the problem is the comparability of biological diversity between groups of samples that come from different habitats, communities, or different research projects. In particular, the ranking of sample groups and the use of some statistical tests will be addressed. We argue that it is necessary to scrutinize differences in sampling design and evaluate the completeness of a survey to estimate species richness and evenness. Failure to do this can lead to spurious results. Finally, some recommendations are made in light of the difficulties posed by the study of these remote communities. We hope that this work will trigger constructive discussion, and ultimately lead to higher quality descriptions and comparisons of the species diversity in these peculiar habitats, both within and between studies.

Some obligatory symbolism

Species richness (S) is the number of species in a community. The true species richness (S_{true}) of a community is the actual number of species composing it. The observed species richness (S_{obs}) is the number of

species found in a sample; it is one of many estimates (S_{est}) of S_{true} . For a sample of N individuals, each species i has an observed abundance of n_i individuals, and the sum of the observed abundances for all species equals the sample size: $\sum_{i=1}^{S_{\text{obs}}} n_i = N$. The relative abundance of each species in the sample (p_i) is simply its abundance divided by the number of individuals in the sample: $p_i = n_i/N$. It is an estimate of the species' true relative abundance in the community, and the p_i of all the species in the community add up to 1: $\sum_{i=1}^{S_{\text{obs}}} p_i = 1$. A community is said to be perfectly 'even' if all species in the community have equal relative abundances, i.e. if $p_i = N/S$ for all $i = 1, 2, \dots, S$. Communities where species frequency distributions depart from this theoretical situation towards one where a minority of species accounts for a majority of the cumulative relative abundance are said to be increasingly 'uneven'. This is often the case in chemosynthetic communities (e.g. Bergquist et al. 2003, 2005, Smith & Baco 2003, Zekely et al. 2006a, Copley et al. 2007, Gollner et al. 2007, Galkin & Goroslavskaya 2008).

DEEP-SEA COMMON SUBMERSIBLE SAMPLING TOOLS & TREATMENT

Push corers (or tube corers; Fig. 1A), blade corers (Bayon et al. 2009; Fig. 1B,C), and Birge-Ekman box corers (Rowe & Clifford 1973; Fig. 1D) are the typical submersible tools that are used to sample fauna in the soft sediments of the deep ocean. These tools are very efficient in sampling a known surface; however, no generic tools are available or are used by scientists from around the world (Table 1). Corers do not necessarily have the same size, and this may pose some comparison concerns. Nevertheless, they are considered to be quantitative, i.e. their results can be easily transformed to account for the area or volume sampled.

Originally designed by Van Dover (2002) and recently upgraded by C. Fisher's team (Pennsylvania State University), the pot sampler (Fig. 1E) is dedicated to the sampling of mussel beds and their associated community on both hard and soft substrata. It consists of a modified 12 q (11.356 l) aluminium-clad stainless pot with a handle, is filled with a Kevlar bag and is deployed by a submersible mechanical arm. After enclosing the assemblage, a handle is turned to cinch up the Kevlar bag, firmly securing the sampled mussels and associated fauna. The sampled area is known (531 cm²), but the sampled volume is variable and depends on the penetration of the pot in the sediments.

Most scientific submersibles are equipped with a suction sampler (Fig. 1F). This sampling tool, which is placed on the vehicle, is composed of a centrifugal pump and a long supple pipe with a nozzle. The suc-

tion sampler is manipulated with the submersible arm and is used on a target surface. Faunal samples go through the pipe and are collected in independent jars. A mesh of varying size is usually placed inside each jar to prevent sample loss. The suction sampler is commonly used to collect smaller organisms, otherwise the pipe could get clogged and the sample lost. It is particularly efficient in collecting smaller animals on hard substratum, but can also be used on soft sediments. Although the area of the sampled surface can be evaluated through video imagery (Sarrazin et al. 1997), this tool is generally not considered to be quantitative. However, as with grab and scoop samples (see below), it can be used to quantify species diversity alongside more quantitative tools, provided that their respective selectivity is acknowledged and data is properly analysed.

The major problem encountered in the deep sea is the sampling of large faunal communities on hard substratum. The Chimney-master and Bush-master (Fig. 1G) are original instruments created by C. Fisher's team to collect vent communities on and around hydrothermal edifices (Hourdez et al. 2000, Bergquist et al. 2002). These large nets are used to collect complete bushes of tubeworms and associated fauna, either on the seafloor (Bush-master) or on sulphide chimneys (Chimney-master). They have also been used on corals. The Bush-master has proven to be very efficient on relatively flat hard bottoms (Bergquist et al. 2002, Urcuyo et al. 2003, Bergquist et al. 2007). On irregular or vertical sulphide surfaces, it is also practical to use the submersible suction sampler in combination with the manipulator arm to collect faunal samples (Sarrazin & Juniper 1999). These so-called 'grab samples' (Fig. 1H) are then placed in collection boxes (Fig. 1I). In these instances, video imagery is used to get an estimation of the sampled area (Sarrazin et al. 1997, Tsurumi 2003, Tsurumi & Tunnicliffe 2003). Many submersibles are also equipped with some kind of nonquantitative scoop (Fig. 1J), which is a canvas rock bag or mesh net with a solid wire frame at the mouth and a T-handle. Finally, the manipulator arm is also regularly used to collect whole organisms (Fig. 1K) and abiotic substrata (Fig. 1L).

Sieving

Each of the tools described above has different capture efficiencies for different species (Table 1) and may lead to loss of individuals, underestimation of the relative abundance of some groups, and overestimation of others. These tools thus differ in selectivity. It may be difficult to distinguish the differences stemming from

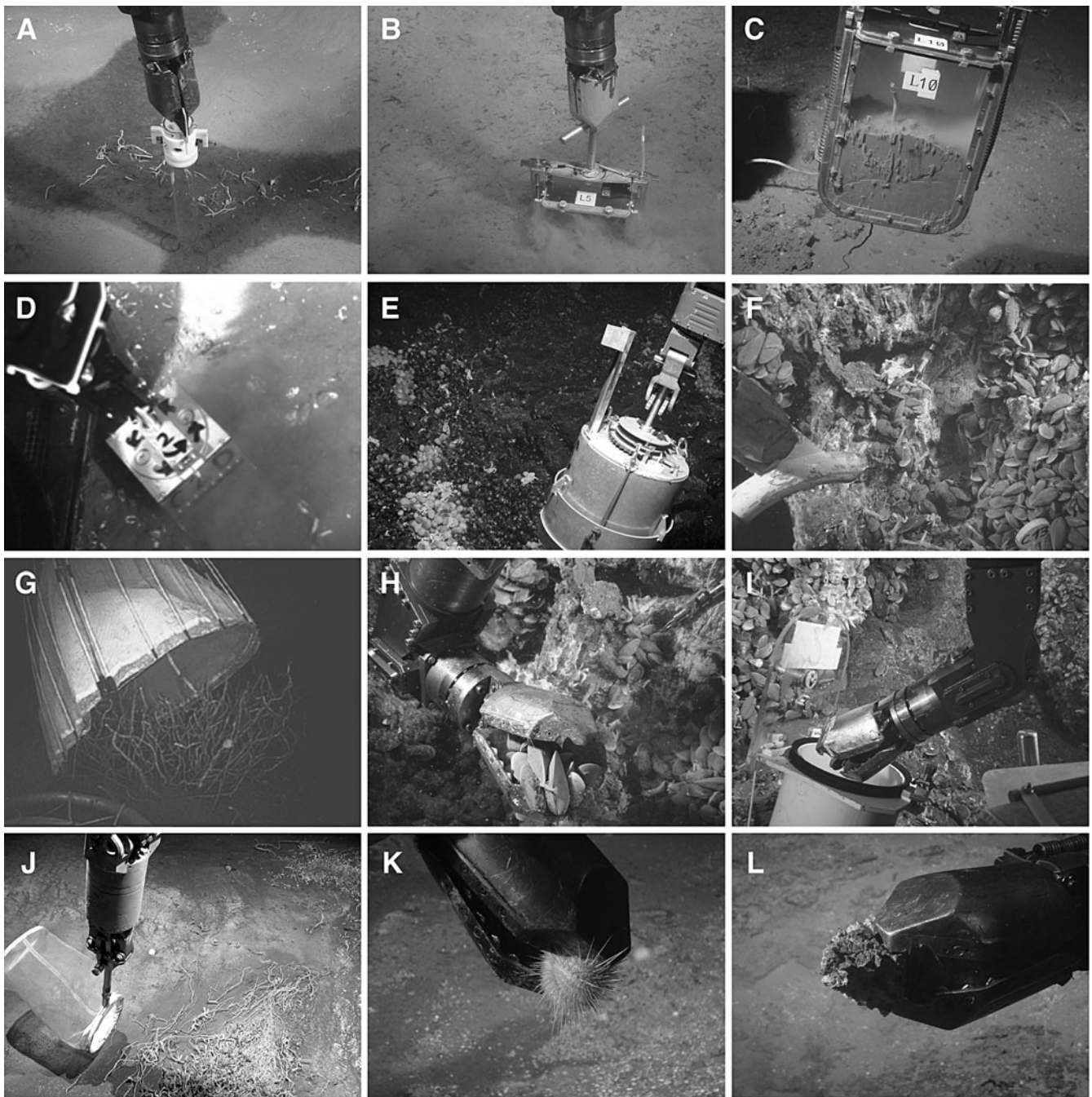


Fig. 1. Common submersible sampling tools used for sampling the fauna in deep-sea chemosynthetic communities: (A) push corer (tube corer) and (B) blade corer, inserted in the sediments, (C) blade corer with a sediment sample, (D) Birge-Ekman box corer (15 × 15 × 20 cm; used extensively in the Atlantic and Pacific Oceans near North America with the submersibles ALVIN and Johnson SeaLink to take quantitative sediment samples; often deployed with 4 metal subcores as inserts, allowing partitioning of sediments for different types of analysis), (E) pot sampler, (F) suction sampler, (G) Bushmaster Sr. used on a tubeworm bush, (H) grab sample in a mussel assemblage, (I) isotherm collection box, (J) scoop sampler, (K) manipulator arm collecting a megafaunal sample (sea urchin), and (L) manipulator arm collecting a carbonate crust. A, K and L were taken on the Napoli mud volcano; B and C on the Amsterdam mud volcano in the Mediterranean Sea during the MEDECO cruise in 2007 (Ifremer); F, H and I on the Tour Eiffel edifice, Lucky Strike vent field, Mid-Atlantic ridge during the MoMARETO cruise in 2006 (Ifremer); D from L. Levin, Scripps Institution of Oceanography; E during a NSF Ridge 2000 cruise to the Eastern Lau Spreading Center in 2005 (C. Fisher, Pennsylvania State University); G during a NSF Johnson Sea Link cruise in the Gulf of Mexico in 2003 (C. Fisher, Pennsylvania State University); and J on the Amon mud volcano in the Mediterranean Sea during the BioNil cruise in 2006 (Marum). See text and Table 1 for additional details

