Environmental and Ecological Statistics **12,** 125–138, 2005

The apportionment of quadratic entropy: a useful alternative for partitioning diversity in ecological data

SANDRINE PAVOINE[1]* and SYLVAIN DOLÉDEC[2]

[1]*Laboratoire de biométrie et biologie évolutive, UMR CNRS 5558, Université Lyon I, 69622, Villeurbanne Cedex, France;*
E-mail: pavoine@biomserv.univ-lyon1.fr
[2]*Laboratoire d'écologie des hydrosystèmes fluviaux UMR CNRS 5023, Université Lyon I, 69622, Villeurbanne Cedex, France*

Received September 2003; Revised September 2004

Many methods that study the diversity within hierarchically structured populations have been developed in genetics. Among them, the analysis of molecular variance (AMOVA) (Excoffier *et al*., 1992) has the advantage of including evolutionary distances between individuals. AMOVA is a special case of a far more general statistical scheme produced by Rao (1982a; 1986) and called the apportionment of quadratic entropy (APQE). It links diversity and dissimilarity and allows the decomposition of diversity according to a given hierarchy. We apply this framework to ecological data showing that APQE may be very useful for studying diversity at various spatial scales. Moreover, the quadratic entropy has a critical advantage over usual diversity indices because it takes into account differences between species. Finally, the differences that can be incorporated in APQE may be either taxonomic or functional (biological traits), which may be of critical interest for ecologists.

Keywords: dissimilarity, functional diversity, macroinvertebrates, Rao's axiomatization, taxonomy

---

1. Introduction

Biodiversity means variability of life in all its form, levels and combination including genetic diversity, species diversity and ecosystem diversity (see e.g. Heywood and Watson, 1995). The study of biodiversity thus covers a wide range of disciplines. Within each discipline, several indices or statistical methods have been developed to measure biodiversity. This enormous quantity of specific biodiversity measurements

---

*Corresponding author.

requires a more general framework since diversity is a property common to any biological element whatever its scale and type. Such a framework was initiated by Rao (1986) who proposed a set of axioms listing the properties required for a given measure to be considered as a measure of diversity.

Several methods dealing with genetic diversity in subdivided populations have been developed (Excoffier, 2001). Every method decomposes genetic diversity into an average genetic diversity within demes and a genetic diversity among demes. If demes are themselves hierarchically structured, genetic diversity within demes can be decomposed as well. Such approaches are currently used for decomposing gene diversity. (Nei, 1973; Weir and Cockerham, 1984; Finkeldey, 1994), nucleotide diversity (Nei and Li, 1979; Nei and Tajima, 1981; Nei and Jin, 1989; Crease *et al.*, 1990; Lynch and Crease, 1990; Nei and Miller, 1990; Holsinger and Mason-Gamer, 1996), microsatellite diversity (Slatkin, 1995), and any kind of genetic diversity (Excoffier *et al.*, 1992).

Similar to genetic data, ecological data are hierarchically structured according to spatial and temporal scale (Frissell *et al.*, 1986; Kolasa, 1989) and taxonomy in case of community ecology. Whittaker (1960, 1972) defined important concepts stating that the total γ-diversity of a region includes two components: α-diversity which represents within-community diversity and β-diversity which characterizes the degree of change in species diversity along environmental gradients. In the traditional multiplicative approach (Whittaker, 1960, 1972), β-diversity is the ratio between total diversity (γ) and α-diversity. Recently many authors stated that the decomposition of the total diversity into within- and between-community diversity should be additive (e.g., Veech *et al.*, 2002; Ricotta, 2003), as formerly suggested by Allan (1975) and Pielou (1975) for example. Such an additive partition of diversity can be expanded to various levels of organization and has potential application to multiple scales (Lande, 1996). Moreover it provides commensurable measures of within- and between-community diversity (Wagner *et al.*, 2000; Veech *et al.*, 2002). Despite the bewildering range of diversity indices, ecologists lack methods that can simultaneously analyze all components of diversity by taking into account both the abundance of the species and the dissimilarities among species.

In this paper, we highlight that Rao's apportionment of quadratic entropy (APQE, Rao, 1982a) is a fundamental basis allowing the partition of diversity suited to any kind of data. Although this method has been used extensively in genetics under the name of AMOVA, it is rather new in Ecology. We illustrate the potential of APQE using an ecological data set including two scales and various types of biological information such as species composition and trait composition.

