LETTER

Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight

Sandrine Pavoine, [1,2]* Milton S. Love[3] and Michael B. Bonsall[1,4]

[1]*Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[2]*Muséum National d'Histoire Naturelle, Département Ecologie et Gestion de la Biodiversité, UMR 7204 MNHN-CNRS-P6, 61 rue Buffon, 75005 Paris, France*
[3]*Marine Science Institute, University of California, Santa Barbara, CA 93160, USA*
[4]*St Peter's College, New Inn Hall Street, Oxford. OX1 2DL, UK*
*Correspondence:
E-mail: pavoine@mnhn.fr*

Abstract

Phylogenetic divergences have recently been included in analyses that aim to elucidate patterns of biodiversity in space and time. We introduce a generalized framework for two widely used phylogenetic diversity (PD) indices: Rao's quadratic entropy (QE) and Faith's PD. We demonstrate how this framework can be used to partition diversity simultaneously across evolutionary periods and spatial (e.g. local communities in a region) and/or time units (e.g. a community investigated yearly). From a study of rockfish hotspot diversity from the Southern California Bight, the analysis of PD revealed a recent decrease in the amount of fish caught from six evolutionary deep lineages, with implications for the community structure of this speciose group. This approach, which can also be applied to trees assembled from functional traits, contributes to our understanding of the mechanisms that underpin community organization and to the description of the consequences of human-driven impacts in the environment.

Keywords

Alpha diversity, beta diversity, biodiversity, community assembly, Faith's phylogenetic diversity, functional diversity, gamma diversity, phylogenetic trees, Rao's quadratic entropy, rockfish.

Ecology Letters (2009) **12**: 898–908

INTRODUCTION

Phylogenetic diversity (PD) represents how species are distributed within a phylogenetic tree. It has become the focus of both new research to understand the evolutionary history of communities (Webb *et al.* 2002) and new strategies for conservation (Faith 2008). Recently, Hardy & Senterre (2007) introduced the idea that a phylogenetic signal in patterns of biodiversity (phylogenetic clustering or overdispersion, e.g. Forest *et al.* 2007) can exist in different parts (or depths) of a phylogenetic tree. The importance of this evolutionary history can only be fully understood by analyses that partition the phylogeny into different lineages and associated evolutionary epochs or time periods. In particular, the PD between local communities (the so called phylo-β-diversity) can determine which lineages drive

temporal and/or spatial patterns within a region (Graham & Fine 2008). By knowing properties about these species lineages, patterns can be interpreted in terms of ecological mechanisms (such as dispersal ability, habitat filtering, competition, phylogenetic niche conservation) that affect the evolutionary organization of species communities. Similarly, by measuring a species community over several years, measures of PD can highlight which species lineages are affected by changes in their local environment.

Many of the past applications of PD measures in ecology have used taxonomic relationships between species as surrogates for evolutionary relationships (Crozier *et al.* 2005). For instance, Shimatani (2001) demonstrated that the taxonomic diversity measured by the quadratic entropy (QE) (Rao 1982) can be divided into Simpson's (1949) indices across the levels of the taxonomic hierarchy.

Taxonomic diversity can also be divided across distinct local communities (Pavoine & Dolédec 2005): the total taxonomic diversity in a region (γ) is equal to the average taxonomic diversity within the local communities ($\bar{\alpha}$) plus the taxonomic diversity between the local communities (β). This link between the decomposition over taxonomic levels and the decomposition over local communities was first demonstrated by Allan (1975). More recently, Ricotta (2005) unified these two approaches by demonstrating that each component of diversity (regional γ, within-communities $\bar{\alpha}$ and between-communities β) can be divided across taxonomic levels. Here, we show how the approach developed by Ricotta (2005) can be usefully extended to phylogenies. We demonstrate that such a partitioning can be extended to any concave index of species diversity, i.e. those indices for which diversity increases by mixing. In particular, we provide a unified framework for two of the most used indices of PD: Rao's (1982) QE and Faith's (1992) PD index. We illustrate how PD can be partitioned between evolutionary periods and between communities defined in terms of spatial (to elucidate community assembly) and time units (to describe impacts of human-driven changes in environments). We applied this methodology to determine how changes through time affect a rockfish (*Sebastes* spp.) assemblage subjected to high fishing pressure.

MATERIAL AND METHODS

A unified framework for phylogenetic diversity

To develop a unified framework for the different indices of PD we build on existing ways of measuring diversity. One of the indices proposed to measure the diversity within a community by regrouping individuals into categories (usually species) and considering their abundances and number is the Havrda & Charvat (1967) index (H_a) (see, e.g. Ricotta & Szeidl 2006):

$$H_a(\mathbf{p}) = \left(1 - \sum_{i=1}^{n} p_i^a\right) \Big/ (a - 1). \tag{1}$$

where $\mathbf{p} = (p_1,...,p_i,...p_n)$, p_i is the relative abundance of the *i*th category in a community of n categories, and a is a scaling constant ($a \geq 0$) that weights the importance of rarity (see below). If a is equal to zero, H_a reduces to the number of categories minus 1 (thereafter referred to as richness). Hence categories that have low abundances in the community are given equal weights with this diversity index as the categories with high abundances. As a increases, the weights of the rarest categories in the community decrease. When a tends to 1, then H_a tends to the Shannon (1948) index. When $a = 2$, then H_a is equal to the Simpson (1949) index. The three most common indices of diversity – the

richness, the Simpson index and the Shannon index – are thus particular cases of this more general scheme (eqn 1), and differ by the weight they give to rarity (Patil & Taillie 1982). We propose to adapt this index to measure PD.

To partition a phylogeny along a hierarchy, first the phylogeny must be ultrametric such that all tips are at equal distance from the root node and we must divide the phylogeny into periods of evolutionary time from the root node to the tips of the tree. The length of the periods can be different but a single period cannot include more than one speciation event (defined as an interior node). Consequently, a key definition for the periods is obtained by fixing the limits between two periods at each speciation event. Periods are numbered from the tips to the root. Let $(t_1, t_2, ..., t_N)$ ($0 \leq t_1 \leq t_2 \leq ... \leq t_N$) be the end time of the evolutionary periods (we fix $t_0 = 0$). We propose a new index of PD that can be adjusted for rarity (I_a) as:

$$I_a = \sum_{K=1}^{N} (t_K - t_{K-1})H_{a,K}. \tag{2}$$

where $H_{a,K}$ is the diversity index H_a (eqn 1) applied to the Kth period. With $a = 0$, H_0 is the richness and I_0 is Faith's (1992) PD index minus the height of the tree. With $a = 2$, H_2 is the Simpson index and I_2 is Rao's QE applied to phylogenetic distances between species. In that latter case, the partitioning is identical to the one proposed by Shimatani (2001, theorem 3, appendix 6) for taxonomies. When QE is applied to measure PD, the phylogenetic distance between two species *i* and *j* is equal to the sum of the branch lengths along the smallest path that connects species *i* and species *j* to their first common ancestor in an ultrametric tree (Pavoine *et al.* 2005; Pavoine & Bonsall 2009). When *a* tends to 1, I_a is thus a generalization of the Shannon index (H_1) to account for evolutionary history. This generalization differs from the one developed by Ricotta & Szeidl (2006). It has the advantage, over Ricotta & Szeidl's approach, in that it provides a direct, unified measure for two of the most used indices of PD, Faith's PD and Rao's QE (see Pavoine *et al.* 2005; Pavoine & Bonsall 2009).