Table 1. Common tools used for sampling chemosynthetic communities from submersibles in the deep sea

Sampling Tool	Area (cm ²)	Substrata	What is sampled
Push corer (tube corer)	Variable	Soft	Epibenthic and infauna
Birge-Ekman box corer	Variable	Soft	Epibenthic and infauna
Blade corer	190 (small) 240 (large)	Soft	Epibenthic and infauna
Pot sampler	531	Soft/Hard Relatively flat surfaces	Mussel beds and associated meiofauna, Infauna
Suction sampler	Variable	Soft/Hard Flat and irregular surfaces	Epibenthic fauna
Arm grabs	Variable	Soft/Hard Flat and irregular surfaces	Epibenthic fauna, Whole organisms, Abiotic substrata, Attached or slow moving organisms
Chimney-master	707	Soft/Hard Flat and irregular surfaces	Bushes of tubeworms and associated meiofauna
Bushmaster Jr.	3848	Soft/Hard Flat and irregular surfaces	Bushes of tubeworms, corals, and associated meiofauna
Bushmaster Sr.	17 671	Soft/Hard Flat and irregular surfaces	Bushes of tubeworms, corals, and associated meiofauna

the underlying species frequency distribution from those resulting from the sampling tools and protocol used. Another major step that may affect comparison between samples and sites is sieving. This obligate step is used to separate different size components of the fauna that are retained on different mesh sizes. The term megafauna usually refers to large size animals that are visible to the naked eye. Macrofauna includes fauna that pass through a 2 cm sieve, but are retained on either a 250 μm (Europe) or 300 μm (North America) sieve. The meiofauna is the compartment where there are more discrepancies. The International Association of Meibenthologists defines meibenthology as the study of small benthic metazoans that pass through a 0.500 mm sieve but are retained on a 0.63 or 0.45 μm sieve (www.meiofauna.org/phyla34.html)! Sieves that have been used to process samples have varied from 74 to 32 μm in chemosynthetic habitats (Copley et al. 2007), and even 20 μm in other deep-sea habitats (Danovaro et al. 2008). These discrepancies tend to obfuscate comparative results. In addition, the meiofaunal compartment in deep-sea hydrothermal vents and other remote marine ecosystems has received much less attention than the larger faunal compartment (Vanreusel et al. 1997, Copley et al. 2007, Danovaro et al. 2008), creating a non-negligible bias in the estimation of diversity. Only recently has the study of meiofauna become almost systematic in vent and seep ecological studies (e.g. Bright et al. 2003, Tsurumi 2003, Sarrazin et al. 2006, Van Gaever et al. 2006, Copley et al. 2007, Limen et al. 2007).

All other things being equal, valid comparisons of diversity between samples, or groups thereof, can only

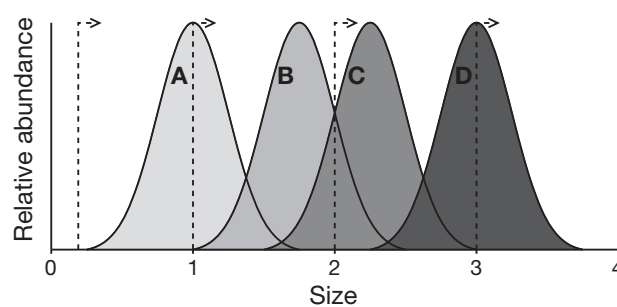


Fig. 2. The selectivity of sieve size and its consequence on observed species richness and species relative abundance. Species A, B, C, and D have different average sizes, but identical Gaussian-shaped size-frequency distributions. They form a perfectly even community ($p_A = p_B = p_C = p_D = 0.25$). Vertical lines represent 4 sieves with different mesh sizes from 0 (finest) to 3 (coarsest). Sieves represent arbitrary partitioning of organisms according to their size. Arrows identify the fraction of the community that is observable for each sieve; the fractions have no upper boundary in this example

be made for identical size components. Fig. 2 illustrates some of the sampling problems that are associated with the use of varying sieve sizes. It shows a fictional and perfectly even community composed of 4 species (i.e. $p_A = p_B = p_C = p_D = 0.25$). These species differ in average size, but otherwise have identical size-frequency distributions. The average body size and sieve size are arbitrary, and only serve to illustrate a point. In this situation, and assuming an exhaustive sampling, different sieve sizes would give markedly different images of this community (Table 2). Sieve size influences sampling results in 2 important ways. Firstly, coarser sieves selectively discard smaller speci-

Table 2. Consequence of sieve sizes on the expected observed species richness (S_{obs}) and relative abundance of observed species (p_i , where $i = A \dots D$) for the hypothetical community in Fig. 2. Relative abundance values are rounded off to add up to 1 for each sieve size

Sieve	S_{obs}	p_A	p_B	p_C	p_D
0	4	0.250	0.250	0.250	0.250
1	4	0.142	0.286	0.286	0.286
2	3	–	0.171	0.329	0.500
3	1	–	–	–	1.000

mens, and are expected to result in lower observed species richness as some species will not be retained on some sieves (e.g. species A with sieve 2; Fig. 2). Secondly (and more subtly), we can expect an underestimation of the relative abundance of the smaller species when mesh size crosses their size-frequency distribution (e.g. species B with sieve 2; Fig. 2). As a corollary, we can expect an overestimation of the relative abundance of larger species. This figure can also be interpreted in terms of the general selectivity differences among sampling tools. For example, relabelling the abscissa as, e.g. mobility, and replacing sieve size with different hypothetical sampling tools, would show that if species differ in terms of mobility, samples obtained using tools that vary in efficiency with the mobility of organisms could lead to very different results, even if said tools were used on the same community. We cannot altogether escape these sampling biases; however, the use of a variety of sampling tools could ultimately lead to better knowledge of the communities. Finally, variation in overall productivity can lead to a shift in average size for particular species and sieve size will always remain somewhat arbitrary. Some investigators have proposed the partitioning of organisms upon both taxonomic and size dependent criteria (e.g. macrofauna sensu strictu, Hessler & Jumars 1974). Nevertheless, a standardization of the size components and sieve sizes to be used in future research would allow comparisons that are more straightforward.

MEASURING DIVERSITY: THE USUAL SUSPECTS

Species (or taxonomic) richness

Observed species richness has been suggested to be the most appropriate measure of diversity (Rosenzweig 1995), and is sometimes regarded as its fundamental measure (Gaston 1996, Martinez 1996). It understandably remains the most often reported measure of diversity since it is a natural result of any field survey. It is also the only measure of species diversity that is available for presence–absence data. Despite its apparent

simplicity, it is the measure that is most dependent on sampling design and intensity, and is greatly affected by the interaction between the evenness of the community and the sampling design. Rare species require more sampling effort to be found, and, by definition, uneven communities contain larger proportions of rare species. Unless we obtain a complete census, S_{obs} is always an underestimation of S_{true} . It is also a very incomplete measure of α -diversity, since it treats all species as equals, regardless of their relative abundance, or importance in terms of processes in the community. A number of numerical methods have been proposed to better estimate S_{true} from a sample. While none seems to be completely satisfactory and all require difficult assumptions about species detection probabilities, they usually provide more accurate lower boundary estimates than S_{obs} (Brose et al. 2003, O'Hara 2005). Finally, mean species richness for a collection of samples is often reported as 'richness' (e.g. Tsurumi 2003, Bergquist et al. 2005, Govenar & Fisher 2007, Levin & Mendoza 2007). However, it is a measure of 'species density', or the mean number of species in a sample (Gotelli & Colwell 2001, Colwell et al. 2004). Species density is an important parameter in itself. It can reflect ecological differences in spatial distributions and faunal densities. However, it can also reflect varying sampling efficiency among habitats. For example, given equal sampling effort in terms of numbers of samples, 2 communities that are identical in all aspects except density of individuals will appear very different with such a measure of 'richness'

Table 3. Observed species richness for samples drawn from constructed communities that are identical in all aspects except density of organisms. Values are mean observed species richness in 1 sample for collections of 1000 samples. All communities contain 12 species ($S_{\text{true}} = 12$) and have identical species abundance distributions ($[p_{[1]}; \dots; p_{[12]}] = [0.4699; 0.2406; 0.0903; 0.0738; 0.0536; 0.0333; 0.0178; 0.0052; 0.0046; 0.0042; 0.0041; 0.0025]$). Variation in density is expressed as a variation in the number of individuals per sample. Mean observed species richness is presented for densities that are increasing at a varying rate to illustrate the diminishing effect of density on S_{obs} for a given S_{true}

Mean sample density (ind. sample ⁻¹)	Mean observed richness (n = 1000 samples)
10	4.12
20	5.50
30	6.28
40	6.83
50	7.24
100	8.52
150	9.22
250	10.12
500	11.29
750	11.70
1000	11.87

(Table 3). This constitutes insufficient information to enable conclusion about differences in richness between these communities. The use of mean number of species by sample in statistical tests of S_{obs} is further discussed later.