2. The apportionment of quadratic entropy

We will restrict our approach to hierarchically structured data (e.g., nested sampling design, taxonomic hierarchy). With this type of data, Rao (1986) provided two rules needed to characterize a measure of diversity (H). The first one is that H must be obviously nonnegative. The second one concerns concavity, which means that the diversity in two mixed sets (communities, regions or taxonomic levels, for example) must be higher than the average diversity within each set in order to avoid negative values for the components of diversity (see also Lande, 1996).

In this context, Rao (1982b, 1984) developed the apportionment of diversity (APDIV) as an appropriate method for partitioning diversity in hierarchically structured data.

Let us consider any N entities distributed among subsets nested into r groups. Group i contains s_i subsets. Each entity belongs to one out of n categories. Let and μ_{ij} be *a priori* probabilities associated with group i and the subset j of group i, respectively. These probabilities usually are the relative size (entity number) of each group or subset. Let $\mathbf{p}_{ij}$ be a vector that contains the frequencies of the categories in the subset j of group i. The frequencies of the categories in group i as a whole are given by the vector $\mathbf{p}_{i\bullet} = \sum_{j=1}^{s_i} \mu_{ij}\mathbf{p}_{ij}$; and their frequencies in all the groups mixed together are given by the vector $\mathbf{p}_{\bullet\bullet} = \sum_{i=1}^{r} \lambda_i \mathbf{p}_{i\bullet}$. Further consider H as a measure of diversity being nonnegative and concave. Then the APDIV is defined as

$$\begin{aligned} H(\mathbf{p}_{\bullet\bullet}) = & \sum_{i=1}^{r} \lambda_i \sum_{j=1}^{s_i} \mu_{ij} H(\mathbf{p}_{ij}) \\ & + \sum_{i=1}^{r} \lambda_i H(\mathbf{p}_{i\bullet}) - \sum_{i=1}^{r} \lambda_i \sum_{j=1}^{s_i} \mu_{ij} H(\mathbf{p}_{ij}) \\ & + H(\mathbf{p}_{\bullet\bullet}) - \sum_{i=1}^{r} \lambda_i H(\mathbf{p}_{i\bullet}). \end{aligned} \tag{1}$$

The index $H(\mathbf{p}_{\bullet\bullet})$ measures the total diversity within all the groups mixed together. In the right-hand side of equation (1), the first row represents the diversity within subsets and within groups, the second corresponds to the diversity among subsets but within groups, and the last one stands for the diversity among subsets and among groups.

Rao and colleagues (Rao and Nayak, 1985; Liu and Rao, 1995) applied this general approach to a particular diversity index called "quadratic entropy". This index was introduced by Rao (1982b) to link diversity measurements with dissimilarity coefficients. Let $\mathbf{D}$ be a $n \times n$ matrix containing the dissimilarities d_{kl} between any categories k and l ($1 \leq k \leq n$ and $1 \leq l \leq n$). Matrix $\mathbf{D}$ is symmetric with null values on the diagonal. The quadratic entropy is defined as

$$H_{\mathbf{D}}(\mathbf{p}) = \mathbf{p}^t \mathbf{D} \mathbf{p} = \sum_{k=1}^{n} \sum_{l=1}^{n} p_k p_l d_{kl}$$

where $\mathbf{p} = (p_1 \cdots p_k \cdots p_n)$ is a frequency vector, either p_{ij}, $p_{i\bullet}$ or $p_{\bullet\bullet}$ with the above notations.

As stated by Rao (1986), a diversity measure can be decomposed along a nested sampling provided that it is nonnegative and concave. H_{D} is always nonnegative because it only sums up frequencies and distances which are nonnegative. In order to assure its concavity, we only consider dissimilarity matrices $\mathbf{D}$ such that the matrix noted $\mathbf{D}^{\frac{1}{2}}$, which contains the square root of the values in $\mathbf{D}$, is Euclidean (Rao, 1984; Rao and Nayak, 1985; Champely and Chessel, 2002) that is to say n points M_k ($k = 1,2,\ldots,n$) can be embedded in a Euclidean space so that the Euclidean distance between M_k and M_l is $\sqrt{d_{kl}}$. (Gower and Legendre, 1986). Note that a nonEuclidean dissimilarity matrix can be transformed into a Euclidean dissimilarity matrix (Lingoes, 1971; Cailliez, 1983). The APDIV applied to the quadratic entropy is called apportionment of quadratic entropy (APQE). APQE generalizes other types of decomposition of diversity indices: the categorical analysis of variance (CATANOVA, Light and Margolin, 1971), which decomposes Gini–Simpson index, and the analysis of variance (ANOVA, Fisher, 1925), which partitions the variance of a quantitative variable. APQE is equal to CATANOVA when the dissimilarities among the categories are all equal to 1 ($\mathbf{D} = \mathbf{11}' - \mathbf{I}$, where $\mathbf{1}$ is a $n \times 1$ vector of units and $\mathbf{I}$ is the $n \times n$ identity matrix) (Nayak, 1986a). It is equal to the ANOVA when the dissimilarity between the entities k and l is equal to $(y_k - y_l)^2$, where y_k and y_l are the values taken by a quantitative variable for the entities k and l, respectively (Rao, 1984).