$H_{a,K}$ is computed for each period, from the number of lineages that descend from the period and from the relative abundances summed within these lineages within the focal community. The branches that lie in the Kth period (from t_{K-1} to t_K) are roots for subtrees that emerged from them in the phylogenetic tree. The tips of each subtree correspond to a lineage that descended from a single ancestor which existed in the Kth period. That way, the Kth period divides extant species into lineages. The contribution of the period to the PD index (I_a) is the product of the length of the period ($t_k - t_{K-1}$) and the ecological diversity index applied to the relative abundances of the lineages of extant species descending from this period ($I_{a,K}$). Accordingly, this

hierarchical partitioning of PD is affected by phylogenetic tree shapes, the length of the periods and the species abundance distributions.

To the analysis of I_a within a community ($\bar{C}$), we can add the analysis of I_a within and between subcommunities (C_i). To do this is, if $I_{a,K} = (t_K - t_{K-1})H_{a,K}$ and $H_{a,K}$ is a concave index of diversity (such that $H_{a,K}$ increases by mixing communities together) then a community ($\bar{C}$) that is a mixture of subcommunities (C_i) (with $w_1, ..., w_r$ as the mixing proportions) can be partitioned over the different levels of diversity.

Subcommunities can be defined in space (e.g. local subcommunities in a region) or through time (e.g. a subcommunity corresponds to the state of a focal community at a given year) and the mixing proportions allow one to choose how to weight subcommunities. For example, weights can be set equal or in terms of sample sizes or the actual subcommunity sizes (Hardy & Jost 2008). The partitioning of I_a over evolutionary periods can then be applied to the averaged community (γ component $= \sum_K I_{a,K}(\bar{C})$), to the subcommunities ($\bar{\alpha}$ component $= \sum_K \sum_i w_i I_{a,K}(C_i)$), and to the PD between subcommunities ($\beta = \gamma - \bar{\alpha} = \sum_K \left[I_{a,K}(\bar{C}) - \sum_i w_i I_{a,K}(C_i) \right]$).

Permutation test

To determine which periods of the phylogeny are affected by differences between subcommunities we develop a Monte Carlo test. The tips of the whole phylogenetic tree (Fig. 1) are randomized so that the connection between species abundance and the phylogenetic tree is permuted (see Hardy 2008 model 1p). The permutation scheme assumes that there is no correlation between the pattern of species abundances in communities and the degree of phylogenetic relatedness between species. For each permutation, we calculated at each evolutionary period K, $(\beta/\gamma)_K = (1 - \alpha/\gamma)_K = 1 - \left[\sum_i w_i I_{a,K}(C_i) / I_{a,K}(\bar{C}) \right]$, which is equivalent to $1 - \left[\sum_i w_i H_{a,K}(C_i) / H_{a,K}(\bar{C}) \right]$, as a statistic of phylogenetic differentiations among subcommunities (Nayak 1983; Hardy & Senterre 2007). A higher value of $(\beta/\gamma)_K$ than expected from the species pool indicates phylogenetic clustering (high β, low α) at period K, which means that the subcommunities have fewer lineages that originated in period K than expected and/or only one or a few of these lineages dominate in abundance. On the contrary, a lower value of $(\beta/\gamma)_K$ indicates phylogenetic overdispersion (low β, high α) at period K, which means that the subcommunities have more lineages that originated in period K than expected and/or the abundance of these lineages are even. Accordingly, a P-value is associated to any period K as the number of permutations of the tips where $(\beta/\gamma)_K$ is higher (for phylogenetic clustering) or alternatively lower (if phylogenetic overdispersion is tested) than

the observed $(\beta/\gamma)_K$ value. A probability of error is associated to each test (= nominal α error). To avoid the chance occurrence of an erroneously significant test, corrections for multiple tests must be applied. Accordingly, we propose to first order the evolutionary periods by their value of $(\beta/\gamma)_K$ from the largest to the lowest if phylogenetic clustering is tested and from the lowest to the largest if phylogenetic overdispersion is tested. Tests are then applied on the ordered periods until the first non-significant test (P-value higher than the nominal α). Then Holm's (1979) correction for multiple non-independent tests can be applied. R scripts and a manual for the implementation of our approach are available in Appendixes S1 and S2.

Case study: hotspot of rockfish diversity

Sebastes species are rockfishes with a high level of diversity especially in terms of morphometry, colours and life history traits. Their range is mostly confined to the North-east Pacific, with a hotspot in the Southern California Bight, but some species occur in Atlantic and South Pacific (Love *et al.* 2002). They have been subjected to intense fishing in recent decades leading to constant declines in communities (Love *et al.* 1998). To determine if the decrease in the number of individuals and species in the Southern California Bight was accompanied by changes in the phylogenetic structure of the community we applied our partitioning of diversity across evolutionary periods, and within the whole period of study (1980–2007; the γ*component), within each year ($\bar{\alpha}*$) and between years ($\beta* = \gamma*-\alpha*$). We use asterisks to designate the component of temporal diversity as the original notations, $\bar{\alpha}$, β, γ, were introduced for biodiversity patterns in space (Whittaker 1972). The abundance data were obtained from the Marine Recreational Fishery Statistics Survey (MRFSS) (Fig. 1). We considered the compositions of rockfish assemblages caught by party and charter boats with hooks and lines from 1980–1986, 1993–1994, 1996, 1998–2007. Years 1987–1989 were not considered because a reduction in the funding of the program might have biased the results. The program was suspended from 1990 to 1992. We discarded years 1995 and 1997 because the numbers of survey hours were low (< 1000 h) (see Appendix S3 for details). Finally, we discarded trips without any rockfish in the catch to avoid counting vessels that shifted away from fishing from rockfishes (see Love *et al.* 1998). The phylogenetic tree was obtained from Hyde & Vetter (2007) (Fig. 1). A total of 52 species were found in the catches.

