



HAL
open science

A simple translation from indices of species diversity to indices of phylogenetic diversity

Sandrine Pavoine, Carlo Ricotta

► **To cite this version:**

Sandrine Pavoine, Carlo Ricotta. A simple translation from indices of species diversity to indices of phylogenetic diversity. *Ecological Indicators*, 2019, 101, pp.552-561. 10.1016/j.ecolind.2019.01.052 . hal-02291153

HAL Id: hal-02291153

<https://hal.sorbonne-universite.fr/hal-02291153v1>

Submitted on 18 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A simple translation from indices of species diversity to indices of**
2 **phylogenetic diversity**

3

4 Sandrine Pavoine^{a,*}, Carlo Ricotta^b

5

6 ^a *Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum national*
7 *d'Histoire naturelle, CNRS, Sorbonne Université, 57 rue Cuvier, 75005 Paris, France.*

8 ^b *Department of Environmental Biology, University of Rome 'La Sapienza', Piazzale Aldo*
9 *Moro 5, 00185 Rome, Italy*

10

11 * Corresponding author at: UMR7204 CESCO, MNHN, 61 rue Buffon, 75005 Paris, France.

12 *E-mail address:* sandrine.pavoine@mnhn.fr (S. Pavoine)

13

14 ABSTRACT

15 Measures of phylogenetic diversity have two main objectives: first disentangling the
16 processes that drive species assemblages and second defining priorities of conservation while
17 considering how much each species might contribute to biodiversity. A now widely used
18 approach for measuring phylogenetic diversity consists in summing branch lengths on
19 phylogenetic trees. Thanks to this approach, a change of perspective has been done since the
20 90s in the ecological literature: to measure biodiversity in sites, species in compositional
21 indices have been replaced, often implicitly and sometimes explicitly, with phylogenetic units
22 (units on the branch lengths of a phylogenetic tree). Here we highlight how any traditional
23 index of species diversity, can be adapted, with this simple change of perspective, to measure
24 phylogenetic diversity. We analyze other potentials of this simple family of indices: the ratio
25 of phylogenetic diversity to the associated species diversity may offer a meaningful way to
26 apprehend phylogenetic uniqueness in communities; and the phylogenetic diversity
27 standardized by its maximum value may provide a meaningful way to apprehend phylogenetic
28 evenness. Theoretically, the same approach could be applied to functional dendrograms to
29 assess functional diversity, evenness and uniqueness. However this requires, somehow, to
30 force functional data to adhere to a hierarchical structure of functional differences between
31 species. As an illustration, we analyze phylogenetic and functional diversity in plant
32 communities along a primary succession in Italy. Overall, this simple framework has a critical
33 potential for future analyses of tree-based diversity as it benefits from myriads of previous
34 researches on the measures of species diversity and species evenness.

35

36 *Keywords:* biodiversity; evenness; functional diversity; Hill numbers; phylogenetic diversity;
37 redundancy

38

39 **1. Introduction**

40

41 Faith (1992) measured phylogenetic diversity as the sum of branch lengths in a
42 phylogenetic tree. His main objective was to introduce a new point of view in conservation
43 studies, where the main measure of biodiversity at large scale was the number of species or
44 “species richness”. He proposed replacing species, as the basic units for conservation, by
45 features of species, where feature means any distinct character state (Faith et al., 2009).
46 According to Faith (1992), “the diversity of features represented by a subset of species
47 provides option value in ensuring not only that one or more members of the subset can adapt
48 to changing conditions, but also that society may be able to benefit (e.g. economically) from
49 features of these species in response to future needs”. His index, the sum of branch lengths in
50 a phylogenetic tree, is thus a proxy for a richness in species’ features. It relies on the
51 assumption that the lengths of a branch in a phylogenetic tree is an indicator for a number of
52 features that have arisen once in the hypothetical ancestral species represented along the
53 branch.

54 Several published coefficients of biodiversity within sites (or communities,
55 assemblages, stations, regions, samples, quadrats, plots, etc.) explicitly exploit the structure of
56 a rooted tree with species as tips. The first index, by Faith (1992) combined those trees with
57 the presence of species in sites, to evaluate the phylogenetic diversity of sites. Other
58 coefficients, as those developed by Allen et al. (2009), Pavoine et al. (2009a) and Chao et al.
59 (2010), also include species’ abundances (or potentially any measure giving the importance of
60 a species in a site, as for example the biomass or the percentage cover for plants). Among
61 these coefficients, those developed by Pavoine et al. (2009a) and Chao et al. (2010) include a
62 parameter to regulate the relative importance given to rare vs. abundant species.

63 A common methodological practice emerges from these developments. Indeed, one
64 way of developing a new field of research is to transpose approaches developed from a
65 previous field to the new one (e.g. Faith et al., 2009). Three concepts have long dominated the
66 research for biodiversity metrics: species richness, evenness, and diversity (e.g. Magurran,
67 2004). As shown above, species richness counts the number of species in a site. Species
68 evenness deals with how abundant species are and, more precisely, it evaluates whether
69 species' abundances are even or whether some species dominate by their high abundance (a
70 widespread profile observed in ecological communities). Species diversity indices more
71 generally measure species richness combined with species evenness.

72 Following Faith (2013) and Pavoine (2016), here we show that several approaches
73 now used to measure phylogenetic diversity within sites are actually strongly rooted on these
74 more traditional species-centered approaches. From a methodological point of view, these
75 approaches replace, in all traditional coefficients, species by units of the branch lengths of a
76 phylogenetic tree. We thus show how traditional measures of species diversity in a site, can be
77 transposed to measure the phylogenetic diversity in a site. In addition, we discuss the
78 possibility of applying similar approaches to functional diversity. Our objective here is to
79 highlight the simplicity and at the same time the potential and limits of this family of indices.

80

81 **2. Methods**

82

83 *2.1. Principle*

84

85 The principle is the same as that considered in Pavoine (2016) for dissimilarity
86 indices. Here it is applied to measure tree-based diversity. Consider a rooted tree T with S
87 species as tips and K branches. A_i is the abundance of species i ; t_k is the set of species

88 descending from branch k ; t_T is the whole set of species (tips of the tree); b_T is the set of
89 branches in the tree T ; L_k is the length of branch k ; a_k is the sum of abundances for all species
90 descending from branch k ($a_k = \sum_{i \in t_k} A_i$). Let $A = \{A_i\}_{i \in t_T}$ be the set of species abundances. The
91 principle of the change of perspective consists in replacing species in diversity coefficients by
92 units of branch lengths in phylogenetic trees. A branch k of a tree with a length equal to L_k
93 will be considered to have L_k units. Each unit on this branch has an abundance value equal to
94 a_k . The sum of abundances for all units supported by the branch is thus $L_k a_k$. To ease the
95 writing below, we use the word "feature" to designate these branch units.

96 Theoretically, these notations and the reasoning below can be applied to any rooted
97 tree with arbitrary or estimated branch lengths and species as tips. For example, ten years after
98 Faith (1992), Petchey and Gaston (2002) suggested summing branch lengths on functional
99 dendrograms instead of phylogenetic trees to obtain an index of functional diversity. They
100 developed this measure as “an amalgam of the diversity caused by species richness, number
101 of functional groups, community composition, and species identity”, useful for example to
102 connect diversity with ecosystem functioning. This approach, however, is not unanimously
103 accepted (see e.g., Poos et al., 2009; Pavoine, 2016). Functional dendrograms can be obtained
104 by first calculating functional dissimilarities between species (e.g. Pavoine et al., 2009b) and
105 then using a clustering approach to obtain a tree from these dissimilarities (e.g. Mouchet et al.,
106 2008). They are thus sensitive to the methodology used to define a tree hierarchy from
107 functional dissimilarities between species (Podani and Schmera, 2007; Mouchet et al., 2008).

108

109 2.2. Richness

110

111 As far as we know, the first index of this general family was thus developed by Faith
112 (1992) as $PD = \sum_{k \in b_T} L_k$, considering a phylogenetic tree. PD stands for "Phylogenetic

113 Diversity". It is an index of feature richness as it counts the number of units on the branch
 114 lengths of a tree, a branch length being assimilated to a richness in features. Then, Petchey
 115 and Gaston (2002) adapted this for a functional tree, leading to $FD = \sum_{k \in b_T} L_k$, where FD
 116 stands for "Functional Diversity". In the framework developed in the present paper, we thus
 117 rename their indices *fric*, for feature richness.

118

119 2.3. Diversity indices

120

121 Considering species abundances and the notations given above, branch k supports L_k
 122 features each of which has an abundance equal to a_k and a relative abundance equal to
 123 $a_k / \sum_{l \in b_T} L_l a_l$. For $a_k / \sum_{l \in b_T} L_l a_l$ to always be lower than, or equal to, unity, we impose a
 124 weak restriction on the definition of branch lengths: the sum of branch lengths on the path
 125 between a species and the root of the tree must be at least equal to 1. A stronger restriction
 126 would be that each branch must support at least one unit, that is to say, $L_k \geq 1$ for all $k \in b_T$;
 127 but this stronger restriction is actually not necessary here (see Appendix A). One can easily
 128 rescale the branches of any tree to respect such restrictions by multiplying them by a constant.
 129 We will show below that not all indices need even the weak restriction but that having it
 130 ensures that any diversity index defined on the relative abundance of some entities can be
 131 applied to the values $a_k / \sum_{l \in b_T} L_l a_l$ for features.

132 For instance, a measure of feature diversity using the Shannon (1948) index would be:

$$133 \quad fdiv_{Shannon}(T, A) = - \sum_{k \in b_T} L_k \frac{a_k}{\sum_{l \in b_T} L_l a_l} \log \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right) \quad \text{Eq. (1)}$$

134 and a measure of feature diversity using the Gini-Simpson index (Gini, 1912; Simpson, 1949)
 135 would be:

136
$$fdiv_{GS}(T, A) = 1 - \sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^2$$
 Eq. (2)

137 In both measures, the relative abundance of features along the tree substitutes the classical
 138 species abundances, which are usually adopted for measuring diversity.

139

140 *2.4. Parametric diversity indices*

141

142 In parametric diversity indices, the entities used for calculating diversity are weighted by a
 143 parameter, here q , that regulates the relative importance given to rare compared to abundant
 144 entities. Applied to species, if $q = 0$, parametric indices are disproportionately sensitive to the
 145 rare species; then if q increases, they are more and more sensitive to the abundant species
 146 compared to the rare species. Applied to features, if $q = 0$, parametric indices are
 147 disproportionately sensitive to the rare features; then if q increases, they are more and more
 148 sensitive to the abundant features compared to the rare features.