3. Use of APQE: from genetics to ecology

Fifteen years before Rao's axiomatization the quadratic entropy was introduced probably for the first time by two ecologists (Hendrickson and Ehrlich, 1971) to take into account differences between species in a diversity index. However, for the last 25 years, this index has been given success mostly in genetics. Two teams of geneticists have contributed to the use of quadratic entropy. First, Nei and collaborators (Nei and Li, 1979; Nei and Tajima, 1981) designed indices similar to quadratic entropy in order to measure nucleotide diversity. In that case, the entities are organisms and the categories represent particular DNA sequences. Second, Excoffier *et al.* (1992) developed a decomposition similar to the apportionment of the quadratic entropy widely used in genetics (see for example Bosch *et al.*, 1999; Olsen *et al.*, 2003; Lecis and Norris, 2004; Qiu *et al.*, 2004; Vences *et al.*, 2004). This decomposition, called analysis of molecular variance (AMOVA), treats data where individuals (entities) generally belonging to the same species are sampled from several populations (subsets). These populations may be grouped into clusters, which may be themselves grouped into larger clusters thus generating a hierarchical structure. Furthermore, each individual is characterized by one or two genetic traits (categories) and genetic dissimilarities among traits are computed.

This data scheme may be easily transposable to ecological data since the analysis of species diversity should take into account the dissimilarities between species, the abundance of species within communities and the hierarchical structure of communities.

Several ecologists have recently rediscovered Rao's work and have suggested applying quadratic entropy to their particular ecological data (Izsak and Papp, 1995; Watve and Gangal, 1996; Izsak and Papp, 2000; Shimatani, 2001; Champely and Chessel, 2002; Izsak and Szeidl, 2002). In that case, the diversity within a fauna or a community is under concern and differences between species are estimated in terms of taxonomy or trait. Izsak and Szeidl (2002) even showed that quadratic entropy may decrease with the number of species. Indeed in a community where species are different in terms of genetics or traits, if the additional species are very close to the others the mean of the between-species dissimilarities decreases. To our knowledge, only Champely and Chessel (2002) have applied the partition of Rao's quadratic entropy to ecological data, as formerly suggested by Woollcott Smith in his discussion following Izsak and Papp (1995) paper.

4. An ecological application in hydrobiology

To illustrate the potential of APQE for analyzing and decomposing diversity in ecological data, we selected a data set published in aquatic ecology by Ivol *et al.* (1997). These authors aimed at analyzing changes in macroinvertebrate assemblages along the course of a large river. Fluvial hydrosystems are in essence hierarchically organized from microhabitats to entire watersheds (Frissell *et al.*, 1986) and thus provide an adequate model for decomposing diversity. Furthermore, approaches at the community level presuppose the use of species lists to compare assemblages among various environmental situations. Such approaches thus imply to consider how species aggregate. Taxonomic aggregation has been mostly used (Corkum and Ciborowski, 1988), however functional aggregation according to biological traits such as size or reproduction (Statzner *et al.*, 1997) or feeding types (Vannote *et al.*, 1980) may enable larger scale comparisons and provide much more general information about ecosystem functioning.

Computations and graphical displays were done using the R statistical software (Ihaka and Gentleman, 1996). Programs and functions for computing AMOVA and quadratic entropy are available in the ade4 package of the R environment; and the APQE function is available by request to the first author of the paper.

4.1 *Data set*

A total of 38 sites were surveyed along 800 km of the Loire River yielding 40 species of Trichoptera and Coleoptera sampled from riffle habitats (see Ivol *et al.*, 1997 for further details on sampling). The river was divided into three regions according to geology: granitic highlands (Region#1), limestone lowlands (Region#2) and granitic lowlands (Region#3).