To analyse changes in species composition across years, we first performed a principal component analysis (PCA) on several parameters of diversity per year (the total catch per unit effort, CPUE values over all species, i.e. the number of fish caught per 1000 fishing hours, the number of species,
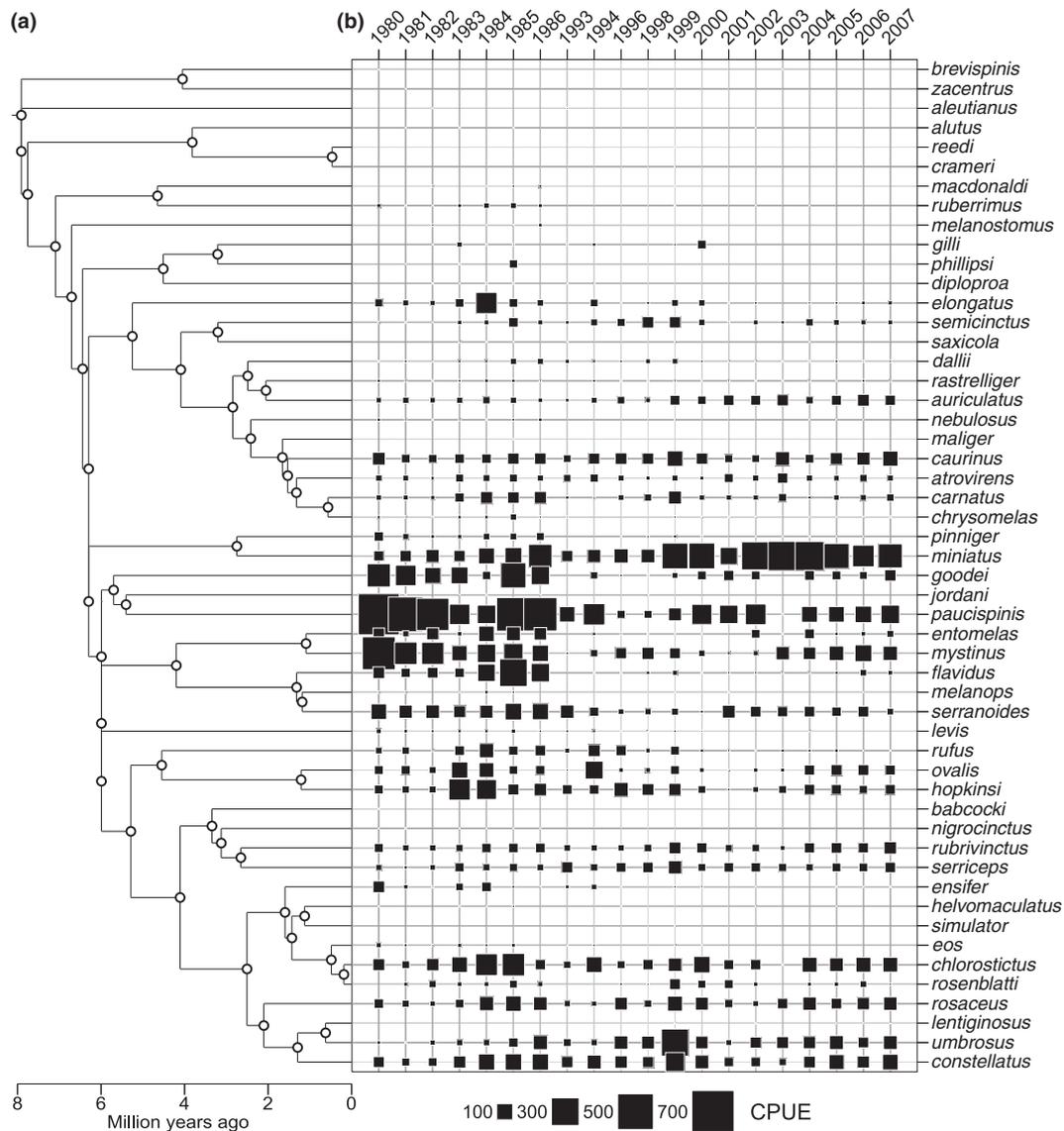
(a)

(b)

Figure 1 Rockfish data set. The phylogenetic tree of the 52 *Sebastes* species included in the analysis is provided. Absolute abundances are measured by CPUE values (number of fish caught per 1000 fishing hours) and given by the size of black squares. The abundances of some species, especially the most basal species in the phylogenetic tree, were very low in the catches (< 1 fish per 1000 fishing hours). The raw data set is provided in Appendix S4 and S5.

the Shannon index, the Simpson index, I_0, I_1, and I_2) and two complementary variables, the mean size of the species (Love *et al.* 2002), and the mean value of the vulnerability index (Cheung *et al.* 2005; Froese & Pauly 2008), both weighted by species' CPUE.

We used our Monte Carlo procedure with Holm's correction and a nominal α error of 5% to test at which periods the phylogenetic structure of the community changed between 1980 and 2007 (phylogenetic clustering hypothesis: lower $\bar{\alpha} * / \gamma *$ and higher $\beta */\gamma *$ than expected). We used the total CPUE values to weight the years:

$w_i = \sum_j CPUE_{ij} / \sum_{ij} CPUE_{ij}$, where $CPUE_{ij}$ is the CPUE value for species j during the year i.

As a comparison, we applied Hardy & Senterre (2007) partial permutation tests within the lineages. We demonstrate that there is no correlation between the pattern of species abundances amongst communities and the degree of phylogenetic relatedness between species (Appendix S2, see Hardy 2008).

Hardy & Senterre (2007) only used the QE (I_a with $a = 2$) metric to estimate the total statistic that corresponds to $(\beta * /\gamma *)_1$ for the first period (where each species corre-

spond to a lineage). They considered partial permutations within each lineage, with the null hypothesis that β^*/γ^* is not structured by the phylogeny within the lineage, and the alternative hypothesis that the phylogenetic differences between the years in this lineage are higher than expected by chance. To compare our methodology to Hardy & Senterre (2007), we used I_0 (based on Faith's PD), I_1 (based on the Shannon index), in addition to I_2. Hardy & Senterre (2007) did not treat multiple tests in their procedure and fixed the nominal α at 0.05. We limited the tests to the lineages that contained at least six species and we fixed the nominal α at 0.01. Even with these restrictions, the number of tests was high (see Appendix S2), which increased the risks of erroneously significant tests.