149 Although other parametric indices exist (e.g. Behara & Chawla, 1975) and could be
 150 applied to features, we concentrate here on parametric generalizations of *fric*, *fdiv_{GS}* and
 151 *fdiv_{Shannon}* (Table 1). For example, a generalization can be obtained using the HCDT entropy
 152 (Havrda & Charvat, 1967; Daróczy, 1970; Tsallis, 1988):

153

154
$${}^q fdiv_{HCDT}(T, A) = \left[1 - \sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right] / (q - 1)$$
 Eq. (3)

155

156 Applying the framework to Hill effective numbers (Hill 1973) leads to Chao et al. (2010)
 157 index, developed in the context of phylogenetic diversity and named ${}^q PD$ ($q \geq 0, q \neq 1$) (Faith,
 158 2013). ${}^q PD$ provides an effective number of features: the number of features with even

159 abundance a theoretical community should have to obtain the same value of HCDT diversity
 160 as that of the real community (see Appendix A for an example). To include qPD fully into the
 161 framework developed in the present paper, we rename it here as the ${}^qfdiv_{Hill}$ index. Its formula
 162 is:

163

$$164 \quad {}^qfdiv_{Hill}(T, A) = \left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right]^{\frac{1}{(1-q)}} \quad \text{Eq. (4)}$$

165

166 Another generalization of $fric$, $fdiv_{GS}$ and $fdiv_{Shannon}$, the Rényi's entropy (Rényi,
 167 1960), could also be applied to feature abundances, leading to

168

$$169 \quad {}^qfdiv_{Rényi}(T, A) = \frac{1}{1-q} \log \left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right] \quad \text{Eq. (5)}$$

170

171 The fact that these parametric indices are generalizations of $fric$, $fdiv_{GS}$ and $fdiv_{Shannon}$ is
 172 shown in Table 2.

173 These three parametric indices are linked. However, they have different properties. For
 174 example, if one multiplies the branch length of the tree by a constant, say N , leading to a new
 175 tree T_N , then, all these indices are increasing functions of N . However, although ${}^qfdiv_{Hill}$ is
 176 simply multiplied by N , the new value for ${}^qfdiv_{HCDT}$ is

177

$$178 \quad {}^qfdiv_{HCDT}(T_N, A) = \left[1 - N^{1-q} \sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right] / (q-1) \quad \text{Eq. (6)}$$

179

180 and that for ${}^q fdiv_{Rényi}$ is

181

$$182 \quad {}^q fdiv_{Rényi}(T_N, A) = \frac{1}{1-q} \log \left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right] + \log N \quad \text{Eq. (7)}$$

183

184 An important consequence of this is that if the weak restriction on branch lengths was not
185 applied, ${}^q fdiv_{HCDT}$ and ${}^q fdiv_{Rényi}$ could have negative values, while ${}^q fdiv_{Hill}$ would still be
186 nonnegative and meaningful (see details in Appendix A).

187 ${}^q fdiv_{Hill}$, ${}^q fdiv_{HCDT}$ and ${}^q fdiv_{Rényi}$ are non-increasing functions of q (Nayak, 1985;
188 proposition 2.4; see also Leinster and Cobbold, 2012). The curve $q \rightarrow$ diversity has often been
189 called a diversity profile, especially in the context of the Hill numbers (e.g. Leinster and
190 Cobbold, 2012). The diversity profiles of these three parametric indices (${}^q fdiv_{Hill}$, ${}^q fdiv_{HCDT}$
191 and ${}^q fdiv_{Rényi}$) are consistent as, for a fixed q , they order communities similarly from the least
192 to the most diverse (Nayak, 1985). With any of these indices, the slope of the diversity profile
193 reflects the unevenness of the relative abundances of evolutionary units. The more uneven the
194 distribution of relative abundances, the more steeply the curve should decline (e.g. Hill,
195 1973). The indices are, however, unequal in the degree of steepness of the diversity profile.
196 For example, for fixed tree and abundance distribution, if $q < 1$, ${}^q fdiv_{Rényi} \leq {}^q fdiv_{HCDT}$, while if
197 $q > 1$, ${}^q fdiv_{Rényi} \geq {}^q fdiv_{HCDT}$ (Nayak, 1985). The two indices merge for $q=1$ where they converge
198 towards $fdiv_{Shannon}$. This implies that, for fixed tree and abundance distribution, the diversity
199 profile obtained with ${}^q fdiv_{HCDT}$ will always be steeper than the diversity profile obtained with
200 ${}^q fdiv_{Rényi}$. Also, ${}^q fdiv_{Rényi} = \log({}^q fdiv_{Hill})$; given the well-known properties of the logarithm
201 function and given that ${}^q fdiv_{Hill}$ takes values at least equal to 1, ${}^q fdiv_{Hill}$ would thus lead to
202 diversity profiles that are also steeper than those obtained with ${}^q fdiv_{Rényi}$.

203

204 2.5. Evenness indices

205

206 Indices of species evenness measure how evenly abundance is distributed among species.

207 Developing an index of feature evenness would thus mean measuring how evenly abundance

208 is distributed among features. Before exploring feature evenness, we first recall a traditional

209 way of measuring species evenness. Species diversity indices are maximized when species

210 have even abundances. Dividing species diversity indices by their maximum could thus

211 theoretically indicate how much abundance is evenly distributed among species. For example,

212 Pielou (1966) proposed the Rényi entropy divided by its maximum as a measure of species

213 evenness; and Taillie (1979) the Hill numbers divided by their maximum (Ricotta and Avena,

214 2002). There are however two limits of this simple evenness approach.

215 The first limit is that many species evenness indices obtained as the ratio of a diversity

216 index by its maximum actually depend on species richness (e.g. Sheldon, 1969). An important

217 property for an evenness index is thus that its value should not change under replication (Hill,

218 1973): replicating the composition of a community N times should give a super-community

219 with the same evenness as the original community. Among the parametric indices discussed

220 above, only the Hill index satisfies this property (as shown in Appendix B; see also Ricotta

221 and Avena, 2002).

222 The second limit is that although the maximum of the Hill numbers and of the Rényi's

223 entropy are constant, irrespective of the parameter q , the maximum of HCDT depends on q .

224 This dependence affects the interpretability of the ratio of HCDT to its maximum as a

225 measure of evenness. Indeed, if q controls the importance given to rarity (for low values of q)

226 vs. commonness (for high values of q), a parametric evenness index defined as the ratio of a

227 diversity index to its maximum, should decrease with q . When, $q=0$, all parametric indices are

228 unaffected by abundance data; the evenness in that case should be maximum and equal to 1.

229 When q tends to infinity, the highest abundance is the most influential in the diversity index.
230 The evenness, in that case, should tend to the minimum. Therefore, parametric evenness
231 should decrease with q . The Hill numbers divided by their maximum and the Rényi's entropy
232 divided by its maximum both respect this property (Appendix A). However, the HCDT
233 entropy divided by its maximum displays non-monotonous variations (Appendix C). As q
234 tends to infinity, the ratio of HCDT entropy to its maximum value tends to 1 (Appendix A).
235 This wrongly suggests high abundance evenness if we consider that evenness varies between
236 0 and 1.

237 Measures of evenness could also be approached by using other indices developed
238 independently of a diversity index as Smith and Wilson's (1996) E_{var} evenness index (see also
239 Magurran, 2004). However, a third limit of the evenness approach appears when this concept
240 is applied to features rather than species: provided at least one branch on any tree T is shared
241 by at least two species, then the abundance of branch units cannot be even. For example,
242 consider two species i and j having the same abundance a . Then any feature (branch unit) on
243 the terminal branches that support the two species will be associated with an abundance of a ,
244 but the features on the shared branches would be associated to a minimum of $2a$ (Fig. 1).

245 Considering this third limit, we propose in Table 3 indices of feature evenness defined
246 as the observed feature diversity divided by the value feature diversity would have if species
247 had same abundance and were independent (species are connected to the root of the
248 phylogenetic tree but do not share branches; e.g. Fig. 2a,b). Trees are said ultrametric if the
249 sum of branch lengths on the shortest path from tip to root is constant. In case of non-
250 ultrametric trees, we can consider another scenario to obtain coefficients of feature evenness.
251 In this second scenario, the diversity is divided by the value it would have if species had same
252 abundance, were independent, and at maximum distance to the root of the tree (this maximum
253 is at least equal to the maximum observed distance; e.g. Fig. 2a,c). Let h_i be the distance (sum

254 of branch lengths) between species i and the root of the tree. Then, using solution 1, leads for
255 example to the evenness indices listed in Table 3. Using the second solution, the formulas are
256 unchanged but $\sum_{i \in \mathcal{T}_r} h_i$ is equal to $S \times \max(h_i)$ or $S \times H$, where S is the number of species and
257 H is the maximum possible height of the tree. For ultrametric trees, the two scenarios are
258 equivalent.

259 Considering the three identified limits to a simple transfer from measuring species
260 evenness to measuring feature evenness, the index ${}^q\text{feve}_{\text{Hill}}$ in Table 3 could thus be a good
261 candidate for measuring feature evenness.

262

263 *2.6. Uniqueness/Redundancy indices*

264

265 Consider a general context where diversity is measured as a function of the abundance of
266 some entities. Here we considered two types of entities: first species and second branch units
267 on a tree with species as tips. A diversity measure is said ‘Schur-concave’ if it increases when
268 abundance is transferred from an entity to another strictly less abundant entity (Dalton’s
269 (1920) principle of transfers; Patil and Taillie, 1982). The HCDT index is Schur-concave if
270 and only if its parameter q is non-negative (Patil and Taillie, 1982). This implies that the
271 richness, the Shannon index and the Gini-Simpson index are all Schur-concave. The Rényi
272 entropy and the Hill index are also Schur-concave for all non-negative values of q (Nayak,
273 1985, proposition 2.6). The framework, below, is close, but not equal to that developed by
274 Ricotta et al. (2016).

275 Consider a tree, T . Consider the associated star-shaped tree, T_{star} , which has the same
276 species as tips, and the same distances (sum of branch lengths) between species and the root
277 of the tree as in T . Compared to T , T_{star} does not have branches shared by species. For a fixed
278 distribution of species abundance, the ratio of any Schur-concave diversity index applied to T

279 versus T_{star} is an index of uniqueness (see details in Appendix A). Its values lies between 0
280 and 1. It is low if species are redundant, being closely related. It is high if species are original
281 or unique, sharing no branches on the tree. Examples of uniqueness indices are given in Table
282 4. These rely on the richness, Shannon, Gini-Simpson, Hill, Rényi and HCDT indices.

283 However, a limit of ${}^qf_{uniHCDT}$ (Table 4), the uniqueness index based on the HCDT
284 entropy, is that it always tends to 1 when q tends to infinity. Inspired by evenness indices, a
285 reasonable property that could thus be required for uniqueness indices is thus as follows:
286 replicating the composition of a community N times as illustrated in Fig. 3 should give a
287 super-community with the same uniqueness as the original community. As for evenness
288 indices, the species and their abundance are replicated N times. As for their associated tree,
289 the tree is also replicated N times and the N resulting trees are connected by their root (see the
290 example in Fig. 3). Only the uniqueness index derived from the Hill numbers (in Table 4)
291 fulfils this replication principle (as shown in Appendix B and illustrated in Fig. 3).