In most accounts of biodiversity research, reported richness is usually observed species richness, or indicated as such, but this is not always the case. In reality, reported richness values do not systematically relate to species or even a constant taxonomic level. The lowest taxonomic level of identification is usually inconsistent within or between publications. Causes for such inconsistencies include insufficient prior knowledge of a given group, the discovery of new species, and the lack of resources or time. These estimates thus represent a lower boundary for S_{obs} and this should be acknowledged. Unfortunately, they are sometimes presented as species richness, and all categories are counted 'as is' in richness estimates, taxon-sampling curves (rarefaction, see 'The completeness of inventories: Rarefaction') and other measures of diversity. Precise description of identification level for different groups should be provided (e.g. Sibuet & Olu 1998, Cao et al. 2002a,b, Levin et al. 2003, Tsurumi & Tunnicliffe 2003, Van Dover & Doerries 2005), and 'taxonomic richness' should be preferred when identification levels vary (e.g. Braby et al. 2007).

Shannon diversity index

Shannon's entropy (H) incorporates evenness and richness in a composite measure by weighting each species according to its relative abundance (Shannon 1948):

$$H = -\sum_{i=1}^S p_i \times \log(p_i) \quad (1)$$

The units are nats or bits, depending on whether natural or base 2 logarithms are used, but these are usually discarded in the ecological literature. H is an entropy, or a measure of uncertainty, that originated in information theory, and this has been a source of criticism. However, like measures based on the Gini-Simpson index (see next subsection), it can be interpreted in terms of the probability of interspecific encounters: the probability that 2 individuals, randomly sampled with replacement from the community, belong to different species (Patil & Taillie 1982, Jost 2007). A common justification for its widespread use in ecological research is that it facilitates comparisons with previous studies. However, its real value strongly depends on unknown parameters in the statistical population: the number of unobserved species in the community along with their relative frequencies. Thus, as with S_{obs} , the accuracy of observed H values increases with sampling effort, but to a lesser degree (Chao & Shen 2003).

Gini-Simpson diversity index

The Gini-Simpson index (D) is the probability of interspecific encounter (Gini 1912, Simpson 1949):

$$D = 1 - \lambda = 1 - \sum_{i=1}^S p_i^2 \quad (2)$$

where λ is called Simpson's concentration, which is a measure of the distribution of individuals among species (Simpson 1949). Compared to the Shannon index, it slightly downweights the less abundant species in favour of the most common ones in its calculations. It is also of common use in diversity surveys and has been advocated for assessing biodiversity with limited time and resources (Lande 1996, Lande et al. 2000, Magurran 2004). Its accuracy is less affected by sample size than that of either S_{obs} or H .

The effective number of species

Not all diversity measures are readily interpretable. All other things being equal, ecologists have no trouble grasping that a community with $S_{\text{true}} = 20$ is twice as rich as one with $S_{\text{true}} = 10$ species. It is also clear that in a community with $D = 0.5$, it is twice as likely to randomly draw 2 individuals belonging to different species than in a community with $D = 0.25$. However, in what way does a community with $H = 1$ nats compare to one with $H = 1.22$ nats? The 'effective number of species' of a community lends itself to simpler interpretations; it is the number of species that would make up a hypothetical perfectly even community (i.e. $p_i = N/S$ for all $i = 1, 2, \dots, S$) that is exactly as diverse as the one under consideration (MacArthur 1965). This so-called 'Hill numbers' (Hill 1973), and 'numbers equivalent' of Patil & Taillie (1982) are both expressed as effective numbers of species. It can be easily computed for most diversity indices (MacArthur 1965, Hill 1973, Patil & Taillie 1982, Keylock 2005, Jost 2006, 2007). It would be generally advisable to report effective numbers of species since their units and meaning are readily interpretable. The effective number of species derived from H is $\exp(H)$; formulas for commonly used entropy measures are found in Jost (2006, 2007).

THE COMPLETENESS OF INVENTORIES

Sample coverage

The validity of comparisons of species diversity estimates between different sites or studies depends largely on the completeness, or coverage, of the different inventories. 'Sample coverage' is defined as the fraction of the total number of species that is observed in a survey, i.e.

$C = S_{\text{obs}}/S_{\text{true}}$ (Chao & Shen 2003). An inventory is said to be complete with regards to species richness if all the species in the community have been observed, i.e. if $C = 1$. S_{true} is usually an unknown parameter, and the only way to know it with certainty is to perform a complete census. This is, at best, impractical in studies of biological diversity, particularly in deep-sea environments where accessibility and time are primary constraints. However, notwithstanding their underlying assumptions, richness estimators S_{est} can be used to evaluate C ; $\hat{C} = S_{\text{obs}}/S_{\text{est}}$. Indeed, Brose et al. (2003) showed that among 12 richness estimators, S_{obs} was always the most biased, and the least accurate when a large number of species were undetected. They also concluded that all these estimators underestimated S_{true} , and must be considered as lower boundary estimates. It is beyond the scope of this paper to provide an in-depth review of richness estimators, but nonparametric estimators, notably the Chao and jack-knife estimators, outperform parametric estimators in most situations (Walther & Moore 2005).

Rarefaction

Taxon-sampling curves, or rarefaction curves (sensu Gotelli & Colwell 2001), are also invaluable tools for assessing inventory completeness. They illustrate the increase in the expected, rather than the observed, species richness with increased sampling effort, up to S_{obs} when all collected samples and individuals are accounted for. If the curve reaches an asymptote at S_{obs} , the reliability of this richness estimate is increased. This procedure is also often used to reduce sampling effort down to its lowest value for groups of samples that are to be compared (Hurlbert 1971, Gotelli & Colwell 2001). However, different habitats do not necessarily require the same amount of effort to be adequately sampled. The detection probabilities of different species are not necessarily constant across habitats (Cao et al. 2007) and rarefaction curves are known to intersect in some situations (Lande et al. 2000, Thompson & Withers 2003). Indeed, the initial slope of a rarefaction curve reflects the evenness of the sampled community. For uneven communities, the curve rises gradually toward its final value because most individuals belong to a few dominant species. In contrast, for even communities, a steep increase is observed from the onset. Hence, given sufficient sampling effort, it is expected that 2 taxon-sampling curves will intersect if the species-poor community is more even than the species-rich one (Lande et al. 2000, Thompson & Withers 2003). In this way, these curves not only convey invaluable information about sample coverage, but also provide a graphical representation of the observed richness and evenness of communities, thereby facilitating comparisons.

However, even the use of the same sampling tools and a standardized effort in terms of surface area, volume, or number of samples collected does not guarantee that comparable coverage will be achieved. Non-comparable coverage can result, for example, from differences in the overall density or patchiness between the habitats, or from the interaction between habitat structure and sampling tools (Thompson & Withers 2003, Underwood et al. 2008). As mentioned in the sampling section, different sampling tools have different selectivities, which can be greatly influenced by the structure of the habitat (e.g. hard vs. soft substratum). In this context, rarefaction curves should be regarded as diagnostic instruments to evaluate the sufficiency of sampling effort within habitats, or groups of sites (e.g. Willott 2001). Ideally, to minimize density and contagious distribution effects, and to focus on species richness, these curves should be computed on samples, but plotted against the expected number of individuals in these samples (Gotelli & Colwell 2001). Indeed, the only option with one sample is to compute the curve by accumulating individuals. When many samples are available, the curve can be computed in 1 of 2 ways: the samples are either (1) pooled and the curve is computed on individuals (individual-based rarefaction), or (2) kept separate, and the curve is computed by accumulating samples (sample-based rarefaction). Notably, individual-based rarefaction effectively blurs any spatial or temporal variations in the distribution of organisms. Therefore, in most applications, sample-based rarefaction should be used when possible. The x-axis of the graphical representation can be scaled on samples, individuals or any other measure relating to the size of the sample (volume, area, etc.). Richness comparisons among groups of samples are made by plotting sample-based accumulation results against the mean number of individuals per sample (Gotelli & Colwell 2001). This representation acknowledges the variation in organism densities among groups of samples (e.g. see Figs. 3 & 4). Alternatively, scaling sample-based accumulation results against the number of samples allows for species density comparisons (Gotelli & Colwell 2001). Readers are referred to Gotelli & Colwell (2001) for a more in-depth description of the use and potential pitfalls of rarefaction curves.

Evaluations of the completeness of surveys with taxon-sampling curves and of sample coverage should be performed for every subset of samples that is to be compared (e.g. Van Dover 2002, Bergquist et al. 2005, Van Dover & Doerries 2005). In practice, however, rarefaction is sometimes either not performed (e.g. Sarrazin et al. 2006) or is computed for the entire dataset only (e.g. Tsurumi 2003, Flint et al. 2006). We computed sample-based rarefaction curves for subsets of samples from hydrothermal vents on the Northern East

Pacific Ridge (NEPR) (Fig. 3; data from Tsurumi 2003). In Tsurumi (2003), rarefaction was only carried out for the entire dataset and subsets were ranked according to their diversity, albeit sampling coverage was unequal within these groups (Fig. 3). The reprocessed data show that the curve for the South Rift Zone, for which the number of individuals by sample is low, still exhibits a very steep slope, while that for Ashes, with a larger number of denser samples, is levelling off. Moreover, it is unclear whether the Cleft and Ashes curves would intersect at higher sampling effort, effectively reversing the ordering of these communities with respect to observed species richness. The initial slopes of these curves differ and reveal the community from Ashes to be more even than that from Cleft (Lande et al. 2000). However, estimation of species richness (Table 4) indicates that with further sampling, these 2 sites might be shown to have comparable species richness. Sample coverage estimates indicate that a maximum of 75 to 93% of all species were observed in this study (Tsurumi 2003; Table 4).