Finally, to describe the differences in the phylogenetic composition of the catches over years, we applied a double principal coordinate analysis (DPCoA, Pavoine *et al.* 2004) to the CPUE values, the presence/absence of species and the phylogenetic distances between species.

RESULTS

The overall values of I_a (sums over all periods), for $a = 0, 1, 2$, and their partition within and among years are given in Table 1.

Diversity within years

The biodiversity of rockfish assemblages caught by party and charter boats changed from 1980 to 2007 in Southern California Bight (Fig. 2). The catches in years 1980 to 1982 are characterized by high richness, PD, and CPUE, but also large-bodied species and high average vulnerability of the species. Years 1983 to 1985 had the most diverse catches, with a peak in 1984, both in terms of richness (species richness, phylogenetic richness (I_0) and number of individuals) and evenness (Simpson, Shannon, I_1 and I_2 indices). The catches in 1999 were also rich and diverse, whereas catches 1996 and 1998 presented similar abundance evenness between species but were less rich. Catches in years 1993–1994 and from 2000 were less rich and less even. Considering all the diversity indices investigated,

Table 1 Overall values of I_a (sums over all periods, $a = 0, 1, 2$) and its partition within and between years (percentages indicate the proportion of total diversity attributable to the components of diversity within and between years)

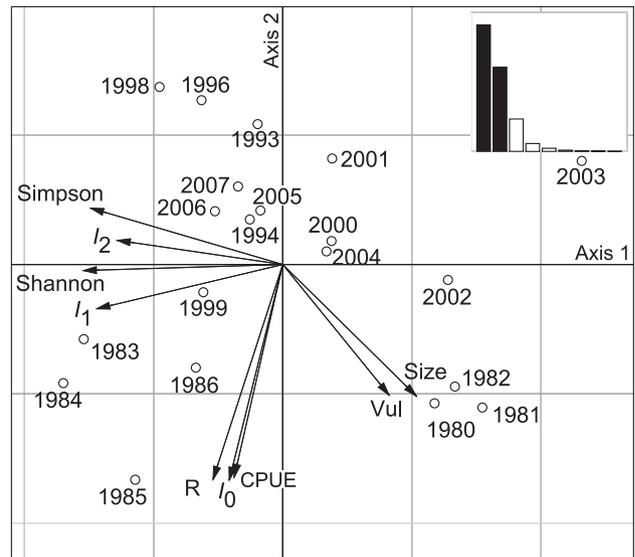|	I_0	I_1	I_2
Total diversity	180.15	20.74	5.38
Within-year diversity	114.36 (63%)	18.87 (91%)	5.13 (95%)
Between-year diversity	65.79 (37%)	1.87 (9%)	0.26 (5%)

Figure 2 First factorial map of the PCA applied to diversity indices: the total CPUE over all species, the species richness (noted R), the Shannon index, the Simpson index, I_0 based on Faith's (1992) PD index, the Shannon-like index of phylogenetic diversity I_1, and I_2 i.e. the quadratic entropy index applied to phylogenetic diversity) and two complementary variables (size = the mean size of the species, and vul = the mean value of the vulnerability index (Cheung *et al.* 2005), both weighted by species' CPUE). The box on the top right-hand corner provides the screeplot (amount of variance per axis).

2003 was the least diverse year, but paradoxically catches in 2003 contained more vulnerable species, than other years.

Diversity between years

The partitions of I_0, I_1 and I_2 between phylogenetic periods are detailed in Fig. 3. The whole contributions of the evolutionary periods depended both on the length of the period (evolutionary history, Fig. 3a) and the ecological index (either the richness, the Simpson index or the Shannon index, Fig. 3b,c,d). By definition, the number of lineages decreases from tips to the root node. Consequently, the ecological contributions of the periods ($H_{\alpha,K}$ in equation 5) decreased from the tips to the root if I_0 was used, and to a lesser extent if I_1 and I_2 were used (Fig. 3b,c,d). The most basal species had higher contributions to the diversity between years when the phylogenetic richness I_0, or the PD based on the Shannon index (I_1) were used (Fig. 3f,g). With the Simpson-like index (I_2), the main differences between years increased from period #1 to periods #41 and #42 and then drastically decreased (Fig. 3e).

The phylogenetic origin of the differentiation between years (high β^*/γ^*, phylogenetic clustering) was significant

Figure 3 Hierarchical partitioning of phylogenetic diversity across years and evolutionary periods: application to the rockfish data set. Periods are ordered from tips to root. (a) provides the length of the periods in million years. (b) gives the decomposition of Simpson diversity into α^* (white bars), β^* (black bars) components and across evolutionary periods (the height of the whole white and black bar is γ^*). (c) and (d) are equivalent bar plots for the Shannon (c) and richness (d) indices. In (e), the difference between the years (measured by β^*/γ^* with the Simpson index) is partitioned between the evolutionary periods. Closed circles highlight significant tests. (f) and (g) are equivalent graphs for the Shannon (f) and richness (g) indices.
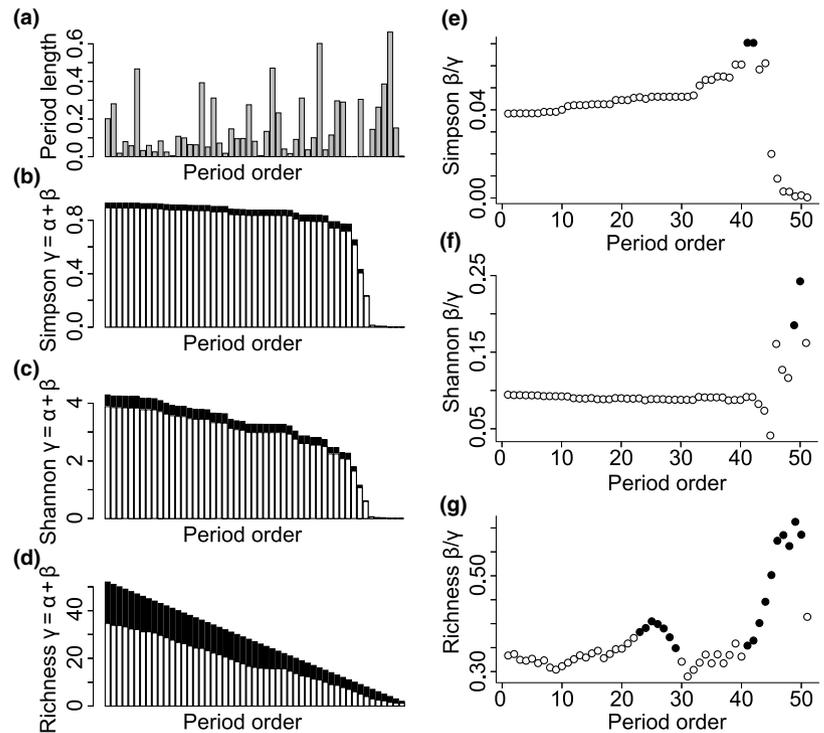
from periods 23 to 29 and from periods 41 to 50 with I_0 (Faith PD based on the richness); at the 49th and 50th periods with I_1 (based on the Shannon index) and at periods 41 and 42 with I_2 (phylogenetic QE based on the Simpson index) (Fig. 4 and Appendix S3). This means that the abundance (presence for I_0) of the lineages that descended from these highlighted significant periods changed over years.