Two species traits were selected from existing databases (Usseglio-Polatera *et al.*, 2000; Statzner *et al.*, 2001; Gayraud *et al.*, 2003). These trait databases summarise the available European knowledge accumulated over the 20th century for all easily identifiable freshwater invertebrates of France. We selected maximal size, which usually indicates the ratio of production/biomass and of production/respiration in lotic invertebrate (Statzner, 1987) and has many implications for many other functions in the ecosystem. We also considered the functional feeding groups, which have been largely documented since the works of Cummins (1974) and represent cornerstone in the river continuum concept (Vannote *et al.*, 1980). In these databases, for each of the two traits, the affinity of each species to each trait category is quantified using a fuzzy coding approach, (Chevenet *et al.*, 1994). The maximal size achieved by the species is divided into five length categories ranging from ≤ 5 to >40 mm. Feeding habits comprise seven categories: engulfers, shredders, scrapers, deposit-feeders, active filter-feeders, passive filter-feeders and piercers. A score is assigned to each species for describing its affinity for a given trait category from "0" which indicates no affinity to "3" which indicates high affinity. These affinities are further transformed into percentage per trait per species. The percentage of affinity of the species k for the category m is noted q_{km}.

4.2 *Dissimilarity measurements*

We used four criteria to compute the dissimilarities among species: equidistance, body size, feeding habits and taxonomy. For the equidistance, the dissimilarity between two species was arbitrarily set to 1 meaning an equivalence between species. To compute dissimilarities from the fuzzy variables (body size and feeding habit), we selected the Manly's distance. (Manly, 1994, formula 5.8 p. 68) defined as

$$d_{kl} = 1 - \sum_{m=1}^{M} q_{km} q_{lm} \Bigg/ \sqrt{\left\{ \sum_{m=1}^{M} q_{km}^2 \sum_{m=1}^{M} q_{lm}^2 \right\}}$$

where d_{kl} is the dissimilarity between species k and l, M is the number of categories (five for the maximal-size trait and seven for the feeding-habit trait), and 9 km and q_{lm} are the percentages of affinities of species k and l for the category m of either the body size or the feeding habit depending upon which criteria is concerned. Quadratic entropy applied to dissimilarities taking into account species traits yields a measure of functional diversity (Petchey and Gaston, 2002). For computing taxonomic dissimilarities we used the index proposed independently by Izsak and Papp (1995) and Warwick and Clarke (1995): the dissimilarity equals 1 between two species of the same genus, 2 between two species of different genera belonging to the same family, 3 between two species of different families belonging to the same order and 4 between two species belonging to a different order.

4.3 *Decomposition of the quadratic entropy*

Since the value of total diversity depends on the type of biological variable we compared the decomposition of the total diversity in terms of percentage. Within-site diversity incorporated from 50 to about 65% of the total diversity (Table 1). Values were very stable across the indices. Only feeding groups demonstrated a lower diversity at this scale linked to a higher diversity among sites. The diversity among sites within regions contained from 24 to 36% of the total diversity. Finally, the diversity among regions ranged from 9 to 14% of the total diversity. For comparing values of diversity across spatial scales, we had to take into account the number of independent items (degree of freedom) at each scale. To test the within-scale differences several methods have been proposed. Nayak's process (1986a, b) tests if the organisms are independently divided into the sites according to a multinomial distribution. This distribution is assumed to be constant first across regions and second across sites within each region. Our data highly differs from both assumptions. Two interpretations are possible: either the sites and regions are indeed different, or the

Table 1. Decomposition of the total diversity according to (a) the frequencies of the species (Gini–Simpson index), the taxonomy, and (b) two biological traits.

Diversity Source	*d.f.*	*Gini–Simpson*	*Ratio*	*P*	*Taxonomy*	*Ratio*	*P*
(a)							
Among regions	2	0.087 (11%)	8.322	< 0.001	0.210 (9%)	5.797	0.022
Among sites/Within regions	35	0.183 (24%)	48.07	0.894	0.635 (28%)	59.84	0.277
Within sites	4628	0.504 (65%)			1.403 (63%)		
Total	4665	0.774			2.248		
	d.f.	Size	Ratio	*P*	Diet	Ratio	*P*
(b)							
Among regions	2	0.013 (9%)	6.558	0.022	0.020 (14%)	6.721	0.012
Among sites/Within regions	35	0.036 (26%)	52.53	0.545	0.051 (36%)	96.08	0.017
Within sites	4628	0.090 (65%)			0.070 (50%)		
Total	4665	0.139			0.141		

The degrees of freedom (d.f.) are indicated. Statistics and results of the permutation tests are given.