292 At this stage, we thus end up with a set of parametric diversity, evenness, and
293 uniqueness indices that exploit the shape of a phylogenetic tree and that differ in their
294 mathematical properties, with consequences on their relative biological interest. Among these
295 indices, those derived from the Hill numbers counts the highest number of essential, useful
296 properties (Table 5, see also Appendix D for illustrative examples).

297

298 2.7. Data analysis

299 We analyzed phylogenetic and functional diversity of plant communities along a primary
300 succession on the foreland of the Rutor glacier in Italy (Caccianiga et al. 2006). The data set
301 comprises 59 plots divided into three succession stages based on the age of the moraine
302 deposits: 17 early-successional plots, 32 mid-successional plots and 10 late-successional
303 plots. Plant species abundances were measured with a five-point ordinal scale transformed to

304 ranks (Caccianiga et al. 2006; see also Ricotta et al. 2016). As in Ricotta et al. (2016), we
305 used Grime (1974, 2001) to classify plant species as competitors (C), stress tolerators (S) and
306 ruderals (R) by means of fuzzy-coded values in the range 0–100, such that $C+S+R = 100$; and
307 we obtained functional dissimilarities between species by applying the Marczewski-Steinhaus
308 (MS) coefficient of dissimilarity to the functional species classification (C, S, R). More details
309 on the data set are given in Caccianiga et al. (2006). We obtained a functional tree for the
310 species of the study area by applying the unweighted pair group method with arithmetic mean
311 (UPGMA) to the functional dissimilarities between species. We considered the phylogenetic
312 tree developed by Durka and Michalski (2012) pruned to retain only the species observed in
313 the Rutor glacier data set.

314 Using different mathematical formulas, Ricotta et al. (2016) analyzed the same data
315 set to evaluate functional uniqueness and Ricotta et al. (2018) the same data set to evaluate
316 phylogenetic uniqueness. These studies concluded for a decline in phylogenetic and
317 functional uniqueness along the primary succession, from the early stage, through the mid
318 stage to the late stage. We re-analyzed these data using indices in Tables 1, 2, 3 and 4 and
319 then focusing on ${}^qdiv_{Hill}$, ${}^qeve_{Hill}$ (species indices), ${}^qfdiv_{Hill}$, ${}^qfeve_{Hill}$, and ${}^qfuni_{Hill}$ (feature
320 indices). Ricotta et al. (2018) developed an index of phylogenetic diversity close to our
321 ${}^qfuni_{HCDT}$ index. Ricotta et al. (2016) based their approach instead on Rao's quadratic entropy
322 (Q , Rao, 1982), an extension of the Gini-Simpson index ($=div_{GS}^2 div_{HCDT}$, see Table 1) where
323 they included functional information on species. Consider that we draw at random two
324 individuals from a community. The Gini-Simpson index is the probability that the two
325 individuals belong to different species, while Rao index in that case is the expected amount of
326 functional difference between the two individuals. If functional differences between species
327 are bounded between 0 and 1, then the maximum possible value for Rao index over all
328 possible functional differences (fixing species abundance) is the value taken by the Gini-

329 Simpson index. Ricotta et al. (2016) thus considered that the ratio of Rao index to Gini-
330 Simpson index ($U=Q/div_{GS}$) is a measure of functional uniqueness. As special cases of the
331 HCDT entropy and its extensions, Rao and Gini-Simpson indices are sensitive to replication
332 in species abundance and patterns of difference: replicating N times the composition of a
333 community as in Fig. 3, Gini-Simpson and Rao's values are not multiplied by N . However,
334 consider transformed versions of these indices driving them as part of the Hill framework:
335 ${}^2div_{Hill}=1/(1-div_{GS})$ for the transformed Gini-Simpson index and $Q^*=1/(1-Q)$ for the
336 transformed Rao index. These transformed versions of the Gini-Simpson and Rao indices
337 fulfil the replication principle: replicating N times the composition of a community as in Fig.
338 3, the values taken by these transformed indices are multiplied by N (Appendix B). To
339 compare Ricotta et al. (2016) approach with our new perspective, we thus re-calculated their
340 results considering the following modified version of their index of functional uniqueness:

341

$$342 \quad U^* = \frac{1/(1-Q)}{1/(1-div_{GS})} = \frac{1-div_{GS}}{1-Q} \quad \text{Eq. (8)}$$

343

344 where Q is applied to the functional distances between species directly obtained with the
345 Marczewski-Steinhaus coefficient. As in Ricotta et al. (2016), the U^* index of functional
346 uniqueness does not require the definition of a functional tree. We thus compared this
347 dissimilarity-based approach with the framework developed here that demand to take the risk
348 of distorting trait data by transforming dissimilarities into a functional dendrogram.

349

350 **3. Results**

351 We synthesize results obtained on ${}^qdiv_{Hill}$, ${}^qeve_{Hill}$ (species indices) and ${}^qfdiv_{Hill}$, ${}^qfeve_{Hill}$, and
352 ${}^qfuni_{Hill}$ (feature indices) in Figures 4 and 5 and give detailed results for all indices in
353 Appendix C. Regarding diversity, as in Ricotta et al. (2016, 2018), we observed reversed U-

354 shaped variations in species diversity, whatever the index used, with the mid-successional
355 stage having the highest species diversity (Fig. 5, Table 6, Appendix C). With $q>1$, as in
356 Ricotta et al. (2016, 2018), we observed a decrease in phylogenetic and functional diversity
357 from early and mid successional stages to the last stage (Fig. 4, Appendix C). With $q=0$
358 (presence-absence data), values of phylogenetic and functional diversity were, on average,
359 more similar between the successional stages (Fig. 4, Appendix C).

360 As regards uniqueness, using Rényi and HCDT indices with $q=1,2$ for functional data
361 or $q=2$ for phylogenetic data, our results were similar as those obtained by Ricotta et al.
362 (2016, 2018): a decrease in uniqueness along the succession from the early stage to the late
363 stage (Appendix C). Using these indices with other values for q , and using Hill numbers, our
364 results were, however, different from those obtained by Ricotta et al. (2016, 2018): we
365 observed U-shaped variations with a decrease in functional uniqueness from the early
366 successional stage to the mid-successional stage, and intermediate values in the late stage,
367 often not significantly different from both the early and mid stages (Fig. 4; Appendix C). We
368 observed similar trends in values of functional and phylogenetic evenness and U-shaped
369 variations in species evenness (Figs. 4, 5; Appendix C). Contrary to functional patterns,
370 species patterns, like phylogenetic patterns, did not remain significant, however, after we
371 applied the correction for multiple tests.

372 Using Rao's quadratic entropy framework expressed in terms of equivalent number of
373 species, we obtained a significant decrease in functional diversity from the early stage to the
374 late stage and a decrease in functional uniqueness from the early to the mid stage, the
375 functional uniqueness of the late stage being not significantly different from both the early
376 and mid stages (Table 6).

377

378 **4. Discussion**

379

380 The framework developed here first considered replacing species in diversity indices
381 by branch units on phylogenetic trees with species as tips. This approach is simple and can be
382 applied to any species diversity index developed so far. In the literature, some have
383 alternatively developed new indices without reference to traditional species diversity indices
384 (e.g. Webb et al., 2002). Others applied traditional indices of diversity to the abundances of
385 branch units normalized by the total abundance of all species in a site. Contrary to the indices
386 developed in the Methods section, these alternative indices are still centered on the species as
387 the target unit to measure diversity, even if they consider the evolutionary differences
388 between these species. Pioneering this alternative approach, Allen (2009) applied Shannon
389 index to the abundance of a feature scaled by the summed abundance of all species, rather
390 than by the summed abundance of all features:

391

$$392 \quad H_p(T, A) = - \sum_{k \in b_T} L_k \frac{a_k}{\sum_{i \in t_T} A_i} \log \left(\frac{a_k}{\sum_{i \in t_T} A_i} \right) \quad \text{Eq. (9)}$$

393

394 Pavoine et al. (2009a) generalized Allen index thanks to the HCDT index, in the
395 particular case of ultrametric trees. More generally, whatever the tree shape, their index could
396 be written as follows:

397

$$398 \quad I_q(T, A) = \left[\frac{\sum_{l \in b_T} L_l a_l}{\sum_{i \in t_T} A_i} - \sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{i \in t_T} A_i} \right)^q \right] / (q-1) \quad \text{Eq. (10)}$$

399

400 If the tree is ultrametric, then the ratio of features' abundance to species' abundance (
401 $\sum_{l \in b_T} L_l a_l / \sum_{i \in t_T} A_i$) is simply equal to the height of the tree and I_0 is independent of
402 abundance data as it is equal to Faith phylogenetic diversity minus the height of the tree. If the
403 tree is not ultrametric, however, I_q depends on the ratio of features' to species' abundance even
404 if $q=0$. If species have similar abundances ($E=(e \dots e \dots e)$), and if they are completely distinct
405 (star-shaped tree T_{star} , with h_i the length of the branch from species i to the root of the tree),
406 then

$$408 \quad I_q(T_{star}, E) = \frac{\sum_{i \in t_T} h_i}{S} \left[\frac{1 - S^{1-q}}{q-1} \right] \quad \text{Eq. (11)}$$

409
410 $\sum_{i \in t_T} h_i / S$ is the average contribution of each species to feature richness in the star-shaped
411 tree. $(1 - S^{1-q}) / (q-1)$ is the maximum value of the HCDT index applied to species. Chao et al.
412 (2010) also generalized Allen index but using instead the Hill index, which led to:

$$414 \quad {}^q D(T, A) = \left[\sum_{k \in b_T} L_k \frac{\sum_{i \in t_T} A_i}{\sum_{l \in b_T} L_l a_l} \left(\frac{a_k}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{(1-q)}} \quad \text{Eq. (12)}$$

415
416 As for I_q , when applied to non-ultrametric trees, ${}^q D$ always depends on species' and features'
417 abundance as ${}^0 D(T, A) = \sum_{k \in b_T} L_k \sum_{i \in t_T} A_i / \sum_{l \in b_T} L_l a_l$. If species have similar abundances and if
418 they are completely distinct (star-shaped tree), then ${}^q D$ is equal to the number of species (S).

419 In contrast, its equivalent in the framework developed here, i.e., ${}^q fdiv_{Hill}$, would be equal to the
420 number of features in the star-shaped tree (sum of branch lengths = $\sum_{i \in \tau_r} h_i$). These two
421 closely related frameworks, which both extend traditional species indices to phylogenetic
422 diversity, are thus competing: the former with each feature's abundance normalized by the
423 summed abundance of all *features* and the latter with each feature's abundance normalized by
424 the summed abundance of all *species*. A main difference between these two alternative
425 approaches is that the former, used in our present framework, completely shifts the focus of a
426 study from species to features. By contrast, the alternative indices, corresponding to the latter
427 approach, are still centred on species richness, and how much each species contributes to
428 feature diversity. Choosing between the two approaches thus depends of which unit is of most
429 interest: species or feature.