Considering I_0, the significant periods mainly reflect lineages (often composed by one or a few species, on the negative part of the axis on Fig. 5b–d) that were present in the catches in the 1980s and in 1999 and had not been caught (or very occasionally) the other years. For I_1, the significant periods separate the following lineages (*S. brevispinis* with *S. zacentrus*), (*S. aleutianus* alone) and (*S. alutus*, *S. crameri* and *S. reedi*), whose abundances were low and variable across years without any trends in time (Fig. 4). For I_2, the significant periods 41 and 42 separate species into 11 and 12 lineages, respectively (Fig. 4). Fig. 5a demonstrates that the main differences between years were focused around lineages 7–12. Closed points in the middle of the graph represent species from lineages 1 to 6 that hardly contribute to differences between years mainly because they have low abundances. Lineages 9, 10 and 11 (species from *S. goodei* to *S. levis*) characterize the catches of the 1980s. Lineages 7 and 12 had the highest increase in 1984 and 1999. Finally, the increase in the catches of *S. miniatus* (lineage 8) characterizes the 2000s.

With Hardy's partial permutation tests, changes between years corresponded to phylogenetic clustering in lineages rooted on node #50 and the root node (all species) with I_0, in lineages rooted on any node between node #43 and the root node with I_1; and in the lineage rooted on node #43 (species from *S. goodei* to *S. constellatus*) with I_2 (Fig. 4).

DISCUSSION

Here, we have illustrated how PD can be partitioned both across periods of evolution and subcommunities (representing units in time or space) to explore evolutionary and ecological processes. We have provided a unified framework that includes Rao's QE applied to phylogenetic distances between species and Faith's index of phylogenetic richness. Within this framework, PD is revealed to be a function both of evolutionary period lengths (evolution) and an index of species diversity (ecology). We demonstrated that the decomposition of diversity into evolutionary periods and ecology can be applied to the three main scales of diversity studies: regional (γ), within communities (α) and between communities (β). We also emphasized that this metric can be usefully applied across different time scales – a long time period (e.g. decades) γ^*, a short time period (e.g. years) α^*, and the turnover between short time periods (e.g. between years) β^*. Here we discuss our results with reference to recent developments in using phylogenies to understand community structure and organization, and to the environ-
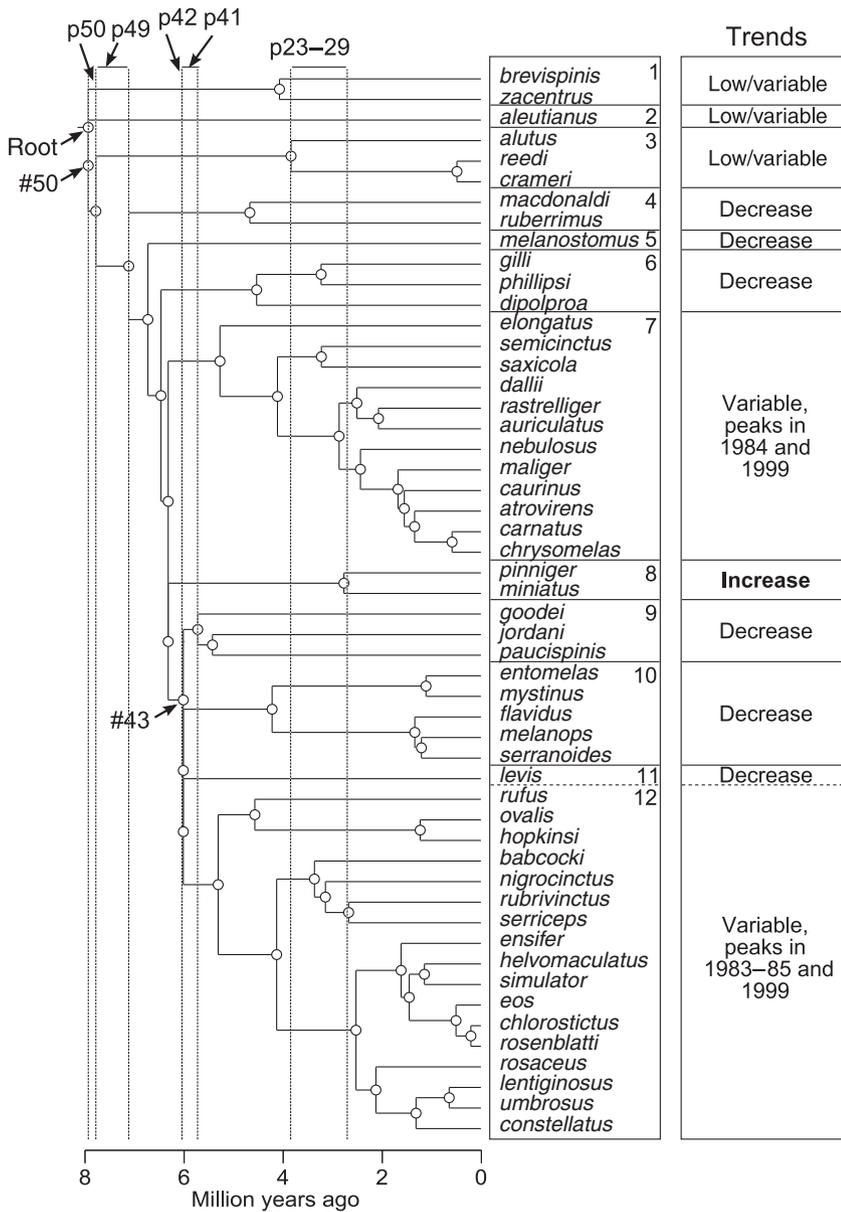
Figure 4 Phylogenetic changes between years: Significant phylogenetic structuring in the quotient β*/γ* are in periods 23 to 29 (p23–29) and 41 to 50 (from p41 to p50) according to I_0, 49 and 50 according to I_1 (p49 and p50) and 41 and 42 according to I_2 (p41 and p42). The periods 41 and 42 divide species into lineages as indicated by boxes. Differences are as follows. The period 41 separates *Sebastes levis* into a single lineage, whereas the period 42 includes *S. levis* with species from *S. ensifer* to *S. hopkinsi*. Trends for the 12 lineages corresponding to period 41 are given. According to Hardy & Senterre's partial tests, phylogenetic clustering (phylogenetic diversity within years lower than phylogenetic diversity between years) is significant in the lineages rooted on node 50 (all species except *S. brevispinis* and *S. zacentrus*) and the root node (all species) according to I_0; in the lineages rooted on any node between #43 and the root node according to I_1 and in the lineage rooted at node #43 (species from *S. goodei* to *S. constellatus*) according to I_2.

mental and human-driven impacts on the temporal structure of rockfish assemblages.