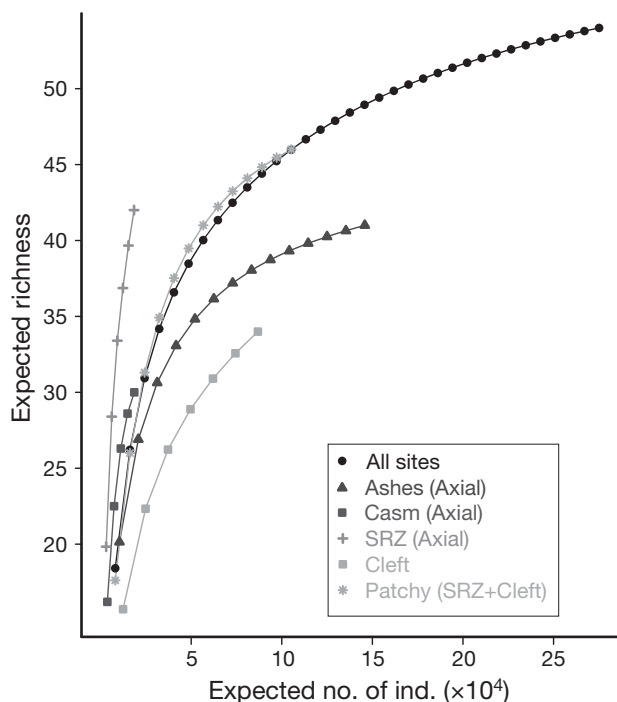


Fig. 3. Sample-based rarefaction curves for groups of samples collected at different vent sites of the Northern East Pacific Ridge (NEPR). Combined macrofaunal (1 mm sieve) and meiofaunal (63 μm sieve) data from vestimentiferan clumps is from Tsurumi (2003). In the original paper, the curve for 'All sites' is compared to that in Grassle & Maciolek (1992) for a very large number of samples from deep-sea sediments across a vast area. Other curves are not presented. Pairwise tests on different aspects of diversity are performed among Cleft, South Rift Zone (SRZ), Cleft+SRZ, and Ashes samples. The 2 samples from Co-Axial are not presented on their own, but are included in the 'All sites' curve

Table 4. Observed taxonomic richness (S_{obs}), number of samples (n_{samples}), number of individuals (n_{ind}), Chao incidence-based estimation of species richness (S_{est}), and sample coverage ($\hat{C} = S_{\text{obs}}/S_{\text{est}}$) for macrofaunal (1 mm sieve) and meiofaunal (63 μm sieve) data from vestimentiferan clumps in Tsurumi (2003). The 2 samples from Co-Axial are included in the 'All sites' sample, but not presented on their own. All taxa are treated as species although a minimum of 8 taxa probably include multiple species. Parentheses: SE of the estimates

Sample	S_{obs}	n_{samples}	n_{ind}	S_{est}	\hat{C} (%)
All sites	54	34	275 125	58.08 (4.05)	93
Ashes (Axial)	41	14	145 810	45.17 (4.88)	91
Casm (Axial)	30	5	18 593	32.72 (2.79)	92
SRZ (Axial)	42	6	18 470	56.00 (9.90)	75
Cleft	34	7	86 818	44.00 (8.37)	77
Patchy (SRZ + Cleft)	46	13	105 288	49.50 (3.50)	93

Coring deeper

Taxon-sampling curves can not only be used to assess the completeness of inventories, but also to evaluate the impact of particular sets of samples or species on diversity. Use of curves can be instructive in partitioning species diversity among separate components to better compare communities from distinct habitats. For example, in a comparison of Florida Escarpment (FE) seep and Snake Pit (SP) vent fauna, Turnipseed et al. (2004) note that 'For sedimentary settings, as at the FE mussel bed, the intent was to sample only the epifaunal mussel bed and its associated community, but some samples penetrated into the mud beneath the mussels'. This is an *a priori* recognition that the sediments on which seep mussel beds rest may constitute a distinct microhabitat that is absent in hard substratum hydrothermal communities; it is also a clear expression of the intention of the authors to compare only the epifaunal components.

The authors identified 5 out of 13 samples that contained some sediment, but the data is analysed without considering this. However, of the 46 taxa that were identified at FE seep, 11 (24%) were present only in samples with sediments. These 'rare' species appeared in no more than 2 samples each and had very low abundances (Table 5). Removing them brings the ratio of total S_{obs} between seeps and vents from 46:23 to 35:23, which is a more reasonable comparison of the epifaunal components. The same change is obtained by entirely removing the 5 samples with sediments. Rarefied values in Turnipseed et al. (2004) likewise supported the view of a doubling in richness from vents to seeps. On the other hand, observed and rarefied richness that were obtained after removing either the species that were found only in samples with sediments or the entire samples with sediments point to a

Table 5. Taxa that only appeared in the 5 samples with sediments from the Florida Escarpment seeps (Turnipseed et al. 2004). Arbitrary tags given in the original paper for incompletely named specimens (A, C, D, 1, 2) are retained here to allow comparison

Phylum	Class	Family	Taxon	No. of samples	Abundance
Annelida	Polychaeta	Capitellidae	<i>Dasybranchus</i> sp. 2	1	1
			<i>Notomastus</i> sp. 2	1	4
		Polynoidae	Polynoid C	2	2
		Pilargidae	<i>Ancistrosyllis</i> sp. 1	1	5
		Cirratulidae	Cirratulid A	1	2
		Ampharetidae	Ampharetid A	1	1
		Terebellidae	<i>Terebellides</i> sp. 1 ^a	1	1
Mollusca	Gastropoda	Xylodisculidae	<i>Xylodiscula</i> n. sp.	1	1
		Cimidae	<i>Cima</i> n. sp.	1	1
Arthropoda	Crustacea	Indeterminate	Isopod C	1	1
			Isopod D	1	1

^aNot tagged as only present in sediments in Table 1 of Turnipseed et al. (2004), but only 1 ind. was recorded from the semi-quantitative scoop sample that contained 0.46 l of sediment

more conservative (~50%) enrichment of epifaunal richness from vents to seeps at this sampling intensity (Fig. 4). The shape of the curves for datasets from which 'sediments-only' species were removed, also suggests that more intensive sampling would provide similar relative gains in richness in FE and SP.

These results support the general trend of a higher richness in seeps than in vents, and the hypothesis that this is a consequence of the lesser toxicity and longer duration of seeps (Sibuet & Olu 1998, Turnipseed et al. 2003, 2004, Levin 2005). However, they also give some evidence that this effect might be better described and understood by partitioning the communities into distinct components before proceeding with comparison since part of the variation in richness is attributable to microhabitat size and diversity. While sedimented hydrothermal sites do exist, the bulk of our current knowledge on hydrothermal ecosystems comes from research on sites with low sedimentation rates that present mostly hard (sulphide or basalt) substrata (Levin et al. 2009). Faunal samples from these ecosystems are often, but not always, limited to epifaunal species (Levin et al. 2009). Except for carbonate crust samples (Fig. 1L), seep samples usually contain some sediments and their associated infauna.

PROFILING DIVERSITY

Inconsistencies can be encountered when trying to order communities in terms of their diversity using different indices; these partly stem from the fact that a multivariate reality has been reduced to a univariate scalar (Patil & Taillie 1979, Tóthmérész 1995). Some information is lost in the process, emphasis is put on different aspects of the multivariate reality, and this

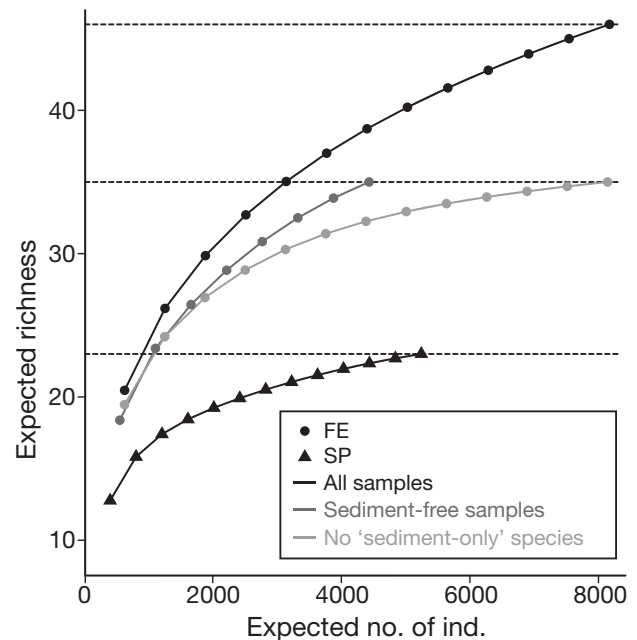


Fig. 4. Sample-based rarefaction curves for fauna from deep-sea mussel beds in cold seeps (Florida Escarpment, FE) and hydrothermal vents (Snake Pit, SP). The cold-seep data are either analysed as a whole (FE – All samples), after removing samples with sediments (FE – Sediment-free samples), or after removing species found only in samples with sediments (FE – No 'sediment-only' species). Data are from Turnipseed et al. (2004). (----) Observed taxonomic richness (S_{obs}) for each group of samples

results in different diversity indices leading to different ordering of the communities (see Patil & Taillie 1979, Tóthmérész 1995 for examples). This has led Tóthmérész (1995) and others (Rousseau et al. 1999, Liu et al. 2007) to further explore the method of 'diversity ordering' proposed by Patil & Taillie (1977, 1979, 1982)

and Solomon (1979). Diversity ordering gives a clearer and more complete picture of species diversity. The first step is to compute a family of diversity indices. Comparisons are then made between communities and a community is said to be more diverse than another if, and only if, this is true according to all of the indices. These indices vary in the weight they assign to abundant and rare species in their quantification of diversity. In this context, 'low order' (also referred to as 'low scale') refers to indices that give more weight to rare species (e.g. S), while 'high order' (also 'high scale') refers to indices that give more weight to dominant species (e.g. the Berger-Parker index, see next subsection). The results are displayed as graphical profiles that allow visual comparisons. Where the lines for 2 communities intersect, they are 'non-comparable' (or 'non-separable'; Liu et al. 2007). This means that such an approach can give a 'partial ordering' of the communities, i.e. an ordering where some uncertainty remains because the position of some of the communities is still undetermined. While this may seem unsatisfactory because it can result in a fuzzy solution, it has the benefit of being more robust than reliance on a few indices.

Liu et al. (2007) have shown that existing diversity ordering methods can be classified within 4 groups. The results from these methods are consistent within groups, but can be incongruent between them. We briefly present here one information-related method; Rényi's generalized entropy (Group 1 in Liu et al. 2007), and one intrinsic diversity-related method, the right tail-sum method (RTS) (Group 4 in Liu et al. 2007).