430 Among the feature-centered indices we explored, our reasoning led us to select the
431 Hill numbers because they fulfill the replication principle, a property which Hill (1973)
432 discovered as critical to the development of evenness indices. Hill (1973) developed a first
433 version of the replication principle in the context of evenness indices applied to species
434 abundances: "Consider a species-abundance relation having the property that each species is
435 matched by a "double" of the same abundance. [...] Intuitively, this has the same evenness as
436 the corresponding species-abundance relation in which each species and its "double" are
437 combined to form one super-species". Hill demonstrated that an evenness index defined as the
438 ratio two Hill numbers with different values of parameter q satisfies this property. More
439 generally, if the species has more than one "double" in a reference community, say if it is
440 replicated N times, the diversity of the reference community is N times that of the community
441 where a species and its replicated species are combined. This property corresponds to the
442 weakest version of the replication principle: if N samples share no species but have the same
443 number of species with same distribution of abundance, then the diversity of the merged

444 samples is N times that of one of the individual samples. Others used a stronger version of the
445 replication principle satisfied by the Hill numbers: if N samples share no species but are
446 equally large and equally diverse, then the diversity of the merged samples is N times that of
447 one of the individual samples (see e.g. Chao et al., 2010, Leinster and Cobbold, 2012, and
448 references therein).

449 Several definitions of a generalized replication principle have been elaborated in the
450 literature in the context of phylogenetic and functional diversity (see Botta-Dukat, 2018, for a
451 review). Compared to these previous definitions, in the present paper, we used a generalized
452 version of the weak replication principle rather than the strong version: if the species of a
453 community and their abundance are replicated N times and if the associated tree with species
454 as tips is also replicated N times (the N resulting trees being connected by their root), then the
455 diversity of resulting super-community must be N times that of the original community. The
456 uniqueness and evenness indices developed in our framework being ratios of diversity indices,
457 the feature uniqueness and evenness in a community do not change by replication if the
458 diversity index used fulfills this replication principle. This property strongly restricted the
459 possible mathematical formulas to the Hill family.

460 The replication principle was useful in the development of feature evenness and
461 feature uniqueness indices using the shift from species to features. In Ricotta et al. (2016), we
462 used the Gini-Simpson index and its generalization to functional diversity (using Rao's
463 quadratic entropy) to calculate functional uniqueness from functional dissimilarity matrices.
464 As underlined in the Method section, the Gini-Simpson index is the special case of the HCDT
465 entropy with $q=2$. In Ricotta et al. (2018), we considered another framework also based on the
466 HCDT entropy to calculate phylogenetic uniqueness from phylogenetic trees. Our results
467 show that using an approach based on the Hill index applied to phylogenetic trees, functional
468 trees, or functional dissimilarities modified our evaluation of the change in functional and

469 phylogenetic uniqueness along the primary succession in Italy. With the frameworks based on
470 HCDDT or Rényi entropy, we observed a decrease in both phylogenetic (as in Ricotta et al.,
471 2018) and functional (as in Ricotta et al., 2016) uniqueness along the primary gradient. By
472 contrast, with the present framework using the Hill index, we observed U-shaped variations
473 with the highest levels of uniqueness at the early stage, the lowest at the mid-stage and from
474 intermediate to high levels at the late stage. Considering corrections for multiple test, only
475 functional patterns remained significant along the successional gradient, which may indicate
476 that species traits are critically important in determining their positions along the gradient.

477 Nonetheless, our results on functional patterns may be partly affected by the use of a
478 functional dendrogram. Using Rao's quadratic entropy framework expressed in terms of
479 equivalent number of species ($Q^*=1/(1-Q)$), we observed similar U-shaped variations in
480 functional uniqueness. Yet, we obtained a significant decrease in functional diversity from the
481 early stage to the late stage. In contrast, using Hill numbers and the framework developed
482 here (${}^qfdiv_{Hill}$), expressed in terms of equivalent number of features, we obtained a significant
483 decrease in functional diversity from the early/mid stages to the late stage: functional diversity
484 was not significantly different between the early and mid-successional stages. This difference,
485 obtained by the use of ${}^qfdiv_{Hill}$ rather than Q^* , could be due to a difference in viewpoint:
486 equivalent number of species for Rao's framework (Q^*) and equivalent number of features for
487 the present framework (${}^qfdiv_{Hill}$).

488 However, the functional tree we used is ultrametric. Chao et al. (2010) underlined that,
489 with ultrametric trees, phylogenetic diversity can be expressed in terms of equivalent number
490 of species by using the mean phylogenetic diversity index, ${}^qD = {}^qfdiv_{Hill}/H$, where H is the
491 height of the tree (from tips to root). As H is a positive constant, patterns in qD along the
492 succession are expected to be similar as those obtained with ${}^qfdiv_{Hill}$. This difference we
493 observed in our results could also be due to the use of different mathematical formulas to

494 measure diversity: Rao' quadratic entropy versus Hill numbers. However, Chao et al. (2010)
495 also demonstrated that when $q=2$ and dissimilarities between species are calculated as the sum
496 of branch lengths on the path that connect the two species in the tree, then ${}^2fdiv_{Hill} = H/(1-$
497 $Q/H)$. As H is a constant, ${}^2fdiv_{Hill}$ and Q^* are both increasing functions of Q and their patterns
498 along the succession should thus be similar. The difference we observed in our results can
499 thus be attributed to the use of a functional dendrogram in ${}^2fdiv_{Hill}$, while we applied Q^*
500 directly on the functional dissimilarities between species.

501 Using another clustering approach (e.g. Mouchet et al., 2008), we might have obtained
502 more similar results between the two approaches. However, this simple study shows the
503 potential impact of methodological choices on our interpretation of diversity patterns.
504 Obtaining a functional dendrogram adds a methodological step in the measurement of
505 functional diversity. Contrary to phylogenetic data, this methodological step is not supported
506 by the biological meaning gained by representing data in a tree shape. Indeed, cladograms
507 aims to depict the tree of life and are thus not artificial. It is legitimate to require that the
508 shared evolutionary history of sister taxa contribute only once to phylogenetic diversity
509 (Faith, 1992; May, 1994). Since a dissimilarity matrix does not contain information on shared
510 evolutionary history, tree-based diversity indices must be used to satisfy this requirement.
511 However, when one aims to compare functional and phylogenetic diversity patterns, an also
512 legitimate temptation is to use the phylogenetic diversity index most appropriate to
513 phylogenetic data and a different index to measure functional diversity more appropriate for
514 functional data. The problem in that case is that the obtained results could be obscured by
515 artifacts due to the use of different mathematical formulas (Pavoine et al., 2013). More
516 research on the comparative analysis of phylogenetic and functional patterns in ecology are
517 thus needed.

518 Overall, our results emphasize that the biological conclusions of a study are critically
519 dependent on the formula used to measure diversity, evenness and uniqueness. This is an
520 obvious consequence of our attempt to reduce multidimensional concepts, such as diversity,
521 evenness and uniqueness into single one-dimensional variation. It is well known for example
522 that different diversity indices, even the Shannon and the Gini-Simpson indices, can rank
523 communities differently (e.g., Patil and Taillie, 1982). Even with the Hill numbers, using
524 diversity profiles as defined above, the profiles of distinct communities can cross. This means
525 that, for example, up to a certain value of q , a community C_1 may be perceived as more
526 diverse than C_2 , while for higher values of q it is perceived as less diverse than C_2 (Leinster
527 and Cobbold, 2012).

528 Our results also highlight the potential and limits of this general family of indices
529 where species are replaced by features in diversity indices. We looked deeply at some
530 properties of these indices, especially at the interpretability of ratios between diversity indices
531 tackling notions of evenness and uniqueness. These results show that speaking generally of
532 'diversity', 'evenness' or 'uniqueness' is probably not enough. We have to go deeper inside the
533 meaning of the different measures. For example, the diversity calculated with Rao's quadratic
534 entropy is an expected distance between species pairs. In our framework, a diversity measure
535 might be an entropy (e.g. Shannon entropy) and/or a probability (e.g. Gini-Simpson index)
536 associated to features. The Hill numbers associated to Rao's quadratic entropy may have a
537 different interpretation from the Hill numbers associated to features, if Rao's quadratic
538 entropy is applied directly to a functional dissimilarity matrix, without the need to obtain a
539 functional tree. A myriad of species diversity indices have already been developed in the
540 literature. The consideration of functional and phylogenetic data opens the way to even more
541 numerous indices of functional and phylogenetic diversity. A detailed interpretation of the
542 very meaning of these metrics, in reference to traditional species diversity indices, is thus of

543 great importance for understanding what the indices are actually measuring and for their
544 sound biological interpretation.

545

546 **Acknowledgements**

547 We thank the referees for their positive and useful comments. Funding: This research did not
548 receive any specific grant from funding agencies in the public, commercial, or not-for-profit
549 sectors.

550

551 **Data Accessibility**

552 All data are publicly available: abundance and trait data from Ricotta et al. (2016, Appendix
553 B); phylogenetic tree from Durka and Michalski (2012); R scripts from package adiv
554 (Pavoine, 2018) and Appendixes E and F.

555

556 **Appendixes A-F. Supplementary data**

557 Supplementary data associated with this article can be found, in the online version, at [...]

558

559 **References**

560 Allen, B., Kon, M., Bar-Yam, Y., 2009. A new phylogenetic diversity measure generalizing
561 the Shannon index and its application to phyllostomid bats. *Am. Nat.* 174, 236–243.

562 Behara, M., Chawla, J.M.S., 1975. Generalized Gamma-Entropy. *Selecta Statistica Canadiana*
563 2, 15–38.

564 Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing
565 under dependency. *Ann. Stat.* 29, 1165–1188.

566 Botta-Dukat, Z., 2018. The generalized replication principle and the partitioning of functional
567 diversity into independent alpha and beta components. *Ecography* 41, 40–50.

568 Caccianiga, M., Luzzaro, A., Pierce, S., Ceriani, R.M., Cerabolini, B., 2006. The functional
569 basis of a primary succession resolved by CSR classification. *Oikos* 112, 10–20.

570 Chao, A., Chiu, C.-H., Jost, L., 2010. Phylogenetic diversity measures based on Hill numbers.
571 *Philos. Trans. R. Soc. B-Biol. Sci.* 365, 3599–3609.

572 Dalton, H., 1920. The Measurement of the Inequality of Incomes. *Econ. J.* 30, 348–361.

573 Daroczy, Z., 1970. Generalized information functions. *Information and Control* 16, 36–51.

574 Durka, W., Michalski, S.G., 2012. Daphne: a dated phylogeny of a large European flora for
575 phylogenetically informed ecological analyses. *Ecology* 93, 2297–2297.

576 Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–
577 10.

578 Faith, D.P., 2013. Biodiversity and evolutionary history: useful extensions of the PD
579 phylogenetic diversity assessment framework. *Ann. N.Y. Acad. Sci.* 1289, 69–89.