distribution of the organisms across the sites is not multinomial. Macroinvertebrates show patchy distributions according to levels of environmental disturbance (Levins and Paine, 1974). Those aggregations imply that, sampling an individual from a species increases the chance of observing other individuals from the same species. The multinomial assumption is thus here invalidated. Excoffier *et al.* (1992) suggested permutation tests to avoid distribution assumptions. They performed test on the differences between regions by permuting sites across regions. We choose this permutation scheme because it suits our data by taking into account the aggregation of individuals. For each permutation, we compute the ratio of the diversity between regions to the diversity between sites within regions. Excoffier *et al.* (1992) tested the differences between sites within regions by permuting the individuals across the sites within each region. This type of permutations does not take into account aggregation and thus could overestimate the real between-site differences. We choose to permute each species' abundance across the sites within each region. For example, for species k in region i, the abundance values n_{ijk} are permuted over sites $1 \leq j \leq s_i$. Once the permutations are done for all species, the ratio of the diversity between sites within regions to the diversity within sites is computed. For selected permutation scheme, the number of simulated values (out of 1000 samples) higher than the observed one is given in Table 1. For the between-region diversity from 12 to 22 simulated values exceeded the observed reference ratio. This suggests significant differences between regions in terms of frequency distributions, taxonomic and size compositions and diets. By contrast, the only significant differences between sites within regions are due to feeding habits.

According to Lande (1996), gamma diversity equals the weighted average alpha diversity plus beta diversity. Computing such values for each of our three regions (Table 2) showed that the three regions were approximately equally balanced according to the abundance of species (Gini–Simpson total diversity). Taxonomic, size, and diet diversities discriminated Region#1 situated upstream better than Gini–Simpson diversity due to higher regional species richness about twice that of the two other regions. Though differing in species richness, Region#2 and Region#3 had similar low diversity in size and diet, whereas similar to other indices Region#1 had the largest value. This result, which contradicts usual knowledge (e.g., Statzner, 1987), should be associated with the further downstream impact of human activities (Region#2 and Region#3) involving a reduction of environmental heterogeneity.

5. Discussion

Few methods for quantifying gamma diversity (i.e., diversity at the landscape level allowing comparison among regions) exist in ecology (Sweeney and Cook, 2001). The selected ecological example was intended to show how the APQE may help to partition diversity. In this case the APQE allowed the computation of two types of global diversity. Diversity at the scale of the entire stream or total diversity could be valuably compared to a similar value computed for some other stream. Regional gamma diversity allowed us the comparison of diversity among regions. As a result, the APQE is a useful tool for estimating biodiversity at a variety of spatial scales, a major issue in both basic and applied ecology (Vinson and Hawkins, 1998).

Table 2. Values of the diversity among sites within regions, the diversity within sites and the total diversity in each selected region.

	d.f.	*Gini–Simpson*	*Taxonomy*	*Size*	*Diet*	*Richness*
Region#1						
Among sites	15	0.25 (28%)	1.28 (33%)	0.08 (31%)	0.11 (44%)	24.00
Within sites	1579	0.63 (72%)	2.64 (67%)	0.18 (69%)	0.14 (56%)	7.00
Total	1594	0.88	3.92	0.26	0.25	31
Region#2						
Among sites	16	0.18 (32%)	0.30 (31%)	0.02 (33%)	0.02 (40%)	13.69
Within sites	2307	0.37 (68%)	0.66 (69%)	0.04 (67%)	0.03 (60%)	4.31
Total	2323	0.55	0.96	0.06	0.05	18
Region#3						
Among sites	4	0.05 (7%)	0.29 (22%)	0.01 (17%)	0.02 (29%)	6.60
Within sites	742	0.66 (93%)	1.05 (78%)	0.05 (83%)	0.05 (71%)	5.40
Total	746	0.71	1.34	0.06	0.07	12

Three diversity indices are used: the Gini–Simpson index, the quadratic entropy applied to three dissimilarity criteria (taxonomy, body size and feeding habits), and the richness. Richness decomposition was given by Lande (1996). The degrees of freedom (d.f.) are indicated.

In our study, diversity was significantly different between regions whereas diversity between sites within regions was significantly dissimilar only for diet composition. Differences in diet composition and resulting diversity are predicted along rivers by the river continuum concept (Vannote *et al.*, 1980). This may explain our results since our sites within regions are distributed from up- to downstream over large distances (> 40 km). The differences observed between regions are probably due to the environmental characteristics of each region such as altitude for example. This latter result is supported by Parsons *et al.* (2003) who have demonstrated a greater similarity in macroinvertebrate assemblages at the site (riffle) scale than at the catchment scale.

Veech *et al.* (2002) states that ecologists should use diversity partitioning as a conceptual framework and an analytical method to address questions pertaining to the relationship between local and regional species diversity. We think that Rao's axiomatization appears as a fundamental basis for analyzing patterns of diversity. Quadratic entropy has an advantage over species richness and Gini–Simpson index because it takes into consideration the dissimilarities among species.