Rényi's generalized entropy

Rényi's generalized entropy (Rényi 1961, Tóthmérész 1995) is one of the better-known information-based families of indices. ${}^{\alpha}H$ is called the entropy of order α :

$${}^{\alpha}H = \left[\ln \left(\sum_{i=1}^S p_i^{\alpha} \right) \right] / (1 - \alpha) \text{ for } \alpha \geq 0 \text{ and } \alpha \neq 1 \quad (3)$$

It gives the same ordering as 'Hill numbers' (Hill 1973), and numbers equivalent (Patil & Taillie 1979, 1982). Jost (2006, 2007) has shown that the choice of the actual entropy is not important. Entropies and numbers equivalents also give the same ordering (Jost 2006, 2007, Liu et al. 2007). The effective numbers of species for Rényi entropies (${}^{\alpha}D$) are their exponential: ${}^{\alpha}D = \exp({}^{\alpha}H)$. Commonly used diversity indices are special cases of Rényi entropies (Rényi 1961, Hill 1973, Tóthmérész 2005, Jost 2006, 2007). For example, at $\alpha = 0$, we get the natural logarithm of richness (${}^0H = \ln[S]$); its numbers equivalent is ${}^0D = S$. Likewise,

all measures that are based on Simpson's concentration are related to order 2 Rényi' entropy (2H), and all give the same effective number of species (Jost 2006). While ${}^{\alpha}H$ is not defined for $\alpha = 1$, its limit is Shannon's entropy. Finally, ${}^{\infty}H = \ln[1/\max(p_i)]$ is the Berger-Parker dominance index, which depends only on the relative abundance of the dominant species ($\max(p_i)$; Berger & Parker 1970). In summary, while low-order measures emphasize the importance of rare species regardless of relative abundance, higher-order measures gradually downweight the contribution of rare species. The pivotal point is found with measures of order 1, like Shannon's entropy, for which the weight given to each species is its relative abundance.

Right tail-sum

The right tail-sum method (RTS; Patil & Taillie 1979, 1982) is a representative of the family of intrinsic diversity-related methods. As defined by these authors, community A is said to be intrinsically more diverse than community B (i.e. $A \geq B$) if A can be derived from B by performing a finite sequence of (1) adding species to B , (2) transferring abundance from a more abundant to a less abundant species in B , and (3) relabelling the species in B . The right tail-sum of rank i (T_i) is obtained by first ranking species in terms of decreasing relative abundance and computing:

$$T_i = \sum_{j=i+1}^S p_{[j]} \text{ for } i = 1, 2, \dots, S-1 \quad (4)$$

where $p_{[j]}$ is the relative abundance of the j^{th} most abundant species. RTS plots are easy to understand, and reflect a clear definition of species diversity.

RTS is connected to other intrinsic diversity-related methods. It is equivalent to logarithmic dominance plots (Tóthmérész 1995), and complementary to majorization (Solomon 1979) and k -dominance plots (Shaw et al. 1983; see Liu et al. 2007 for a concise description of these and other diversity ordering methods). These methods all give the same ordering of communities. In RTS and logarithmic dominance plots, the curves of the more diverse communities are above those of the less diverse ones, just as in Rényi's generalized entropy profiles. The opposite is true in majorization and k -dominance plots. Intrinsic diversity-based methods are the most stringent: communities that can be ordered by intrinsic diversity can also be ordered by other methods but the reverse is not necessarily true (Liu et al. 2007). Tóthmérész (1995) recommends RTS ordering for species-poor communities, and the results of Liu et al. (2007) indicate that RTS might be the method of choice for all communities given its straightforward interpretation and the inconsistencies of results given by other groups of methods.

Profiles in practice

Diversity ordering has appeared in diverse fields of ecology (Lewis et al. 1988, Kendall & Aschan 1993, Gove et al. 1995, Solomon & Gove 1999, Ricotta et al. 2002, 2003, Danovaro et al. 2008, 2009), but apart from a few notable exceptions, it has not become common practice in deep-sea chemosynthetic environments (Zekely et al. 2006a,b, Copley et al. 2007, Galkin & Goroslavskaya 2008). Carney (2007) makes an appeal for the use of multiple indices that give variable weights to rare and abundant species. He also points out that 'ease of computation has made this the standard approach in benthic ecology' but, apart from a reference to Hill (1973), makes no mention of diversity profiles in his timely historical review. Similarly, Tóthmérész (1995) states that diversity profiles are not of widespread use in ecology because 'they involve more calculations than a simple diversity index' and 'few of these methods are included in standard computer packages'. While these reasons might explain why we do not see more diversity profiles in published work, they are only of a practical nature. The widely distrib-

uted and freely available EstimateS (Colwell 2006) already performs rarefaction for species richness, Shannon's entropy, and the inverse of Simpson's concentration ($1/\lambda$). Finally, the R language provides access to numerous functions for the computation and plotting of a variety of classical and rarefied diversity measures (R Development Core Team 2008). The vegan (Oksanen et al. 2008) and BiodiversityR (Kindt & Coe 2005) libraries are useful in producing Rényi and Hill profiles, as well as accumulation curves for all orders.

As an electronic supplement to this paper, we provide an R source file defining functions to easily compute and plot intrinsic diversity profiles for groups of samples (www.int-res.com/articles/suppl/m402p285_app/). Along with functions in the Biodiversity R package (Kindt & Coe 2005), these were used to produce the diversity profiles in Fig. 5 for data from the Unimak Margin, Alaska (at depths >3000 m), that was taken from the first published seep macrofaunal data set (Levin & Mendoza 2007, their Appendix 3). Richness and numbers equivalents of Shannon's entropy and Simpson's diversity can be read on the numbers equiv-

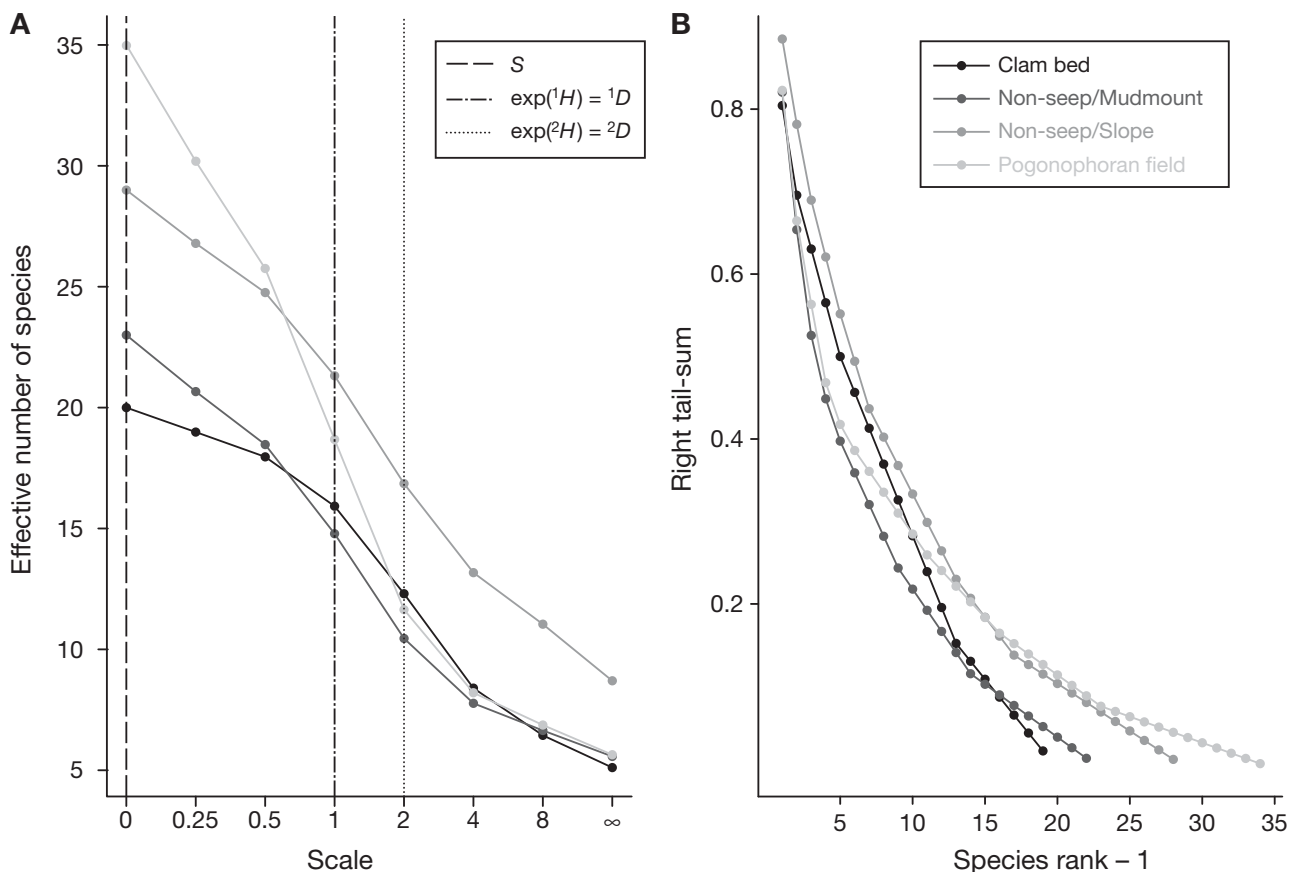


Fig. 5. Effective number of species (A) and right tail-sum (RTS) (B) diversity profiles for data from the Unimak Margin seep (Alaska; Levin & Mendoza 2007). Note that the x-axis is distorted in (A) in order to illustrate variation over the most commonly used range of orders. S : richness; 1D and 2D : effective numbers of species for Rényi entropies of order 1 and 2 respectively 1H and 2H

alents profile (Fig. 5A) for orders of 0, 1 and 2, respectively. Except for non-seep samples, they do not allow a coherent ordering of communities. The profiles, however, reveal that the most important differences lie in the number of observed species: the value at an order of 0 in Fig. 5A, and the length of the 'tail' in Fig. 5B. For example, non-seep samples from slopes are more diverse than those on mud mounds mostly because they harbour more species, although both have similarly shaped profiles. Unfortunately, we were unable to assess the completeness of this survey because the published data table contains mean densities per core. We assumed comparable sampling to illustrate the usefulness of diversity profiles.