580 Faith, D.P., Lozupone, C.A., Nipperess, D., Knight, R., 2009. The cladistic basis for the
581 Phylogenetic Diversity (PD) measure links evolutionary features to environmental
582 gradients and supports broad applications of microbial ecology’s “Phylogenetic Beta
583 Diversity” framework. *Int. J. Mol. Sci.* 10, 4723–4741.

584 Gini, C., 1912. Variabilita e mutabilita. *Studi economicoaguridici delle facolta di*
585 *giurizprudenza dell, Universite di Cagliari III, Parte II.*

586 Grime, J.P., 1974. Vegetation classification by reference to strategies. *Nature* 250, 26–31.

587 Grime, J.P., 2001. *Plant Strategies, Vegetation Processes and Ecosystem Properties.* Wiley,
588 Chichester, UK.

589 Havrda, M., Charvat, F., 1967. Quantification method of classification processes: concept of
590 structural alpha-entropy. *Kybernetika* 3, 30–35.

591 Hill, M.O., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology*
592 54, 427–432.

593 Leinster, T., Cobbold, C.A., 2012. Measuring diversity: the importance of species similarity.
594 Ecology 93, 477–489.

595 Magurran, A.E., 2004. Measuring biological diversity. Blackwell Publishing, Oxford, U.K.

596 May, R.M., 1994. Conceptual aspects of the quantification of the extent of biological
597 diversity. Phil. Trans. R. Soc. Lond. 345, 13–20.

598 Mouchet, M., Guilhaumon, F., Villéger, S., Mason, N. W., Tomasini, J. A., Mouillot, D.,
599 2008. Towards a consensus for calculating dendrogram-based functional diversity
600 indices. Oikos 117, 794–800.

601 Nayak, T.K., 1985. On diversity measures based on entropy functions. Commun. Stat.-Theory
602 Methods 14, 203–215.

603 Patil, G.P., Taillie, C., 1982. Diversity as a concept and its measurement. J. Am. Stat. Assoc.
604 77, 548–561.

605 Pavoine, S., 2016. A guide through a family of phylogenetic dissimilarity measures among
606 sites. Oikos 125, 1719–1732.

607 Pavoine, S., 2018. adiv: Analysis of Diversity. R package version 1.2. [https://CRAN.R-](https://CRAN.R-project.org/package=adiv)
608 [project.org/package=adiv](https://CRAN.R-project.org/package=adiv)

609 Pavoine, S., Gasc, A., Bonsall, M.B., Mason, N.W.H., 2013. Correlations between
610 phylogenetic and functional diversity: mathematical artefacts or true ecological and
611 evolutionary processes? J. Veg. Sci. 24, 781–793.

612 Pavoine, S., Love, M., Bonsall, M.B., 2009a. Hierarchical partitioning of evolutionary and
613 ecological patterns in the organization of phylogenetically-structured species
614 assemblages: application to rockfish (genus: *Sebastes*) in the Southern California
615 Bight. Ecol. Lett. 12, 898–908.

616 Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S., Daniel, H., 2009b. On the challenge of
617 treating various types of variables: application for improving the measurement of
618 functional diversity. *Oikos* 118, 391–402.

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

621 Pielou, E.C., 1966. The measurement of diversity in different types of biological collections.
622 *J. Theor. Biol.* 13, 131–144.

623 Podani, J., Schmera, D., 2007. How should a dendrogram-based measure of functional
624 diversity function? A rejoinder to Petchey and Gaston. *Oikos* 116, 1427–1430.

625 Poos, M.S., Walker, S.C., Jackson, D.A., 2009. Functional-diversity indices can be driven by
626 methodological choices and species richness. *Ecology* 90, 341–347.

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

629 Rényi, A., 1960. On measures of entropy and information. *Proc. 4th Berkeley Symp. Math.*
630 *Stat. Probability* 1, 547–561.

631 Ricotta, C., Avena, G.C., 2002. On the information-theoretical meaning of Hill's parametric
632 evenness. *Acta Biotheor.* 50, 63–71.

633 Ricotta, C., Bacaro, G., Caccianiga, M., Cerabolini, B., Pavoine, S., 2018. A new method for
634 quantifying the phylogenetic redundancy of biological communities. *Oecologia* 186,
635 339–346.

636 Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B.E., Pavoine, S., 2016.
637 Measuring the functional redundancy of biological communities: a quantitative guide.
638 *Methods Ecol. Evol.* 7, 1386–1395.

639 Schmera, D., Eros, T., Podani, J., 2009. A measure for assessing functional diversity in
640 ecological communities. *Aquat. Ecol.* 43, 157–167.

- 641 Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical
642 Journal 27, 379–423, 623–656.
- 643 Sheldon, A.L., 1969. Equitability indices: dependence on the species count. Ecology 50, 466–
644 467.
- 645 Simpson, E.H., 1949 Measurement of diversity. Nature 163, 688.
- 646 Smith, B., Wilson, J.B., 1996. A consumer's guide to evenness measures. Oikos 76, 70–82.
- 647 Taillie, C., 1979. Species equitability: a comparative approach. In: Grassle, J.F., Patil, G.P.,
648 Smith, W.K., Taillie, C. (Eds.), Ecological Diversity in Theory and Practice.
649 International Cooperative Publishing House, Fairland, MD, pp. 51–62.
- 650 Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. 52,
651 480–487.
- 652 Webb, C.O., Ackerly, D.D., McPeck, M.A., Donoghue, M.J., 2002. Phylogenies and
653 community ecology. Annu. Rev. Ecol. Syst. 33, 475–505.
- 654

655 **Table 1**

656 From species to feature diversity indices.

Original Index	Species diversity*	Feature diversity*
Richness	$ric = \text{card}(t_T)$	$fric = \sum_{k \in b_T} L_k$
Shannon	$div_{Shannon} = -\sum_{i \in t_T} P_i \log(P_i)$	$fdiv_{Shannon} = -\sum_{k \in b_T} L_k p_k \log(p_k)$
Gini-Simpson	$div_{GS} = 1 - \sum_{i \in t_T} (P_i)^2$	$fdiv_{GS} = 1 - \sum_{k \in b_T} L_k (p_k)^2$
HCDT, $q \neq 1^\dagger$	${}^q div_{HCDT} = \left[1 - \sum_{i \in t_T} (P_i)^q \right] / (q-1)$	${}^q fdiv_{HCDT} = \left[1 - \sum_{k \in b_T} L_k (p_k)^q \right] / (q-1)$
Hill, $q \neq 1$	${}^q div_{Hill} = \left[\sum_{i \in t_T} (P_i)^q \right]^{\frac{1}{1-q}}$	${}^q fdiv_{Hill} = \left[\sum_{k \in b_T} L_k (p_k)^q \right]^{\frac{1}{1-q}}$
Hill, $q \rightarrow 1$	${}^1 div_{Hill} = \exp \left[-\sum_{i \in t_T} P_i \log(P_i) \right]$	${}^1 fdiv_{Hill} = \exp \left[-\sum_{k \in b_T} L_k p_k \log(p_k) \right]$
Rényi, $q \neq 1^\dagger$	${}^q div_{Rényi} = \frac{1}{1-q} \log \left[\sum_{i \in t_T} (P_i)^q \right]$	${}^q fdiv_{Rényi} = \frac{1}{1-q} \log \left[\sum_{k \in b_T} L_k (p_k)^q \right]$

657 *Most notations are similar as those used in the main text. However, to simplify the
658 equations, we use $p_k = a_k / \sum_{l \in b_T} L_l a_l$, the relative abundance of any feature (unit) on branch k ,
659 and $P_i = A_i / \sum_{j \in t_T} A_j$, the relative abundance of species i .

660 $^\dagger q$ tending to 1 leads to the Shannon index.

661

662 **Table 2**

663 Links between parametric indices of feature diversity and the richness, Gini-Simpson and

664 Shannon indices applied to feature abundance.

Index	$q = 0$	$q \rightarrow 1$	$q = 2$
HCDT	${}^0fdiv_{HCDT} = fric - 1$	${}^1fdiv_{HCDT} = fdiv_{Shannon}$	${}^2fdiv_{HCDT} = fdiv_{GS}$
Hill	${}^0fdiv_{Hill} = fric$	${}^1fdiv_{Hill} = \exp(fdiv_{Shannon})$	${}^2fdiv_{Hill} = 1 / (1 - fdiv_{GS})$
Rényi	${}^0fdiv_{Rényi} = \log(fric)$	${}^1fdiv_{Rényi} = fdiv_{Shannon}$	${}^2fdiv_{Rényi} = -\log(1 - fdiv_{GS})$

665

666 **Table 3**

667 From species evenness to potential feature evenness indices analyzed in the paper.

Original Index	Species evenness	Feature evenness*
Shannon	$eve_{Shannon} = \frac{-\sum_{i \in \mathcal{T}_r} P_i \log(P_i)}{\log(S)}$	$feve_{Shannon} = \frac{-\sum_{k \in \mathcal{B}_r} L_k p_k \log(p_k)}{\log\left(\sum_{i \in \mathcal{T}_r} h_i\right)}$
Gini-Simpson	$eve_{GS} = \frac{1 - \sum_{i \in \mathcal{T}_r} (P_i)^2}{1 - S^{-1}}$	$feve_{GS} = \frac{1 - \sum_{k \in \mathcal{B}_r} L_k (p_k)^2}{1 - \left(\sum_{i \in \mathcal{T}_r} h_i\right)^{-1}}$
HCDT, $q \neq 1^\dagger$	${}^q eve_{HCDT} = \frac{1 - \sum_{i \in \mathcal{T}_r} (P_i)^q}{1 - S^{1-q}}$	${}^q feve_{HCDT} = \frac{1 - \sum_{k \in \mathcal{B}_r} L_k (p_k)^q}{1 - \left(\sum_{i \in \mathcal{T}_r} h_i\right)^{1-q}}$
Hill, $q \neq 1$	${}^q eve_{Hill} = \frac{\left[\sum_{i \in \mathcal{T}_r} (P_i)^q\right]^{\frac{1}{1-q}}}{S}$	${}^q feve_{Hill} = \frac{\left[\sum_{k \in \mathcal{B}_r} L_k (p_k)^q\right]^{\frac{1}{1-q}}}{\sum_{i \in \mathcal{T}_r} h_i}$
Hill, $q \rightarrow 1$	${}^1 eve_{Hill} = \frac{\exp\left[-\sum_{i \in \mathcal{T}_r} P_i \log(P_i)\right]}{S}$	${}^1 feve_{Hill} = \frac{\exp\left[-\sum_{k \in \mathcal{B}_r} L_k p_k \log(p_k)\right]}{\sum_{i \in \mathcal{T}_r} h_i}$
Rényi, $q \neq 1^\dagger$	${}^q eve_{Rényi} = \frac{\frac{1}{1-q} \log\left[\sum_{i \in \mathcal{T}_r} (P_i)^q\right]}{\log(S)}$	${}^q feve_{Rényi} = \frac{\frac{1}{1-q} \log\left[\sum_{k \in \mathcal{B}_r} L_k (p_k)^q\right]}{\log\left(\sum_{i \in \mathcal{T}_r} h_i\right)}$

668 *Most notations are similar as those used in the main text. However, to simplify the
 669 equations, we use $p_k = a_k / \sum_{l \in \mathcal{B}_r} L_l a_l$, the relative abundance of any feature (unit) on branch k
 670 and $P_i = A_i / \sum_{j \in \mathcal{T}_r} A_j$, the relative abundance of species i .