Other recent studies are linked to APQE. Their goal was not to describe diversity but to test differences between groups of sites estimated through Bray–Curtis index (Bray and Curtis, 1957). Legendre and Anderson (1999), McArdle and Anderson (2001) and Anderson (2001) have tackled the question of factorial designs instead of nested designs. In fact, their analyses correspond to another part of Rao's axiomatization namely the analysis of diversity (ANODIV). In that case, up to now the constraint of orthogonal sampling is needed. Ways for performing permutation tests are still debated. Anderson (2001) did not restrict the partition of diversity to Euclidean matrix of dissimilarity, but stated that any matrix can be used. Since the primary focus of the author was to test differences between groups of sites, then the question of negative diversity was of minor importance. In this paper, we have restricted our analysis to Euclidean matrix (e.g., Rao and Nayak, 1985; Schneider *et al.*, 2000) to remove the non-interpretable result such as negative diversity between communities.

Cousins (1991) underlined the contradiction that traditional ecological indices treat species on an equal basis whereas species identification put to the fore differences between species. Shimatani (2001) also considered that species differences should be included in biodiversity indices to provide better ecological applications. Finally, Watve and Gangal (1996) noticed that "an information-based index would treat a community of four different biotypes of coliforms identical to another community consisting of one species of coliforms, one of actinomycetes, one of myxobacteria, and one of archaebacteria, whereas we feel that the latter should be treated as more diverse."

The use of APQE allows the introduction of phylogenetic or taxonomic distances between species, which helps to overcome the above drawbacks. In our example, we used a taxonomic distance computed according to the taxonomic tree. An alternative could be to decompose diversity according to the taxonomic level (e.g., genus, family). Furthermore, ecologists do need methods that simultaneously evaluate diversity at different scales (Ricklefs, 1987) and APQE may help to reach this objective.

As stated in the introduction, several methods for decomposing diversity were developed mainly in genetics and transferred to ecology. For example, Allan (1975) compared two methods developed in ecology by Pielou (1967) and Levins (1968) to one method designed in genetics by Lewontin (1972). Furthermore, Lande (1996) introduced in ecology the additive decomposition of the Gini–Simpson index proposed in genetics by Nei (1973). In this paper, we suggest that the apportionment of the quadratic entropy of Rao (1982a), also at work in AMOVA (Excoffier *et al.*, 1992) a method designed for genetics, can be efficient for ecological data. All these decompositions of diversity represent particular cases of Rao's apportionment of diversity.

Acknowledgments

Many thanks to the "BIOFUN" group who kindly provided some additional information on the trait data. The data set on the Loire River was kindly provided by J.M. Ivol.

References

Allan, J.D. (1975) Components of diversity. *Oecologia*, **18**, 359–67.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.

Bosch, E., Calafell, F., Santos, F.R., Perez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., and Bertranpetit, J. (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *American Journal of Human Genetics*, **65**, 1623–38.

Bray, J.R. and Curtis, J.T. (1957) An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, **27**, 325–49.

Cailliez, F. (1983) The analytical solution of the additive constant problem. *Psychometrika*, **48**, 305–10.

Champely, S. and Chessel, D. (2002) Measuring biological diversity using Euclidean metrics. *Environmental and Ecological Statistics*, **9**, 167–77.

Chevenet, F., Dolédec, S., and Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology*, **31**, 295–309.

Corkum, L.D. and Ciborowski, J.J.H. (1988) Use of alternative classifications in studying broad-scale distributional patterns of lotic invertebrates. *Journal of the North American Benthological Society*, **7**, 167–79.

Cousins, S.H. (1991) Species diversity measurement: Choosing the right index. *Trends in Ecology and Evolution*, **6**, 190–92.

Crease, T.J., Lynch, M., and Spitze, K. (1990) Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of *Daphnia pulex*. *Molecular Biology and Evolution*, **7**, 444–58.

Cummins, K.W. (1974) Structure and function of stream ecosystems. *BioScience*, **24**, 631–41.

Excoffier, L. (2001) Analysis of population subdivision in *Handbook of Statistical Genetics*, D.J. Balding, M. Bishop and C. Cannings (eds.), John Wiley & Sons, Ltd, New York, pp. 271–307.

Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–91.

Finkeldey, R. (1994) A simple derivation of the partitioning of genetic differentiation within subdivided populations. *Theoretical and Applied Genetics*, **89**, 198–200.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.

Frissell, C.A., Liss, W.J., Warren, C.E., and Hurley, M.D. (1986) A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management*, **10**, 199–214.