Case study

Rockfish have been subjected to intense fishing pressure in recent decades and were affected by a period of warming ocean temperature (Harvey *et al.* 2006). We effectively found a decrease in all diversity indices since 1980. In particular, all large-bodied species were shown to decline (except *S. miniatus*) which is corroborated by the widespread decrease in fish body size observed along the U.S. Pacific coast (Harvey *et al.* 2006). As such, substantial changes to

the fishing regulations were applied in 2000–2001, were increased 2003, and relaxed in 2004. Although these fishing regulations might have been expected to affect the CPUE values in 2003, our results demonstrate that the changes in the amount and composition of fish caught since the beginning of the conservation measure are far less important than the changes that occurred since 1980. Consequently, recent changes in fishing practice did not affect our main conclusions.

In our study of phylo-β-diversity, the use of the richness, Simpson or Shannon index in our PD framework (eqn 2) revealed that different processes affect the rockfish assemblage. The significant periods separate species into lineages
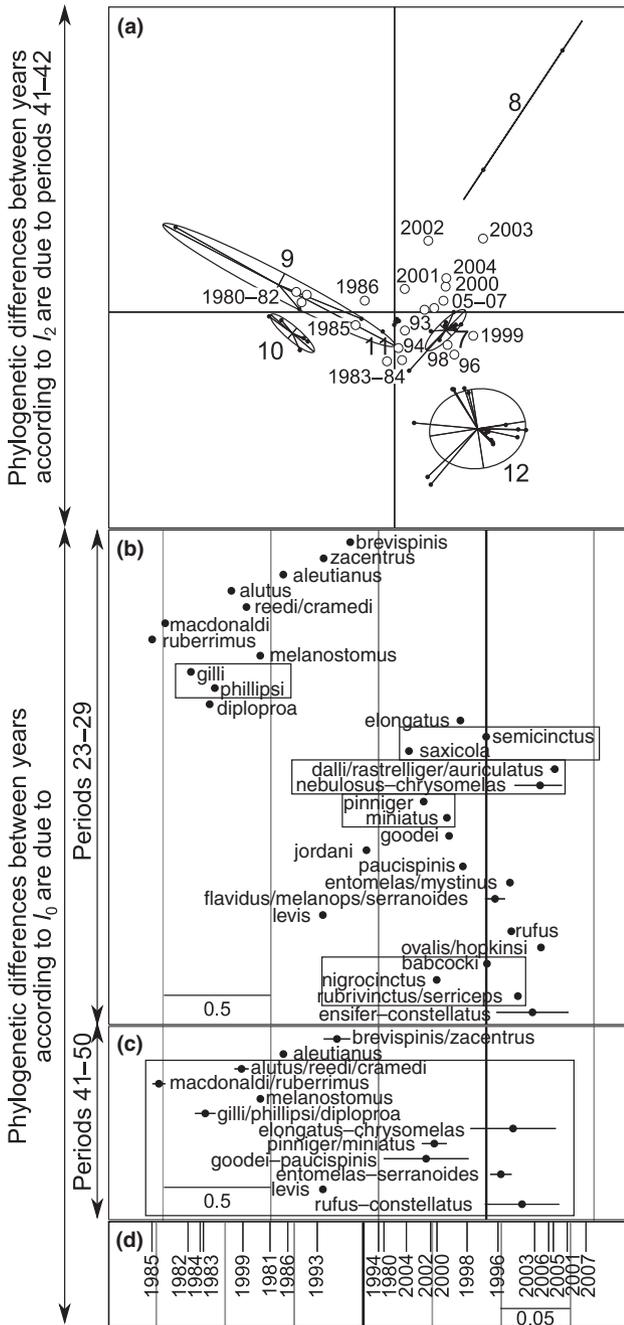
Figure 5 Application of the double principal coordinate analysis (DPCoA) to the rockfish data set. The DPCoA provides a graphical description of the phylogenetic differences between years (i.e. diversity β^*) calculated from abundances (I_2, panel a) and presence/absence (as in I_0, panels b–d). The relative positions of the years on the graphs (open circles on panel a, dashes on panel d) reflect their differences in terms of phylogenetic composition. The positions of the species (closed circles on panels a–c) help to explain differences between years. In panel a, the first (horizontal) axis and the second (vertical) axis expresses 61% and 20% of β^* as measured by I_2. The lineages 7–12 corresponding to periods 41 and 42 (Fig. 4) are indicated through ellipses and lines that connect the species to the centre of the lineage. The other lineages defined by periods 41 and 42 are clustered close to the centre of the graph, which indicates low impact on differences between years. Panels (b)–(d) detail the first axis of the DPCoA where the β^* was measured with presence/absence data. This first axis express the highest amount (35%) of β^*. Panel (b) provides the coordinates of the lineages defined by periods 23 to 29 and (c) those defined by periods 41–50 (significant periods, see Fig. 4). Labels written as 'species*i*-species*j*' indicate the lineage from species *i* to species *j* on Fig. 4 while 'species*i*/species*j*' denotes the lineages constituted by species *i* and species *j*. The rectangles highlight how lineages are nested from period 23–29 (panel b) and from period 41 to 50 (panel c). Panel (d) gives the coordinates of the years.

restricted to northern parts of the rockfish range (around Alaska or from British Columbia to northern-central California) (species from *S. brevispinis* to *S. ruberrimus*) or because they are occasionally or even rarely taken in the recreational fishery (species from *S. melanostomus* to *S. diploproa*, Love *et al.* 2002). Second, the other part of the phylogenetic signal in β^*/γ^* (periods 23 to 29 on Fig. 4) was due to species that occurred during all or almost all years and were clustered in the same lineages. Finally, the species that were caught in the 1980s but not recently (mainly *S. levis* and *S. jordani*) did not show strong phylogenetic signal. Although significant phylogenetic differences between years were highlighted by the analysis of I_0, these differences corresponded to decline in the catches from 1980 to 2007 only for species from *S. macdonaldi* to *S. diploproa* (Figs 4 and 5b,c).