671 $\dagger q$ tending to 1 leads to the Shannon index.

672

673 **Table 4**

674 Potential feature uniqueness indices analyzed in the paper.

Original Index	Feature uniqueness*
Richness	$funi_{richness} = \frac{\sum_{k \in b_T} L_k}{\sum_{i \in t_T} h_i}$
Shannon	$funi_{Shannon} = \frac{\sum_{k \in b_T} L_k p_k \log(p_k)}{\sum_{i \in t_T} h_i w_i \log(w_i)}$
Gini-Simpson	$funi_{GS} = \frac{1 - \sum_{k \in b_T} L_k (p_k)^2}{1 - \sum_{i \in t_T} h_i (w_i)^2}$
HCDT, $q \neq 1^\dagger$	${}^q funi_{HCDT} = \frac{\left[1 - \sum_{k \in b_T} L_k (p_k)^q \right]}{\left[1 - \sum_{i \in t_T} h_i (w_i)^q \right]}$
Hill, $q \neq 1$	${}^q funi_{Hill} = \left[\frac{\sum_{k \in b_T} L_k (p_k)^q}{\sum_{i \in t_T} h_i (w_i)^q} \right]^{\frac{1}{(1-q)}}$
Hill, $q \rightarrow 1$	${}^1 funi_{Hill} = \frac{\exp \left[\sum_{k \in b_T} L_k p_k \log(p_k) \right]}{\exp \left[\sum_{i \in t_T} h_i w_i \log(w_i) \right]}$
Rényi, $q \neq 1^\dagger$	${}^q funi_{Rényi} = \frac{\log \left[\sum_{k \in b_T} L_k (p_k)^q \right]}{\log \left[\sum_{i \in t_T} h_i (w_i)^q \right]}$

675 *Most notations are similar as those used in the main text. However, to simplify the
 676 equations, we use $p_k = a_k / \sum_{l \in b_T} L_l a_l$, the relative abundance of any feature (unit) on branch k ,
 677 and $w_i = A_i / \sum_{j \in t_T} h_j A_j$, the relative abundance of any feature on the branch that sustain
 678 species i in the star-shaped tree.

679 $\dagger q$ tending to 1 leads to the Shannon index.

680

681 **Table 5**

682 Properties fulfilled by the parametric indices of feature diversity, evenness and uniqueness

Property*	HCDT	Hill	Rényi
Diversity (^q <i>fdiv</i>)			
D1. Non-increasing function of q	Y	Y	Y
D2. Unaffected if a species is split into two similar subspecies supported by zero-length branches	Y	Y	Y
D3. Schur-concave†	Y	Y	Y
D4. Dependent of the measurement unit of branch length	Y	Y	Y
D5. Independent of q at the maximum	N	Y	Y
D6. Expressed as an effective number of features	N	Y	N
D7. Multiplied by N if the composition of a community is replicated N times as in Fig. 3	N	Y	N
D8. Multiplied by λ if the branch lengths are multiplied by λ ($\lambda > 0$)	N	Y	N
Evenness (^q <i>feve</i>)			
E1. Constrained to a 0-1 range for convenience	Y	Y	Y
E2. Reflects the evenness in the abundances of features (considering the amount of features the tree would have if species had not shared any branch since the root of the tree)	N	Y	Y
E3. Non-increasing function of q	N	Y	Y
E4. Unchanged if the composition of a community is replicated N times as in Fig. 3	N	Y	N
Uniqueness (^q <i>funi</i>)			
U1. Constrained to a 0-1 range for convenience	Y	Y	Y
U2. Increases with the amount of differences between species and is maximum when species are maximally dissimilar (sharing no branches on the tree, and being at a maximum distance from the root of the tree)	N	Y	Y
U3. Unchanged if the composition of a community is replicated N times as in Fig. 3	N	Y	N

683 * "Y", for "Yes", indicates that an index has the property, "N", for "No", that it does not.

684 Proofs can be found in the main text and illustrative examples in Appendix D.

685 † being increased when abundance is transferred from a feature to another strictly less

686 abundant feature

687

688 **Table 6**

689 Tests for differences, between successional stages (early, mid and late), in the level of species
 690 diversity, functional diversity and functional uniqueness. The tests use the quadratic entropy
 691 (Q) framework with the Gini-Simpson index (div_{GS}) as a special case, both transformed as
 692 equivalent numbers of species. We provide mean (s.d.) value taken by each index in each
 693 successional stage. Mean values associated with different letters were significantly different
 694 ($P < 0.05$; pairwise comparisons by Wilcoxon test). Italic letters indicate individually
 695 significant tests, whereas bold letters indicate that the test remained significant after P was
 696 controlled for false discovery rate using Benjamini and Yekutieli (2001) approach.

	Early	Mid	Late
${}^2div_{Hill} = 1/(1-div_{GS})$	10.514 ^a (3.676)	12.768 ^b (2.897)	10.152 ^a (3.565)
$Q^*=1/(1-Q)$	1.764 ^a (0.114)	1.613 ^b (0.069)	1.475 ^c (0.062)
$U^*=(1-div_{GS})/(1-Q)$	0.187 ^a (0.064)	0.133 ^b (0.030)	0.159 (0.047)

697

698 **Figure Legends**

699

700 **Fig. 1.** Illustration of the fact that, provided at least one branch on any tree T is shared by at
701 least two species, then the abundance of branch units cannot be even. Here species (tips of the
702 tree) are given equal abundance, a . The abundance associated with a branch is indicated
703 below this branch. It varies from a , to $3a$.

704

Single-column figure

705

706 **Fig. 2.** Illustration of a theoretical tree (a) modified so that species are independent with no
707 shared branches (the distance from tip to root is unchanged) (b), or independent and
708 maximally dissimilar (the distance from tip to root is elongated so that all species are equally
709 distant from the root node; broken lines indicate how much the branches were elongated) (c).
710 In (c) the new height of the tree may be equal to the maximum distance from tip to root
711 observed in the original tree (solution 1) or higher (solution 2).

712

Single-column figure

713

714 **Fig. 3.** Example of replicated data set with a theoretical tree and abundance data: (a) original
715 data set giving the composition of a theoretical community with five species; (b) data set
716 obtained by replicating 3 times the composition of the original community. With Hill index,
717 the values of uniqueness for the two data sets are equal (for example, ${}^0f_{uni_{Hill}}=0.756$,
718 ${}^1f_{uni_{Hill}}=0.748$, ${}^2f_{uni_{Hill}}=0.751$). With the Rényi and HCDT indices, the uniqueness of the
719 original data is different from that of the replicated data. For example, with the original data,
720 ${}^0f_{uni_{HCDT}}=0.735$, ${}^1f_{uni_{HCDT}}=0.870$, ${}^2f_{uni_{HCDT}}=0.951$; but with the replicated data,

721 ${}^0f_{uni_{HCDT}}=0.749$, ${}^1f_{uni_{HCDT}}=0.913$, ${}^2f_{uni_{HCDT}}=0.985$. With the original data, ${}^0f_{uni_{Rényi}}=0.889$,
722 ${}^1f_{uni_{Rényi}}=0.870$, ${}^2f_{uni_{Rényi}}=0.861$; but with the replicated data, ${}^0f_{uni_{Rényi}}=0.923$,
723 ${}^1f_{uni_{Rényi}}=0.913$, ${}^2f_{uni_{Rényi}}=0.909$.

724 *1.5-column figure*

725

726 **Fig. 4.** Box plots for feature indices applied to the three successional stages identified on the
727 foreland of the Rutor glacier and to (a) phylogenetic data, (b) trait data. Here we used the Hill
728 numbers with $q = 0$, $q = 1$ and $q = 2$ as indicated on the graphs. "E" stands for the early-
729 successional stage, "M" for the mid-successional stage and "L" for the late-successional stage.
730 For each index, box plots associated with different letters differed significantly in their
731 median ($P < 0.05$; pairwise comparisons by Wilcoxon test). Italic letters indicate individually
732 significant tests, whereas bold letters indicate that the test was also significant after P was
733 controlled for false discovery rate using Benjamini and Yekutieli (2001) approach. Removing
734 italic letters thus indicates which tests remained significant after the correction for multiple
735 tests.

736 *2-column figure*

737

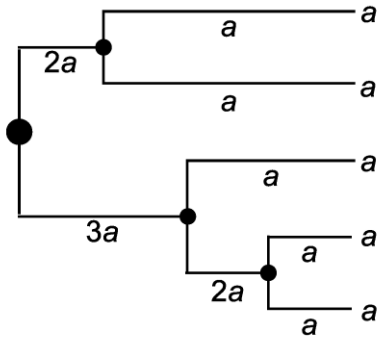
738 **Fig. 5.** Box plots for species indices applied to the three successional stages identified on the
739 foreland of the Rutor glacier. Here we used the Hill numbers with $q = 0$, $q = 1$ and $q = 2$ as
740 indicated on the graphs. "E" stands for the early-successional stage, "M" for the mid-
741 successional stage and "L" for the late-successional stage. As in Figure 4, for each index, box
742 plots associated with different letters differed significantly in their median ($P > 0.05$; pairwise
743 comparisons by Wilcoxon test corrected for potential ties). We used italic letters to indicate
744 individually significant tests; none of the tests was significant after P was controlled for false
745 discovery rate using Benjamini and Yekutieli (2001) approach.