Rarefying profiles

Rarefaction is usually performed only for species richness. It often reveals that some groups of samples clearly have not levelled off, but tests are nonetheless performed on the total observed species richness,

Shannon entropy, or other measures of diversity. However, as shown in Fig. 6, the bias and accuracy of all measures of diversity are naturally affected by the completeness of inventories. Moreover, low-order entropies and Hill numbers (Fig. 6A), and high rank values in intrinsic diversity profiles such as RTS (Fig. 6B), including richness itself are the most affected.

This sampling effect is expected from the definition of these measures, and has been illustrated before for measures other than richness (Soetaert & Heip 1990). Previous research has advocated the use of the Gini-Simpson index, rather than its Shannon counterpart, for assessing biodiversity with limited time and resources because it converges more rapidly (Lande 1996, Lande et al. 2000, Magurran 2004). While this recommendation is generally applicable in practice, chemosynthetic communities are known to be uneven and can harbour numerous rare species (e.g. Bergquist et al. 2003, 2005, Smith & Baco 2003, Zekely et al. 2006a, Copley et al. 2007, Gollner et al. 2007, Galkin & Goroslavskaya 2008). Since the latter are difficult to sample, low-order measures can give positively mis-

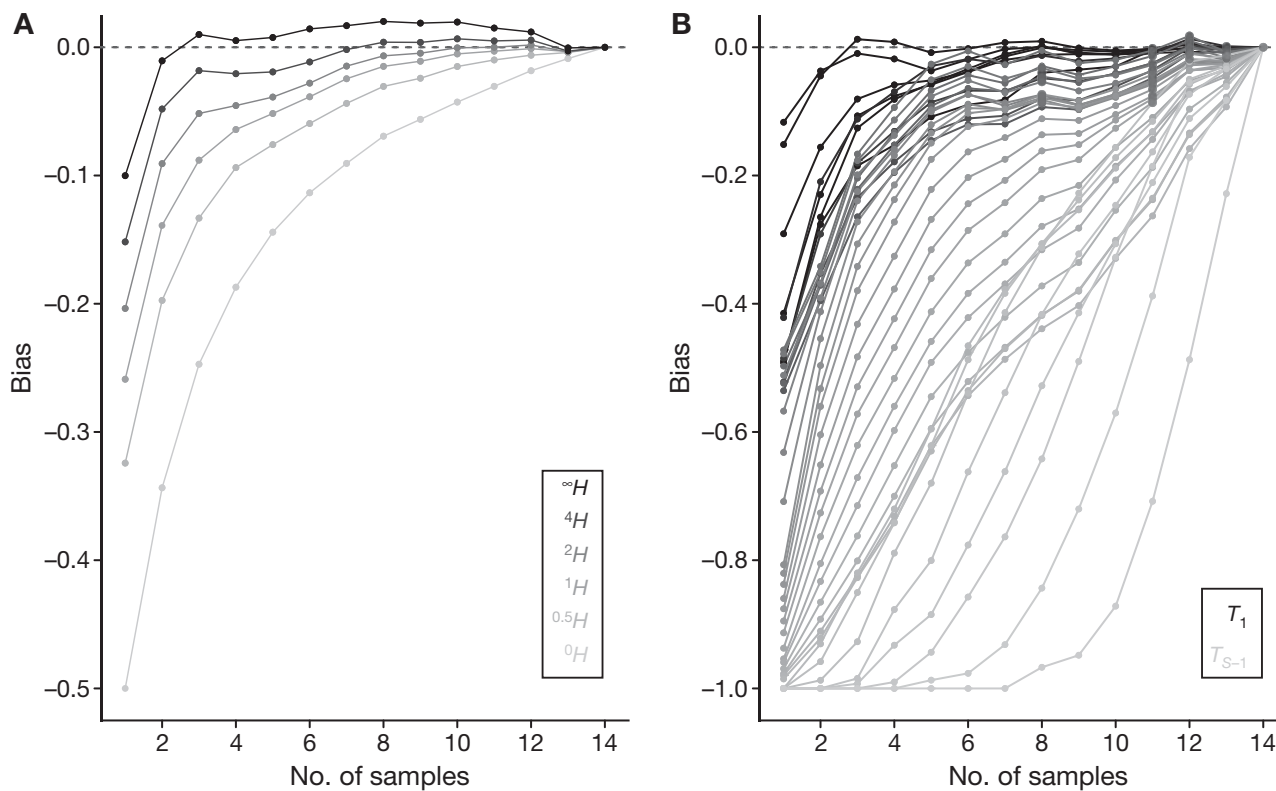


Fig. 6. Bias in diversity indices with increasing sample size. (A) Rényi entropies (${}^{\infty}H$) of orders 0, 0.5, 1, 2, 4 and infinity, (B) right tail-sum (T) intrinsic diversity of ranks 1, 2, ..., $S-1$. Lighter lines correspond to measures that are most affected by rare species (e.g. $\ln[S] = {}^0H$, observed taxonomic richness), darker lines to those that are affected primarily by the most abundant species (e.g. T_1). For both types of measures, all orders are affected by sampling intensity. Bias is expressed as the proportional deviation from the 'true' value (Bias = (estimate-true)/true; Walther & Moore 2005). Data used are the 14 samples from the Axial segment of the Northern East Pacific Ridge (NEPR) in Tsurumi (2003). Diversity measures for the entire dataset are treated as the 'true' values

leading results, and rarefaction should be applied to 'all' diversity measures (i.e. even those that down-weight rare species more than the Gini-Simpson index). This is particularly important when rarefaction performed for richness indicates unequal completeness, which is a common conclusion in deep-sea research. Rarefaction allows a more robust and complete ordering based on the data at hand. If diversity profiles overlap (we have reason to believe that the accuracy of these estimations varies between groups), we can make preliminary conclusions based on the ordering of profiles at higher orders (or lower ranks). Rarefaction also allows bringing these estimates down to a common sampling effort for curves that have not levelled off.

When comparing 2 or more communities, it can be misleading to draw conclusions on low-order diversities without assessing completeness regardless of whether sampling has been homogeneous or not. However, high-order/low-rank measures are usually much less informative of the overall community diversity. Computing rarefied diversity profiles obviously requires more computations and data manipulation, but tools are available for the calculation of Rényi entropies and effective numbers of species in the BiodiversityR package (Kindt & Coe 2005). Furthermore, functions to compute rarefied intrinsic diversity profiles, including RTS, are also included in the electronic supplement.

THE MEANING OF DIFFERENCES

It has become common to look for statistical differences in average richness or other measures of diversity (e.g. Shannon's index or Simpson's concentration) with parametric or bootstrapped *t*-tests or nonparametric *U*-tests (e.g. Levin et al. 2000, Turnipseed et al. 2003, 2004, Zekely et al. 2006a,b, Gollner et al. 2007, Galkin & Goroslavskaya 2008). The purpose of such tests is usually to assess whether different groups of samples exhibit differences in these characteristics, and to rank them accordingly. The results of these analyses are interesting in their own right. However, it is important to point out that these tests do not allow the ranking of communities with regards to the desired parameter (richness, Shannon diversity, etc.). Indeed, what is being tested is whether there are significant differences in the average values of the parameter for single samples. Thus, the tests are performed on the worst estimates available that correspond to the left-most values in rarefaction plots.

For example, a good estimation of species richness usually requires repeated sampling within the community. Mean richness in a single sample is not only affected by total richness but also by overall density,

evenness, the patchiness of distributions and sampling design (Tables 2 & 3, Figs. 2 & 7). Moreover, as explained above, rarefaction curves can intersect with an increase in sampling effort (Lande et al. 2000, Thompson & Withers 2003). Fig. 7 illustrates a hypothetical situation where testing for average differences leads to spurious conclusions. It shows rarefaction curves for richness (Fig. 7A) and the number equivalent of Shannon's entropy (Fig. 7B) for 2 collections of 10 samples with a total richness of $S_A = 20$ and $S_B = 11$, respectively, and an equal Shannon entropy of ${}^1H_A = {}^1H_B = 1.84$ ($\exp(1.84) = 6.3$ species). Significant differences are found for the average value of both parameters (S : $t = -21.019$, $p < 0.001$; $\exp({}^1H)$: $t = -14.43$, $p < 0.001$). An incorrect interpretation of these results would be that B is richer and has higher diversity of order 1 than A when this is clearly not the case. The correct statistical interpretation is that the mean observed richness in one sample from B is greater than that in one sample from A, and this can lead to different ecological inter-

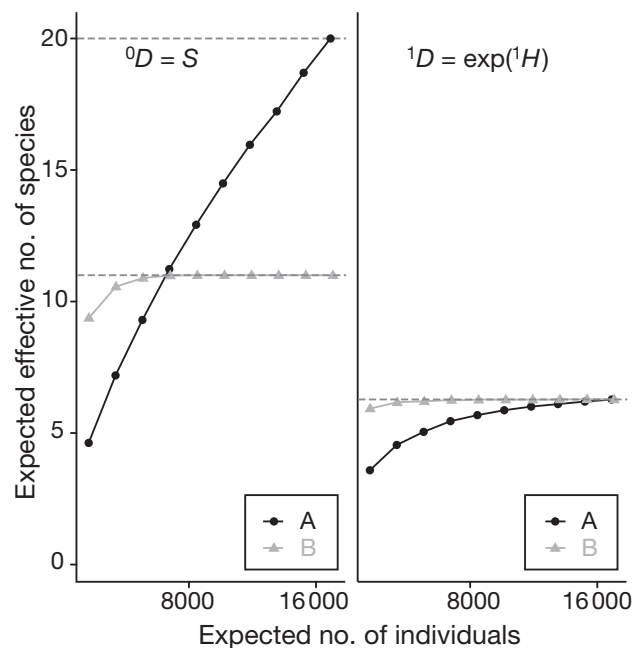


Fig. 7. Sample-based rarefaction curves for samples drawn from 2 different communities illustrating the problem of testing mean differences between groups of samples. S : richness (${}^0D = S$); $\exp({}^1H)$: effective number of species of order 1 (${}^1D = \exp[{}^1H]$). Pooled samples from communities A and B both have $\exp({}^1H) = 6.3$, but $S_A = 20$ and $S_B = 11$, as indicated by dashed lines. In A, majority of species are rare, appearing in only 1 sample. This reflects very patchy distributions. In B, species appear in nearly all samples. Parametric (*t*-test) and nonparametric (*U*-test) tests performed on these indices evaluate the significance of differences in the worst estimation of diversity available from the survey (value for 1 sample, left-hand side of each graph) and do not reflect actual differences in richness or diversity of order 1. Significant differences can stem from differences in overall density or spatial distribution. Example data are from Tóthmérész (1998)

pretations. In this constructed example, we know that species exhibit different distribution patterns in communities A and B. Nearly all species are present in each sample from B (homogeneous distributions) while only a few are present in each sample from A (contagious distributions). While this example is undeniably of an extreme nature, it nevertheless illustrates the fallacy of using such tests to draw conclusions on species diversity ranking of communities.