When $a = 1$ (I_1 = phylogenetic version of the Shannon index), the significant periods 49 and 50 show that the pattern of phylo-β-diversity was due to the most basal species (*S. brevispinis*, *S. zacentrus* and *S. aleutianus*). As highlighted above, these species are more likely found from the Gulf of Alaska to central California, northern California and central Oregon, respectively (Love *et al.* 2002). The presence of few individuals in Southern California may be maintained by low dispersal. Accordingly, we were unable to find, with I_1, any continuous phylogenetic change in species composition from 1980 to 2007.

The impact of fishing on the assemblage from 1980 to 2007 was revealed by the most abundant species through the

whose first common ancestors originated during these periods. The changes in abundance of these lineages and consequently their importance in differentiating the years have been obtained from Fig. 1 and Fig. 5, which led to the following conclusions. With $a = 0$ (I_0 based on Faith PD index) part of the phylogenetic signal in β^*/γ^* (periods 41 to 51 on Fig. 4) was due to the effects associated with basal (close-to-root) species (from *S. brevispinis* to *S. diploproa* on Fig. 4) that occurred temporarily in the catches with very low CPUE value. These species are rare either because they

use of the index I_a when $a = 2$ (i.e. the phylogenetic QE index). The decomposition of the corresponding phylo-β-diversity index (β^*/γ^*) revealed important, continuous shifts in the composition of anglers' catches over years. Half of the lineages defined by periods 41 and 42 on Fig. 4 declined (either continuously or abruptly) in the catches after the 1980s (lineages 4–6 and 9–11 on Fig. 4). Every species in these lineages decreased in abundance. They covered a broad variety of life styles (epibenthic and benthic fishes, solitary and schooling forms, large-bodied and short bodied fishes, midwater feeders and benthic feeders). In particular, these lineages include rare and endangered species among them. *S. paucispinis* has been classified as critically endangered in the IUCN redlist since 1996 (IUCN 2008). The PD (as measured by Faith's index) within these lineages (15 species) amounts to 47% of the total PD of the whole tree (52 species).

Given that these species represent deep lineages in the phylogenetic tree and show declining abundances, changes in the phylogenetic composition of the assemblages are likely to lead to the loss of substantial evolutionary history and diversity. The time for recovery for the species that are declining might be longer than previously believed (Hutchings 2000). Apart from general trends of decrease for six lineages, species from two other lineages (numbered 7 and 12 on Fig. 4) were characterized by high peaks of abundance in the period from 1983 to 1985 (especially in 1984), and in 1999. Years 1984 and 1999 (La Niña conditions) were strong year classes and some species might have benefited from favourable conditions. Finally, a single species *S. miniatus*, with very good recruitment and less vulnerability to trawls (Love *et al.* 2002), exhibited high continuous increase over the survey period.

We applied, as a comparison, Hardy & Senterre (2007) test for phylogenetic clustering within lineages. This test, with I_0, shows phylogenetic clustering was found in the close-to-root nodes, and accordingly resulted from the chance occurrence of basal species. With I_2, the phylogenetic clustering was found in the lineage from *S. goodei* to *S. constellatus* and resulted mainly from the clustering of the species that exhibited the highest decline. With I_1, the significant lineages highlight that the Shannon-like phylogenetic index is influenced by both the basal species (like the richness-like index: Faith's index, I_0) and the clustering of the species that exhibited high decline (like the Simpson-like index: QE I_2).

A unified approach for analyzing phylo-β-richness and -diversity

For each evolutionary period, our phylo-β-diversity analysis tests to see if the changes between communities (either time or spatial units) correspond to changes between the lineages

that descended from this period. It can thus identify shifts in lineages at any part of the phylogeny. The parameter a in eqn 2 allows to give different weights to rare vs abundant species. In the rockfish case study, the periods where the phylogenetic differences between years were significant (phylogenetic clustering) changed according to the value of a ($a = 0$, 1, 2). Because temporal variation in species composition are likely to affect differently rare and abundant species, varying parameter a through field studies and/or simulations from 0 (rare and abundant species are evenly considered) to high values (rare species are given less importance) would allow to estimate the impact of species rarity (parameter a) on the analysis of the changes in the phylogenetic composition between subcommunities. Indeed, different selections of values for a can highlight different ecological mechanisms that underpin phylo-β-diversity. Most studies of PD have dealt with presence/absence data, which are likely to miss important ecological patterns (Vamosi *et al.* 2009). Moreover, presence/absence data are highly sensitive to the chance occurrence of a few individuals in an unfavourable environment (Vamosi *et al.* 2009). Previous studies confirmed that the strength of phylogenetic signal in communities can be changed by considering the presence/absence data vs. abundance data (Hardy & Senterre 2007; Helmus *et al.* 2007). In our case study, the effect of fishing pressure on the phylogenetic structure of the assemblage was highlighted by focusing on the most abundant species.

Further, we considered that the relative abundances based on sample data reflected actual relative abundances. However, in general some species might not be detected or the abundance underestimated because species have different probabilities of being detected. Complementary methods, which have been covered extensively elsewhere (Borchers *et al.* 2004), could be used to better estimate and model species abundance distributions. For instance, bias corrections for the Shannon ($a = 1$) diversity index due to unobserved rare species are available (e.g. Chao & Shen 2003). Their application and use to improve the analysis of the method presented here requires further research.

Our approach is clearly related to two previous developments. While Hardy & Senterre (2007) also proposed a partial analysis of the PD, their approach is different to that we presented here as they do not partition the diversity index but instead use within-lineage permutation schemes. Although the objectives of these two methods are different, this makes them entirely complementary for understanding patterns in community ecology and evolution. The second related development is the DPCoA used in Fig. 5 to complement the analysis of phylo-β-diversity as measured by I_a. In their seminal paper, Webb *et al.* (2002) proposed, as a direction for future work, the use of ordination methods on intersample distances that reflect net phylogenetic

dissimilarity, which is now possible using the DPCoA with both presence/absence and abundance data (Pavoine *et al.* 2004).