746

single-column figure

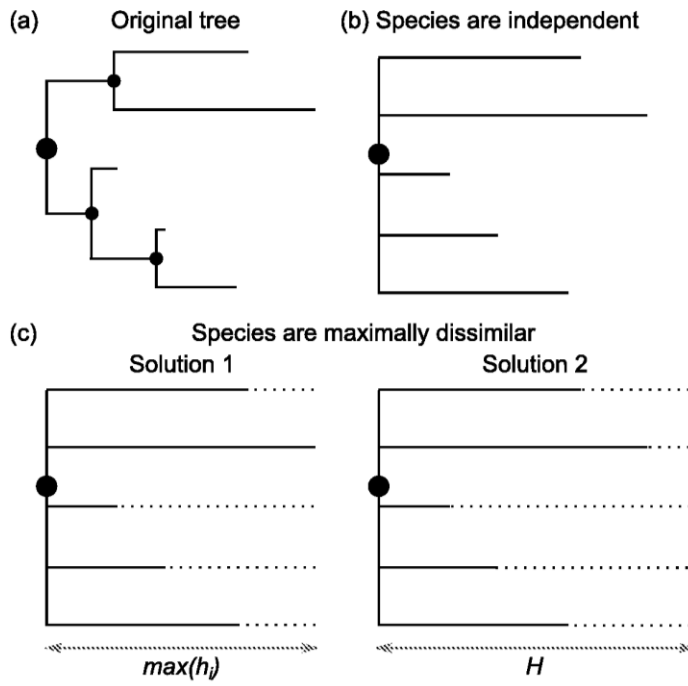
747

748 Figure 1



749

750

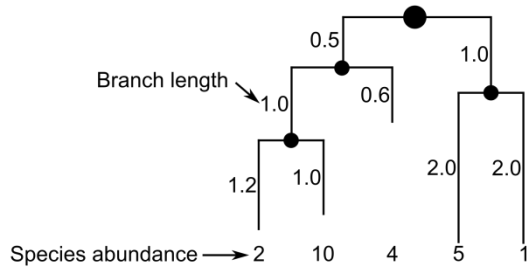


752

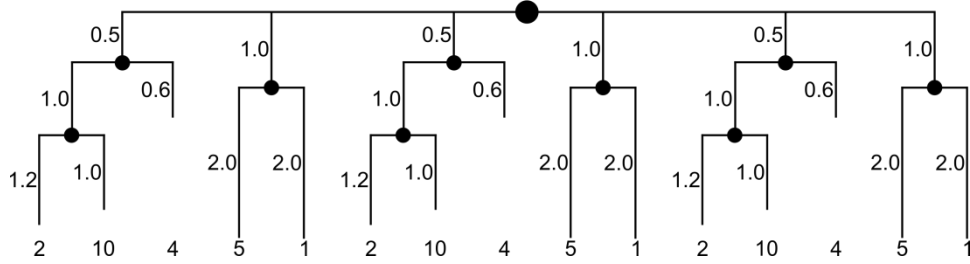
753

754 Figure 3

(a) Original data



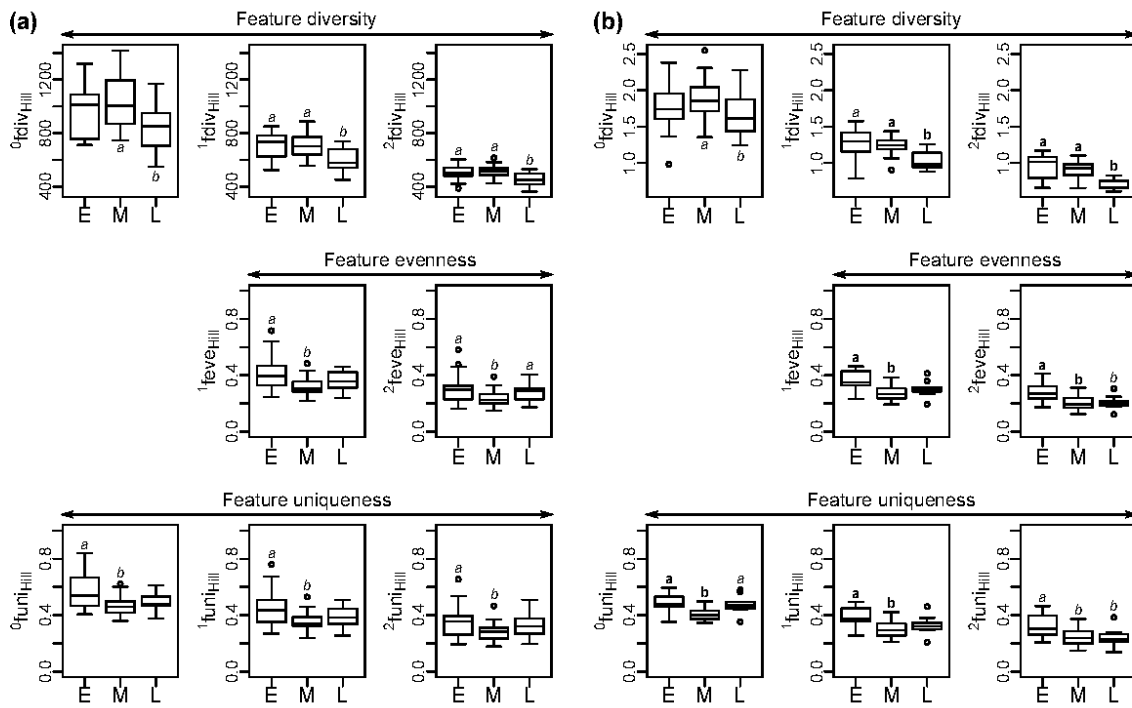
(b) Replicated data (3 times)



755

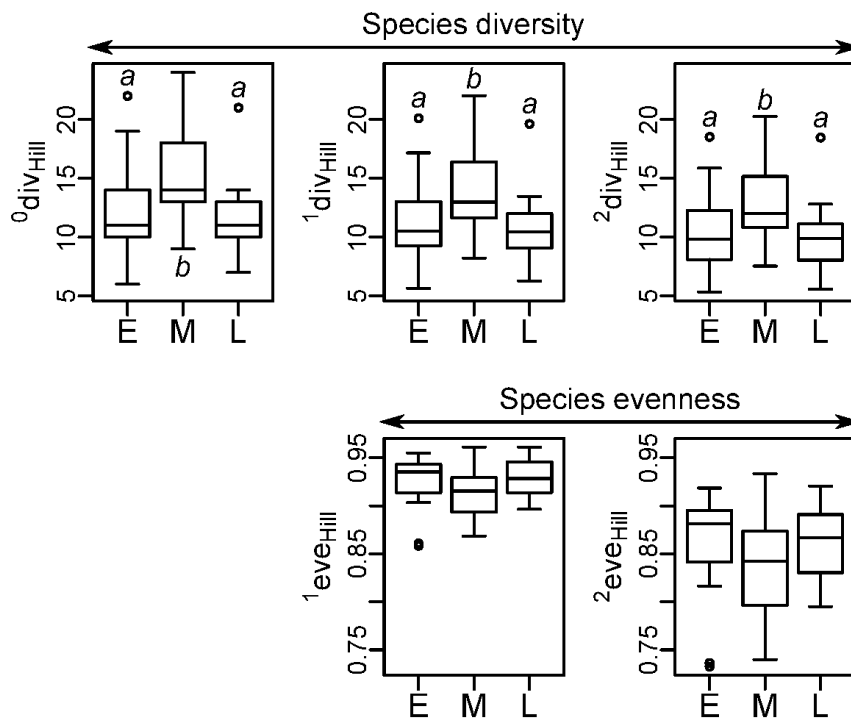
756

757 Figure 4



758

759



761

762

763

Appendix A. Proofs

We consider here the same notations as in the main text.

1. If the smallest sum of branch lengths between a species and the root of the tree is at least equal to 1, then $a_k / \sum_{k \in b_T} L_k a_k$ is always lower than or equal to unity.

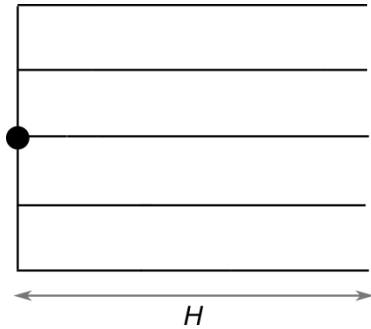
Let $C(i, Root)$ be the set of branches on the shortest path from a species to the root of the tree.

$$\frac{a_k}{\sum_{k \in b_T} L_k a_k} = \frac{\sum_{i \in t_k} A_i}{\sum_{k \in b_T} L_k \sum_{i \in t_k} A_i} = \frac{\sum_{i \in t_k} A_i}{\sum_{i \in t_T} A_i \left(\sum_{k \in C(i, Root)} L_k \right)} \quad (\text{eq. A.1})$$

Let $h_i = \sum_{k \in C(i, Root)} L_k$ be the smallest sum of branch lengths between a species i and the root of the tree. If $h_i \geq 1$, then eq. A.1 is ≤ 1 .

2. If the weak restriction on branch lengths was not applied, ${}^q fdiv_{HCDT}$ and ${}^q fdiv_{Rényi}$ could have negative values, while ${}^q fdiv_{Hill}$ would still be nonnegative and meaningful

As an example, consider a theoretical case study of a star-shaped phylogeny T_{star} with S species as tips and branch lengths all equal to H :



Consider also that the species have even abundance, say equal to 1: abundance distribution $E=(1 \dots 1)$. In that case,

$${}^q fdiv_{HCDT}(T_{star}, E) = \left[1 - (S \times H)^{1-q} \right] / (q-1)$$

$${}^q fdiv_{Rényi}(T_{star}, E) = \log(S \times H)$$

If for example, $H=0.01$, $S=2$, and $q=2$, then, ${}^q fdiv_{HCDT}=-49$ and ${}^q fdiv_{Rényi}=-3.91$. As for ${}^q fdiv_{Hill}$, multiplying all branch lengths of a tree by a constant leads to ${}^q fdiv_{Hill}$ multiplied by this constant.

3. ${}^q fdiv_{Hill}$ provides an effective number of features

Consider a real community where species are related according to a defined rooted phylogenetic tree, with the notations used in the main text (T is the tree, k is a branch of the tree, L_k the length of branch k , b_T the set of branches on the phylogeny, a_k the abundance associated with branch k , A the set of species' abundances, q a parameter). Consider a theoretical community where species have even abundance and are independently related to a theoretical ancestor (star-shaped phylogenetic tree, T_{star}). Let e be the abundance of any of these species and E the distribution of abundance of all branches (each branch is also associated with an abundance of e). Let D be the number of features in the star-shaped phylogeny ($D = \text{sum of branch lengths on the star-shaped phylogeny}$). If the phylogenetic diversity of the communities is measured by ${}^q fdiv_{HCDT}$, then the phylogenetic diversity of the real community would be

$${}^q fdiv_{HCDT}(T, A) = \left[1 - \sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right] / (q-1)$$

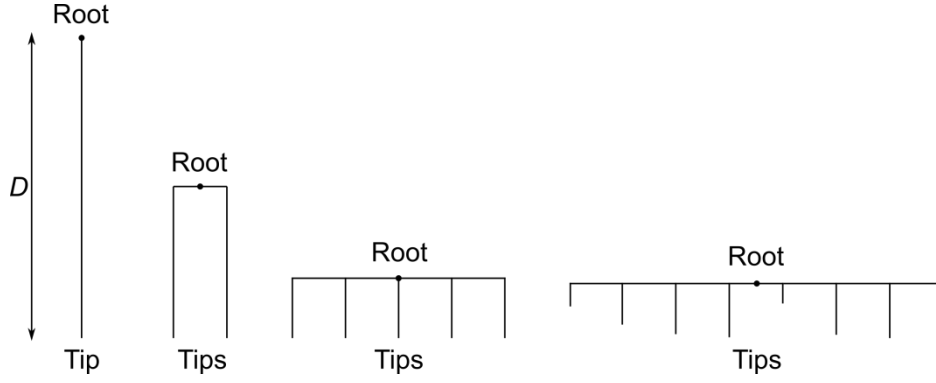
and that of the theoretical community would be

$${}^q fdiv_{HCDT}(T_{star}, E) = \left[1 - D \left(\frac{e}{De} \right)^q \right] / (q-1) = [1 - D^{1-q}] / (q-1)$$

For ${}^q fdiv_{HCDT}(T, A)$ to be equal to ${}^q fdiv_{HCDT}(T_{star}, E)$, then D must be equal to

$$D = \left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right]^{\frac{1}{1-q}} = {}^q fdiv_{Hill}(T, A)$$

The star-shaped phylogeny may have different shapes and different numbers of species, provided the number of features (sum of branch lengths) is D and the features have equal abundances. For example, let's consider the data set in Fig. 3a of the main text. The number of features in this data set is ${}^0 fdiv_{Hill} = 11.300$. The value of ${}^2 fdiv_{Hill}$ for this data set is 6.909. This means that a community with 6.909 features with even abundances would have the same value of ${}^2 fdiv_{HCDT}$ as that of the community described in Fig. 3a. There are multiple ways of having D features with even abundances, including those displayed below where each tips is considered to have an abundance of e :



4. When $q \rightarrow +\infty$, ${}^q feve_{Hill}$ and ${}^q feve_{Renyi}$ tend to their minimum while ${}^q feve_{HCDT}$ tends to 1.