CONCLUSION

We have examined the current state of practices in species diversity measurement and comparison based on quantitative samples from deep-sea chemosynthetic communities. These data are becoming both more frequent and extensive, and our analysis has revealed some problems in their use and analysis. In light of our conclusions, we make 5 recommendations for future and ongoing studies:

(1) Be specific. When different taxonomic levels are reported, this should always be clearly stated. Species richness should be used only when all reported taxa are species. Taxonomic richness is vague enough that it can be used even when mixed levels are found, but should always come with a detailed account of actual taxonomic levels. In this instance, it should be acknowledged that these values represent lower bound estimates of S_{true} . Ideally, all data tables should be made available after publication through the use of online supplements, or centralized databases.

(2) Avoid the protocol conundrums. Numerical data treatment is never a panacea. It is imperative to consider all aspects of sampling before comparing results. Some further efforts should be made within the research community to use standardized size compartments and sieves.

(3) Embrace diversity. Rather than restricting the description of species diversity to just a few indices, use diversity profiles. Emphasize comparisons at those orders for which estimations are more reliable. Intrinsic diversity-related methods are readily interpretable and stricter compared to other methods. They can be computed, plotted and rarefied in R using the functions provided as an online supplement.

(4) Rarefy everything. Or, at least, consider that all species diversity measures based on richness and evenness are more or less affected by sampling effort and inventory completeness. Evaluate how robust these estimates are in groups of samples of interest through rarefaction curves. If the curves have not levelled off, gather more samples or use caution in your interpretation.

(5) Pass on the tests. Beware of tests that do not answer the biological question of interest. The results

of tests of average differences in species diversity can actually reflect differences in many parameters other than species diversity itself. A good estimate of species diversity requires repeated samplings (or one very large sample). Mean values for one sample are the worst estimates available.

Our knowledge of species diversity in deep-sea chemosynthetic communities has come a long way since their discovery in the late 70s. These communities appear to display opposite patterns of diversity and biomass from their abyssal counterparts. While they have only recently been discovered and some of the species they harbour are still unknown to science, chemosynthetic ecosystems are already threatened by current or upcoming anthropogenic impacts such as deep-sea fisheries, exploration and exploitation for mineral and hydrocarbon sources as well as global warming. The development and use of approaches that allow more accurate and robust assessments and comparisons of their species diversity is a necessary step towards their management and protection. It is our hope that the views and tools presented here will contribute positively to species diversity research in deep-sea chemosynthetic communities.

Acknowledgements. We thank P. Girard, M. Morineaux, B. Ritt, and 2 anonymous reviewers for constructive comments on an earlier version of this manuscript. This work was made possible by ANR Deep Oases (ANR06 BDV005) and Ifremer post-doctoral fellowships to O.G.; it also received support from the GDR Ecchis, European network MARBEF (FP6-GOCE-CT-2003-505446) and the CoML CHESs project.

LITERATURE CITED

- Baco AR, Smith CR (2003) High species richness in deep-sea chemoautotrophic whale skeleton communities. *Mar Ecol Prog Ser* 260:109–114
- Bayon G, Loncke L, Dupré S, Caprais JC and others (2009) Multi-disciplinary investigation of fluid seepage on an unstable margin: the case of the Central Nile deep sea fan. *Mar Geol* 261:92–104
- Berger WH, Parker FL (1970) Diversity of planktonic Foraminifera in deep sea sediments. *Science* 168:1345–1347
- Bergquist DC, Urcuyo IA, Fisher CR (2002) Establishment and persistence of seep vestimentiferan aggregations on the upper Louisiana slope of the Gulf of Mexico. *Mar Ecol Prog Ser* 241:89–98
- Bergquist DC, Ward T, Cordes EE, McNelis T and others (2003) Community structure of vestimentiferan-generated habitat islands from Gulf of Mexico cold seeps. *J Exp Mar Biol Ecol* 289:197–222
- Bergquist DC, Fleckenstein C, Knisel J, Begley B, MacDonald IR, Risher CR (2005) Variations in seep mussel bed communities along physical and chemical environmental gradients. *Mar Ecol Prog Ser* 293:99–108
- Bergquist DC, Eckner JT, Urcuyo IA, Cordes EE, Hourdez S, Macko SA, Fisher CR (2007) Using stable isotopes and quantitative community characteristics to determine a local hydrothermal vent food web. *Mar Ecol Prog Ser* 330:49–65

- Braby CE, Rouse GW, Johnson SB, Jones WJ, Vrijenhoek RC (2007) Bathymetric and temporal variation among *Osedax* boneworms and associated megafauna on whale-falls in Monterey Bay, California. *Deep-Sea Res II* 54:1773–1791
- Bright M, Arndt C, Keckeis H, Felbeck H (2003) A temperature-tolerant interstitial worm with associated epibiotic bacteria from the shallow water fumaroles of Deception Island, Antarctica. *Deep-Sea Res II* 50:1859–1871
- Brose U, Martinez ND, Williams RJ (2003) Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* 84:2364–2377
- Cao Y, Larsen D, Hughes R, Angermeier P, Patton TM (2002a) Sampling effort affects multivariate comparisons of stream communities. *J N Am Benthol Soc* 21:701–714
- Cao Y, Williams D, Larsen D (2002b) Comparison of ecological communities: the problem of sample representativeness. *Ecol Monogr* 72:41–56
- Cao Y, Hawkins CP, Larsen DP, Van Sickle J (2007) Effects of sample standardization on mean species detectabilities and estimates of relative differences in species richness among assemblages. *Am Nat* 170:381–395
- Carney R (2007) Use of diversity estimations in the study of sedimentary benthic communities. *Oceanogr Mar Biol Annu Rev* 45:139–172
- Chao A, Shen TJ (2003) Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ Ecol Stat* 10:429–443
- Colwell RK (2006) EstimateS: statistical estimation of species richness and shared species from samples, version 8.2.0. <http://viceroy.eeb.uconn.edu/estimates>
- Colwell RK, Mao CX, Chang J (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85:2717–2727
- Copley JTP, Flint HC, Ferrero TJ, Van Dover CL (2007) Diversity of meiofauna and free-living nematodes in hydrothermal vent mussel beds on the northern and southern East Pacific Rise. *J Mar Biol Assoc UK* 87:1141–1152
- Cordes EE, Hourdez H, Predmore BL, Redding ML, Fisher CR (2005) Succession of hydrocarbon seep communities associated with the long-lived foundation species *Lamellibrachia luymesii*. *Mar Ecol Prog Ser* 305:17–29
- Danovaro R, Gambi C, Lampadariou N, Tselepides A (2008) Deep-sea nematode biodiversity in the Mediterranean basin: testing for longitudinal, bathymetric and energetic gradients. *Ecography* 31:231–244
- Danovaro R, Corinaldesi C, Luna GM, Magagnini M, Manini E, Pusceddu A (2009) Prokaryote diversity and viral production in deep-sea sediments and seamounts. *Deep-Sea Res II* 56:738–747
- Flint HC, Copley JTP, Ferrero TJ, Van Dover CL (2006) Patterns of nematode diversity at hydrothermal vents on the East Pacific Rise. *Cah Biol Mar* 47:365–370
- Galkin SV, Goroslavskaya EI (2008) Bottom fauna associated with mussel beds and alvinellid communities in the hydrothermal field at 9° N of the East Pacific Rise. *Oceanology* 48:509–516
- Gaston KJ (1996) Species richness: measure and measurement. In: Gaston KJ (ed) *Biodiversity: a biology by numbers and difference*. Blackwell Science, Oxford, p 77–113
- Gini C (1912) *Variabilità e Mutabilità*. Tipografia di Paolo Cuppini, Bologna
- Gollner S, Zekely J, Govenar B, Le Bris N, Nemeschkal HL, Fisher CR, Bright M (2007) Tubeworm-associated permanent meiobenthic communities from two chemically different hydrothermal vent sites on the East Pacific Rise. *Mar Ecol Prog Ser* 337:39–49
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species. *Ecol Lett* 4:379–391
- Gove JH, Patil GP, Taillie C (1995) A mathematical programming model for maintaining structural diversity in uneven-aged forest stands with implications to other formulations. *Ecol Modell* 79:11–19
- Govenar B, Fisher CR (2007) Experimental evidence of habitat provision by aggregations of *Riftia pachyptila* at hydrothermal vents on the East Pacific Rise. *PSZNI: Mar Ecol* 28:3–14
- Grassle JF, Maciolek NJ (1992) Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *Am Nat* 139:313–341
- Hessler RR, Jumars PA (1974) Abyssal community analysis from replicate box cores in the central North Pacific. *Deep-Sea Res* 21:185–209
- Hill M (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427–432
- Hourdez S, Lamontagne J, Peterson P, Weber RE, Fisher CR (2000) Hemoglobin from a deep-sea hydrothermal-vent copepod. *Biol Bull* 199:95–99
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577–586
- Jost L (2006) Entropy and diversity. *Oikos* 113:363–375
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427–2439
- Kendall MA, Aschan M (1993) Latitudinal gradients in the structure of macrobenthic communities: a comparison of arctic, temperate and tropical sites. *J Exp Mar Biol Ecol* 172:157–169
- Keylock C (2005) Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. *Oikos* 109:203–207
- Kindt R, Coe R (2005) *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), Nairobi
- Lande R (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76:5–13
- Lande R, DeVries PJ, Walla TR (2000) When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos* 89:601–605
- Levin LA (2005) Ecology of cold seep sediments: interactions of fauna with flow, chemistry and microbes. *Oceanogr Mar Biol Annu Rev* 43:1–46
- Levin LA, Mendoza GF (2007) Community structure and nutrition of deep methane-seep macrobenthos from the North Pacific (Aleutian) Margin and the Gulf of Mexico (Florida Escarpment). *PSZNI: Mar Ecol* 28:131–151
- Levin LA, Gage JD, Martin C, Lamont PA (2000) Macrobenthic community structure within and beneath the oxygen minimum zone, NW Arabian Sea. *Deep-Sea Res II* 47:189–226
- Levin LA, Ziebis W, Mendoza GF, Growney VA and others (2003) Spatial heterogeneity of macrofauna at northern California methane seeps: influence of sulfide concentration and fluid flow. *Mar Ecol Prog Ser* 265:123–139
- Levin LA, Mendoza GF, Konotchick T, Lee R (2009) Macrobenthos community structure and trophic relationships within active and inactive Pacific hydrothermal sediments. *Deep-Sea Res II* 56:1632–1648
- Lewis CE, Swindel BF, Tanner GW (1988) Species diversity and diversity profiles: concept, measurement, and application to timber and range management. *J Range Manage* 41:466–469
- Limen H, Levesque C, Jumiper SK (2007) POM in macro/meiofaunal food webs associated with three flow regimes