Finally, we have illustrated the application of our methodology for studying phylo-β-diversity with a hierarchical time series approach. Nevertheless, our methodology has a broader application that can be extended to spatial and spatio-temporal analyses. It can thus integrate geographical space, current history (ecological time) and evolutionary history, all of them influence patterns of α, β, and γ diversity. Depending on the scale of the study, phylo-β-diversity will contribute to our understanding of ecological and historical, biogeographic influences that have created and that continuously modify the assemblages of species. For example, in spatial analyses, the skew of phylo-β-diversity towards the root of the phylogenetic tree can result from habitat filtering and phylogenetic niche conservatism or, at larger scale, biogeographic regions and barriers to dispersal. In contrast, the skew of phylo-β-diversity to the tips of a tree can occur due to competition and replacement by less similar species from the same lineage. From temporal analyses, such as the rockfish example presented here, high β differences between years are expected to have the strongest impact on the phylogenetic structure of the focal community if they are skewed to the root. Such effects can occur from a competitive invader that impacts only part of the phylogenetic tree, or from a phylogenetic signal in species vulnerability to habitat modification.

CONCLUSION

In conclusion, we have developed a robust method for measuring phylogenetic biodiversity that can be applied in space and/or time to obtain a detailed description of evolutionary diversity in species communities. Phylogenetic partitions are essential to find the mechanisms that underpin dynamics in community assembly. Moreover, our approach allows heterogeneity in phylogenetic compositions due to modifications of the environment to also be carefully evaluated. The same methodology could be applied to functional trees (see e.g. Petchey & Gaston 2002). Functional partitions will improve the evaluation of human-driven environmental changes in conservation studies. Clearly these approaches have wide-ranging uses in understanding and managing ecological communities.

ACKNOWLEDGEMENTS

REFERENCES

Allan, J.D. (1975). Components of diversity. *Oecologia*, 18, 359–367.

Borchers, D.L., Buckland, S.T. & Zucchini, W. (2004). *Estimating Animal Abundance*. Springer, London.

Chao, A. & Shen, T.-J. (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ. Ecol. Stat.*, 10, 429–443.

Cheung, W.W.L., Pitcher, T.J. & Pauly, D. (2005). A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. *Biol. Conserv.*, 124, 97–111.

Crozier, R.H., Dunnett, L.J. & Agapow, P.-M. (2005). Phylogenetic biodiversity assessment based on systematic nomenclature. *Evol. Bioinform. Online*, 1, 11–36.

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, 61, 1–10.

Faith, D.P. (2008). Threatened species and the potential loss of phylogenetic diversity: conservation scenarios derived from estimated extinction probabilities and phylogenetic risk analysis. *Conserv. Biol.*, 22, 1461–1470.

Forest, F., Grenyer, R., Rouget, M., Davies, J., Cowling, R., Faith, D.P. *et al.* (2007). Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, 445, 757–760.

Froese, R. & Pauly, D. (2008) *FishBase*. World Wide Web electronic publication. Available at: http://www.fishbase.org (accessed on 1 June 2008).

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

Hardy, O.J. (2008). Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J. Ecol.*, 96, 914–926.

Hardy, O.J. & Jost, L. (2008). Interpreting and estimating measures of community phylogenetic structuring. *J. Ecol.*, 96, 849–852.

Hardy, O.J. & Senterre, B. (2007). Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecol.*, 95, 493–506.

Harvey, C.J., Tolimieri, N. & Levin, P.S. (2006). Changes in body size, abundance, and energy allocation in rockfish assemblages of the northeast pacific. *Ecol. Appl.*, 16, 1502–1515.

Havrda, M. & Charvat, F. (1967). Quantification method of classification processes: concept of structural α-entropy. *Kybernetik*, 3, 30–35.

Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007). Phylogenetic measures of biodiversity. *Am. Nat.*, 169, E68–E83.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 6, 65–70.

Hutchings, J.A. (2000). Collapse and recovery of marine fishes. *Nature*, 406, 882–885.

Hyde, J.R. & Vetter, R.D. (2007). The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). *Mol. Phylogenet. Evol.*, 44, 790–811.

IUCN (2008) 2008 IUCN red list of threatened species. URL http://www.iucnredlist.org

Love, M.S., Caselle, J.E. & Buskirk, W.V. (1998). A severe decline in the commercial passenger fishing vessel rockfish (*Sebastes* spp.) catch in the Southern California Bight, 1980–1996. *Calif. Coop. Ocean. Fish. Invest. Rep.*, 39, 180–195.

Love, M.S., Yoklavich, M. & Thorsteinson, L. (2002). *The Rockfishes of the Northeast Pacific*. University of California Press, Berkeley, Los Angeles, CA and London.

Nayak, T.K. (1983) *Applications of Entropy Functions in Measurement and Analysis of Diversity*. PhD Thesis, University of Pittsburgh, Pittsburgh, PA.

Patil, G.P. & Taillie, C. (1982). Diversity as a concept and its measurement. *J. Am. Stat. Assoc.*, 77, 548–561.

Pavoine, S. & Bonsall, M.B. (2009). Biological diversity: distinct distributions can lead to the maximization of Rao's quadratic entropy. *Theor. Popul. Biol.*, 75, 153–163.

Pavoine, S. & Dolédec, S. (2005). The apportionment of quadratic entropy: a useful alternative for partitioning diversity in ecological data. *Environ. Ecol. Stat.*, 12, 125–138.

Pavoine, S., Dufour, A.B. & Chessel, D. (2004). From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *J. Theor. Biol.*, 228, 523–537.

Pavoine, S., Ollier, S. & Pontier, D. (2005). Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarity indices suitable? *Theor. Popul. Biol.*, 67, 231–239.

Petchey, O.L. & Gaston, K. (2002). Functional diversity (FD), species richness and community composition. *Ecol. Lett.*, 5, 402–411.

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

Ricotta, C. (2005). Additive partitioning of Rao's quadratic diversity: a hierarchical approach. *Ecol. Modell.*, 183, 365–371.

Ricotta, C. & Szeidl, L. (2006). Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. *Theor. Popul. Biol.*, 70, 237–243.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech.*, 27, 379–423. 623-656

Shimatani, K. (2001). On the measurement of species diversity incorporating species differences. *Oikos*, 93, 135–147.

Simpson, E.H. (1949). Measurement of diversity. *Nature*, 163, 688.

Vamosi, S.M., Heard, S.B., Vamosi, J.C. & Webb, C.O. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.*, 18, 572–592.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.

Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, 21, 213–251.

SUPPORTING INFORMATION

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

Appendix S1 The R scripts (function.R).
Appendix S2 The manual that describes how to use the R scripts with an application to the rockfish data set (Manual.pdf).
Appendix S3 Details on the species Trends in CPUE values from 1980 to 2007 and on the total hours of fishing survey and number of anglers surveyed from 1980 to 2007 (AppendixS3.pdf).
Appendix S4 The phylogenetic tree as a newick format (phywick.txt).
Appendix S5 The CPUE values (number of fish caught per 1000 fishing hours) per year and *Sebastes* species (tabsp.txt).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.