Gayraud, S., Statzner, B., Bady, P., Haybach, A., Schöll, F., Usseglio-Polatera, P., and Bachi, M. (2003) Invertebrate traits for the biomonitoring of European large rivers: An initial assessment of alternative metrics. *Freshwater Biology*, **48**, 2045–64.

Gower, J.C. and Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, **3**, 5–48.

Hendrickson, J.A.J. and Ehrlich, P.R. (1971) An expanded concept of "species diversity". *Notulae Naturae*, **439**, 1–6.

Heywood, V.H. and Watson, R.T. (1995) *Global biodiversity assessment*, Cambridge University Press, Cambridge.

Holsinger, K.E. and Mason-Gamer, R.J. (1996) Hierarchical analysis of nucleotide diversity in geographically structured populations. *Genetics*, **142**, 629–39.

Ihaka, R. and Gentleman, R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Ivol, J.M., Guinand, B., Richoux, P., and Tachet, H. (1997) Longitudinal changes in Trichoptera and Coleoptera assemblages and environmental conditions in the Loire River (France). *Archiv für Hydrobiologie*, **138**, 525–57.

Izsak, J. and Papp, L. (1995) Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. *Environmental and Ecological Statistics*, **2**, 213–24.

Izsak, J. and Papp, L. (2000) A link between ecological diversity indices and measures of biodiversity. *Ecological Modelling*, **130**, 151–56.

Izsak, J. and Szeidl, L. (2002) Quadratic diversity: Its maximization can reduce the richness of species. *Environmental and Ecological Statistics*, **9**, 423–30.

Kolasa, J. (1989) Ecological systems in hierarchical perspective: breaks in community structure and other consequences. *Ecology*, **70**, 36–47.

Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, **76**, 5–13.

Lecis, R. and Norris, K. (2004) Population genetic diversity of the endemic Sardinian newt *Euproctus platycephalus*: Implications for conservation. *Biological Conservation*, **119**, 263–70.

Legendre, P. and Anderson, M.J. (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.

Levins, R. (1968) *Evolution in changing environments. Monographs in population biology*, Princeton University Press, Princeton.

Levins, S.A. and Paine, R.T. (1974) Disturbance, patch formation, and community structure. *Proceedings of the National Academy of Sciences of the USA*, **71**, 2744–47.

Lewontin, R.C. (1972) The apportionment of human diversity. *Evolutionary Biology*, **6**, 381–98.

Light, R.J. and Margolin, B.H. (1971) An analysis of variance for categorical data. *Journal of the American Statistical Association*, **66**, 534–44.

Lingoes, J.C. (1971) Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika*, **36**, 195–203.

Liu, Z.J. and Rao, C.R. (1995) Asymptotic distribution of statistics based on quadratic entropy and bootstrapping. *Journal of Statistical Planning and Inference*, **43**, 1–18.

Lynch, M. and Crease, T.J. (1990) The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, **7**, 377–94.

Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*, 2 Chapman & Hall, London.

McArdle, B.H. and Anderson, M.J. (2001) Fitting multivariate models to community data: comment on distance-based redundancy analysis. *Ecology*, **82**, 290–97.

Nayak, T.K. (1986a) An analysis of diversity using Rao's quadratic entropy. *Sankhya: The Indian Journal of Statistics*, **48B**, 315–30.

Nayak, T.K. (1986b) Sampling distributions in analysis of diversity. *Sankhya: The Indian Journal of Statistics*, **48B**, 1–9.

Nei, M. (1973) Analysis of gene diversity in subdivised populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–23.

Nei, M. and Jin, L. (1989) Variances of the average numbers of nucleotide substitutions within and between populations. *Molecular Biology and Evolution*, **6**, 290–300.

Nei, M. and Li, W.-H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–73.

Nei, M. and Miller, J.C. (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics*, **125**, 873–9.

Nei, M. and Tajima, F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145–63.

Olsen, J.B., Miller, S.J., Spearman, W.J., and Wenburg, J.K. (2003) Patterns of intra- and inter-population genetic diversity in Alaska coho salmon: Implications for conservation. *Conservation Genetics*, **4**, 557–69.

Parsons, M., Thoms, M.C., and Norris, R.H. (2003) Scales of macroinvertebrate distribution in relation to the hierarchical organization of river systems. *Journal of the North American Benthological Society*, **22**, 105–22.

Petchey, O.L. and Gaston, K. (2002) Functional diversity (FD), species richness and community composition. *Ecology Letters*, **5**, 402–11.

Pielou, E.C. (1967) The use of information theory in the study of the diversity of biological populations. In *Proc. Fifth Berkeley Symposium on Math. Stat. and Prob.* pp. 163–77.