- at deep-sea hydrothermal vents on Axial Volcano, Juan de Fuca Ridge. *Mar Biol* 153:129–139
- Liu C, Whittaker RJ, Ma K, Malcolm JR (2007) Unifying and distinguishing diversity ordering methods for comparing communities. *Popul Ecol* 49:89–100
- MacArthur R (1965) Patterns of species diversity. *Biol Rev Camb Philos Soc* 40:510–533
- Magurran AE (2004) *Measuring biological diversity*. Blackwell Science, Oxford
- Martinez ND (1996) Defining and measuring functional aspects of biodiversity. In: Gaston KJ (ed) *Biodiversity: a biology of numbers and difference*. Blackwell Science, Oxford, p 114–148
- McIntosh RP (1967) An index of diversity and the relation of certain concepts to diversity. *Ecology* 48:392–404
- O'Hara RB (2005) Species richness estimators: How many species can dance on the head of a pin? *J Anim Ecol* 74: 375–386
- Oksanen J, Kindt R, Legendre P, O'Hara RB, Simpson GL, Stevens MHH (2008) *vegan: community ecology package*. R package version 1.11-0. <http://cran.r-project.org/package=vegan>
- Patil GP, Taillie C (1977) Diversity as a concept and its implications for random communities. *Bull Int Stat Inst* 47:497–515
- Patil GP, Taillie C (1979) A study of diversity profiles and orderings for a bird community in the vicinity of Colstrip, Montana. In: Patil GP, Rosenzweig M (eds) *Contemporary quantitative ecology and related ecometrics*. International Cooperative Publishing House, Fairland, MD, p 23–48
- Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Assoc* 77:548–561
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405:212–219
- R Development Core Team (2008) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. www.r-project.org
- Rényi A (1961) On measures of entropy and information. *Proc 4th Berkeley Symp Mathematical Statistics and Probability* 1:547–561
- Ricotta C, Pacini A, Avena G (2002) Parametric scaling from species to growth-form diversity: an interesting analogy with multifractal functions. *Biosystems* 65:179–186
- Ricotta C, Avena G, Chiarucci A (2003) An index of divergence from species to life-form diversity based on the notion of intrinsic diversity ordering. *Plant Ecol* 165:217–222
- Rosenzweig ML (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge
- Rousseau R, Van Hecke P, Nijssen D, Bogaert J (1999) The relationship between diversity profiles, evenness and species richness based on partial ordering. *Environ Ecol Stat* 6:211–223
- Rowe GT, Clifford CH (1973) Modifications of the Birge-Ekman box corer for use with SCUBA or deep submergence research vessels. *Limnol Oceanogr* 18:172–175
- Sarrazin J, Juniper SK (1999) Biological characteristics of a hydrothermal edifice mosaic community. *Mar Ecol Prog Ser* 185:1–19
- Sarrazin J, Robigou V, Juniper SK, Delaney JR (1997) Biological and geological dynamics over four years on a high-temperature sulfide structure at the Juan de Fuca Ridge hydrothermal observatory. *Mar Ecol Prog Ser* 153:5–24
- Sarrazin J, Walter C, Sarradin PM, Brind'Amour A and others (2006) Community structure and temperature dynamics within a mussel assemblage on the southern East Pacific Rise. *Cah Biol Mar* 47:483–490
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423, 623–656
- Shaw KM, Lamshead PJD, Platt HM (1983) Detection of pollution-induced disturbance in marine benthic assemblages with special reference to nematodes. *Mar Ecol Prog Ser* 11:195–202
- Sibuet M, Olu K (1998) Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. *Deep-Sea Res II* 45:517–567
- Simpson EH (1949) Measurement of diversity. *Nature* 163:688
- Smith CR, Baco AR (2003) Ecology of whale falls at the deep-sea floor. *Oceanogr Mar Biol Annu Rev* 41:311–354
- Soetaert K, Heip C (1990) Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Mar Ecol Prog Ser* 59:305–307
- Solomon DL (1979) A comparative approach to species diversity. In: Grassle JF, Patil GP, Smith W, Taillie C (eds) *Ecological diversity in theory and practice*. International Cooperative Publishing House, Fairland, MD, p 29–35
- Solomon DS, Gove JH (1999) Effects of uneven-age management intensity on structural diversity in two major forest types in New England. *For Ecol Manage* 114:265–274
- Sutherland WJ (1996) The twenty commonest censusing sins. In: Sutherland WJ (ed) *Ecological census techniques*. Cambridge University Press, Cambridge, p 317–318
- Thompson GG, Withers P (2003) Effect of species richness and relative abundance on the shape of the species accumulation curve. *Austral Ecol* 28:355–360
- Tóthmérész B (1995) Comparison of different methods for diversity ordering. *J Veg Sci* 6:283–290
- Tóthmérész B (1998) On the characterization of scale-dependent diversity. *Abstr Bot* 22:149–156
- Tóthmérész B (2005) Diversity characterizations in R. In: Lóvei GL, Toft S (eds) *European carabidology 2003*. DIAS Reports 114:333–344
- Tsurumi M (2003) Diversity at hydrothermal vents. *Glob Ecol Biogeogr* 12:181–190
- Tsurumi M, Tunnicliffe V (2003) Tubeworm-associated communities at hydrothermal vents on the Juan de Fuca Ridge, northeast Pacific. *Deep-Sea Res I* 50:611–629
- Turnipseed M, Knick K, Lipcius R, Dreyer J (2003) Diversity in mussel beds at deep-sea hydrothermal vents and cold seeps. *Ecol Lett* 6:518–523
- Turnipseed M, Jenkins C, Van Dover CL (2004) Community structure in Florida Escarpment seep and Snake Pit (Mid-Atlantic Ridge) vent mussel beds. *Mar Biol* 145:121–132
- Underwood AJ, Chapman MG, Cole VJ, Palomo MG (2008) Numbers and density of species as measures of biodiversity on rocky shores along the coast of New South Wales. *J Exp Mar Biol Ecol* 366:175–183
- Urcuyo IA, Massoth GJ, Julian D, Fisher CR (2003) Habitat, growth and physiological ecology of a basaltic community of *Ridgeia piscesae* from the Juan de Fuca Ridge. *Deep-Sea Res I* 50:763–780
- Van Dover CL (2002) Community structure of mussel beds at deep-sea hydrothermal vents. *Mar Ecol Prog Ser* 230: 137–158
- Van Dover CL, Doerries MB (2005) Community structure in mussel beds at Logatchev hydrothermal vents and a comparison of macrofaunal species richness on slow- and fast-spreading mid-ocean ridges. *PSZNI: Mar Ecol* 26:110–120
- Van Dover CL, Trask JL (2000) Diversity at deep-sea hydrothermal vent and intertidal mussel beds. *Mar Ecol Prog Ser* 195:169–178
- Van Gaever S, Moodley L, de Beer D, Vanreusel A (2006) Meiobenthos at the Arctic Håkon Mosby Mud Volcano, with a parental-caring nematode thriving in sulphide-rich sediments. *Mar Ecol Prog Ser* 321:143–155

- Vanreusel A, Van den Bossche I, Thiermann F (1997) Free-living marine nematodes from hydrothermal sediments: similarities with communities from diverse reduced habitats. *Mar Ecol Prog Ser* 157:207–219
- Walther BA, Moore JL (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28:815–829
- Warwick RM, Clarke KR (1996) Relationships between body-size, species abundance and diversity in marine benthic assemblages: Facts or artefacts? *J Exp Mar Biol Ecol* 202:63–71
- Willott SJ (2001) Species accumulation curves and the measure of sampling effort. *J Appl Ecol* 38:484–486
- Zekely J, Van Dover CL, Nemeschkal H, Bright M (2006a) Hydrothermal vent meiobenthos associated with mytilid mussel aggregations from the Mid-Atlantic Ridge and the East Pacific Rise. *Deep-Sea Res I* 53:1363–1378
- Zekely J, Gollner S, Van Dover CL, Govenar B, Le Bris N, Nemeschkal HL, Bright M (2006b) Nematode communities associated with tubeworm and mussel aggregations on the East Pacific Rise. *Cah Biol Mar* 47:477–482

*Editorial responsibility: Paul Snelgrove,
St. John's, Canada*

*Submitted: April 24, 2009; Accepted: October 28, 2009
Proofs received from author(s): February 17, 2010*