Pielou, E.C. (1975) *Ecological diversity*, Wiley & Sons, New York.

Qiu, Y.-X., Hong, D.-Y., Fu, C.-X., and Cameron, K.M. (2004) Genetic variation in the endangered and endemic species *Changium smyrnioides* (Apiaceae). *Biochemical Systematics and Ecology*, **32**, 583–96.

Rao, C.R. (1982a) Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, **21**, 24–43.

Rao, C.R. (1982b) Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhya: The Indian Journal of Statistics*, **A44**, 1–22.

Rao, C.R. (1984) Convexity properties of entropy functions and analysis of diversity. *Inequalities in statistics and probability*, **5**, 68–77.

Rao, C.R. (1986) Rao's axiomatization of diversity measures in *Encyclopedia of Statistical Sciences*, S. Kotz and N.L. Johnson (eds.), Wiley & Sons, New York, pp. 614–17.

Rao, C.R. and Nayak, T.K. (1985) Cross entropy, dissimilarity measures, and characterizations of Quadratic Entropy. *IEEE Transactions on Information Theory*, **IT-31**, 589–93.

Ricklefs, R.E. (1987) Community diversity: Relative roles of local and regional processes. *Science*, **235**, 167–71.

Ricotta, C. (2003) Additive partition of parametric information and its associated beta-diversity measure. *Acta Biotheoretica*, **51**, 91–100.

Schneider, S., Roessli, D., and Excoffier, L. (2000) *Arlequin Ver 2.000: A Software for Population Genetics Data Analysis*, Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Shimatani, K. (2001) On the measurement of species diversity incorporating species differences. *Oikos*, **93**, 135–47.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–62.

Statzner, B. (1987) Characteristics of lotic ecosystem and consequences for future research directions. *Ecological studies*, **61**, 365–90.

Statzner, B., Bis, B., Dolédec, S., and Usseglio-Polatera, P. (2001) Perspectives for biomonitoring at large spatial scales: a unified measure for the functional composition of invertebrate communities in European running waters. *Basic and Applied Ecology*, **2**, 73–85.

Statzner, B., Hoppenhaus, K., Arens, M.-F., and Richoux, P. (1997) Reproductive traits, habitat use and templet theory: A synthesis of world-wide data on aquatic insects. *Freshwater Biology*, **38**, 109–35.

Sweeney, B.A. and Cook, E. (2001) A landscape-level assessment of understory diversity in upland forests of North-Central Wisconsin, USA. *Landscape Ecology*, **16**, 55–69.

Usseglio-Polatera, P., Bournaud, M., Richoux, P., and Tachet, H. (2000) Biological and ecological traits of benthic freshwater macroinvertabrates: relationships and definition of groups with similar traits. *Freshwater Biology*, **43**, 175–205.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.A., and Cushing, C.E. (1980) The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–37.

Veech, J.A., Summerville, K.S., Crist, T.O., and Gering, J.C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, **99**, 3–9.

Vences, M., Chiari, Y., Raharivololoniaina, L., and Meyer, A. (2004) High mitochondrial diversity within and among populations of Malagasy poison frogs. *Molecular Phylogenetics and Evolution*, **30**, 295–307.

Vinson, M.R. and Hawkins, C.P. (1998) Biodiversity of stream insects: variation at local, basin, and regional scales. *Annual Review of Entomology*, **43**, 271–93.

Wagner, H.H., Wildi, O., and Ewald, K.C. (2000) Additive partitioning of plant species diversity in an agricultural mosaic landscape. *Landscape Ecology*, **15**, 219–27.

Warwick, R.M. and Clarke, K.R. (1995) New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Marine Ecology Progress Series*, **129**, 301–5.

Watve, M.G. and Gangal, R.M. (1996) Problems in measuring bacterial diversity and a possible solution. *Applied and Environmental Microbiology*, **62**, 4299–301.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–70.

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

Whittaker, R.H. (1972) Evolution and measurement of species diversity. *TAXON*, **21**, 213–51.

Biographical sketches

S. Pavoine is Ph.D. student at the "statistical ecology" team in the UMR CNRS 5558. Her research interest concerns the statistical methods for measuring diversity including test procedures and validation with the aim of providing a theoretical unifying point of view available to various almost separate disciplines such as genetics and community ecology.

S. Dolédec teaches biology, ecology and data analysis at the University of Lyon1. He belongs to a CNRS ecological research unit in the Laboratoire d'Ecologie des Hydrosystèmes Fluviaux of the same university. His research incorporates data analysis and current ecological theories to serve ecologically oriented management.