We consider here species with positive abundance and their phylogenetic connections. Species with zero abundance are discarded (see e.g. Leinster and Cobbold 2012). When $q \rightarrow +\infty$, the Hill index applied to species tends to 1 divided by the highest relative abundance of a species (Hill, 1973). ${}^q div_{Hill}$ thus tends to 1 divided by the highest relative abundance of a species and ${}^q eve_{Hill}$ (species evenness: ratio of ${}^q div_{Hill}$ to its maximum) tends to 1 divided by the product of the highest relative abundance and the richness ($S =$ number of species). The minimum value for the richness times the highest relative abundance is 1. Indeed, if a community has only 1 species its relative abundance is 1; if a community has 2 species, the minimum relative abundance of the most abundant species is 0.5; more generally, if a community has S species, the minimum relative abundance of the most abundant species is $1/S$. Whatever the number of species in the community, the maximum relative abundance of the most abundant species tends to 1. The evenness of a community with one species only is 1. As soon as the community contains more than 1 species and the abundances are not all equal, when $q \rightarrow +\infty$, ${}^q eve_{Hill}$ takes thus values bounded in $]1/S; 1[$.

Extending the notion of species evenness to that of feature evenness, we developed index ${}^q feve_{Hill}$:

$${}^q feve_{Hill} = \frac{\left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right]^{\frac{1}{(1-q)}}}{\sum_{i \in t_T} h_i}$$

where h_i is the distance from species i to the root of the tree (as in the main text T is the tree, k is a branch of the tree, L_k the length of branch k , b_T the set of branches on the phylogeny, t_T the set of tips (species), a_k the abundance associated with branch k , q a parameter). The numerator of ${}^q feve_{Hill}$ is ${}^q fdiv_{Hill}$. It can be re-written as:

$$\left[\sum_{k \in b_T} L_k \left(\frac{a_k}{\sum_{l \in b_T} L_l a_l} \right)^q \right]^{\frac{1}{(1-q)}} = \left[\sum_{k \in b_T} \frac{L_k a_k}{\sum_{l \in b_T} L_l a_l} \left(\frac{\sum_{l \in b_T} L_l a_l}{a_k} \right)^{1-q} \right]^{\frac{1}{(1-q)}}$$

which shows that ${}^q fdiv_{Hill}$ is the weighted generalized mean of values $\sum_{l \in b_T} L_l a_l / a_k$, where the weights are the $w_k = L_k a_k / \sum_{l \in b_T} L_l a_l$. According to the known limits of the generalized mean (Hardy et al., 1952, p. 15),

$${}^q fdiv_{Hill} \xrightarrow{q \rightarrow +\infty} \min_{k \in b_T} \left(\frac{\sum_{l \in b_T} L_l a_l}{a_k} \right)$$

When $q \rightarrow +\infty$, ${}^q fdiv_{Hill}$ thus tends to 1 divided by the highest relative abundance of a feature. If all features have even abundance, then ${}^q feve_{Hill}=1$. Otherwise, when $q \rightarrow +\infty$, ${}^q feve_{Hill}$ takes thus values bounded in $]1/\sum_{i \in t_T} h_i ; 1[$.

By definition, ${}^q feve_{Renyi}$ tends to the ratio of the logarithm of 1 divided by the highest relative abundance to the logarithm of the maximum feature richness. Applying the same reasoning as above shows that, if all features have even abundance, then ${}^q feve_{Hill}=1$, otherwise, when $q \rightarrow +\infty$, ${}^q feve_{Renyi}$ can take values bounded in $]0; 1[$. Indeed,

$${}^q feve_{Renyi} \xrightarrow{q \rightarrow +\infty} \frac{\ln(\frac{1}{p_{max}})}{\ln(\sum_{i \in t_T} h_i)}$$

where p_{max} is the relative abundance of the most abundant feature. For a given value of S , the minimum possible value for p_{max} is $1/\sum_{i \in t_T} h_i$. This implies that the maximum possible value for $\ln(\frac{1}{p_{max}})/\ln(\sum_{i \in t_T} h_i)$ is 1. High values for p_{max} can be expressed as $1-\varepsilon$, with ε close to zero. For high values of p_{max} , $\ln(\frac{1}{p_{max}})$ is thus close to zero.

However, when $q \rightarrow +\infty$, ${}^q feve_{HCDT}$ tends to 1. Indeed,

$${}^q feve_{HCDT} = \frac{\left[1 - \sum_{k \in b_T} L_k (p_k)^q \right]}{\left[1 - \left(\sum_{i \in t_T} h_i \right)^{1-q} \right]} = \frac{\left[1 - \sum_{k \in b_T} L_k (p_k)^q \right]}{\left[1 - \left(\sum_{i \in t_T} h_i \right) \left(\frac{1}{\sum_{i \in t_T} h_i} \right)^q \right]}$$

Both the numerator and the denominator of ${}^q feve_{HCDT}$ tend to 1 if $q \rightarrow +\infty$, given that a^q tends to 0 if $q \rightarrow +\infty$ and $0 \leq a < 1$.

The same reasoning holds if $\sum_{i \in t_T} h_i$ is replaced, in the above equations, with $S \times H$ as also proposed in the main text (H is the maximum possible height of the tree).

5. The ratio of any Schur-concave diversity index applied to branch units of a tree to the same index applied to the tree modified so that species are independent lies between 0 and 1.

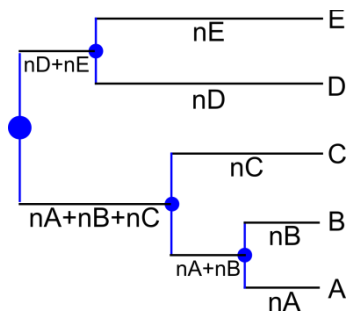
Consider a (phylogenetic) tree and the notations used in the main text. t_k is the set of species that descend from branch k . The length of this branch is L_k . The framework considered in this

paper assumes that there are L_k theoretical features (or branch units) that have a relative abundance of $a_k / \sum_{l \in b_r} L_l a_l = \sum_{i \in t_k} A_i / \sum_{l \in b_r} L_l a_l$. If the species were maximally dissimilar then there would have been not one branch to sustain all these species but one branch for each of these descending species in t_k . Instead of having L_k features of relative abundance $a_k / \sum_{l \in b_r} L_l a_l$, there would have been $n_k = \text{card}(t_k)$ branches of length L_k . The L_k features which would support species i would have a relative abundance of $A_i / \sum_{l \in b_r} L_l a_l$. According to Dalton's (1920) principle of transfers, this transfer of abundance from a single branch to several branches with zero abundances increases Schur-concave diversity indices.

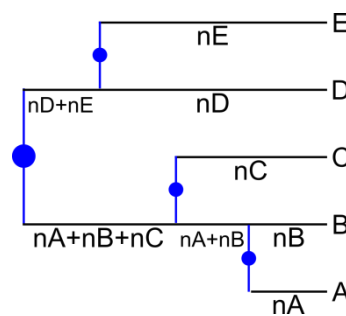
□

An example is given below:

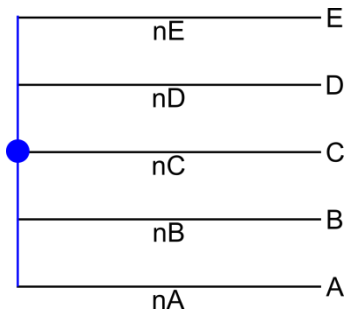
Consider five species, named A, B, C, D and E, and their abundances $n_A, n_B, n_C, n_D,$ and n_E , respectively. Consider that the five species can be placed at the tips of the following tree, which describes their similarities:



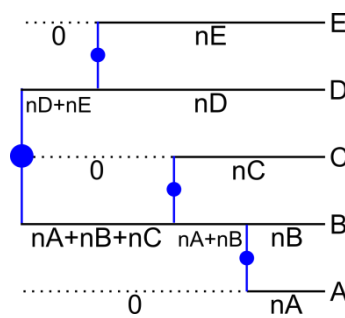
Let H be the height of the tree. We indicated above the abundance associated with each branch of the tree (equal to the summed abundance of its descending species). Because the tree is ultrametric, the total abundance of the features is $H \cdot (n_A + n_B + n_C + n_D + n_E)$. We now reorganize the way the tree is displayed to simplify the demonstration. This reorganization does not change the topology nor the branch lengths:



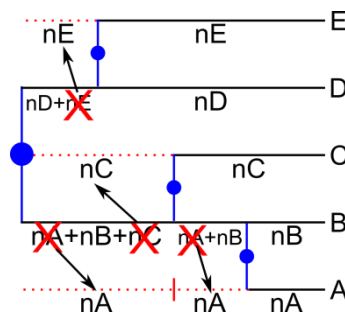
If the species were maximally dissimilar; they would have emerged from the root of the tree, lying thus on the tips of a star-shaped tree, as in the following graph:



The star-shaped tree above has $5 \times H$ features or branch units, while the original tree had less ($3.5 \times H$) features because some were shared between two or more species. We represent below, on the original tree, the additional $1.5 \times H$ features using black broken lines and indicate that their abundance is zero.



Measuring diversity from the star-shaped tree compared to measuring it from the original tree, with the framework developed in the main text, corresponds to transferring part of the abundance of the original features to these zero abundance features, as shown below:



According to Dalton's (1920) principle of transfers, this transfer of abundance from features with positive abundance to features with zero abundance increases feature-based Schur-concave diversity indices.

References:

Dalton, H., 1920. The Measurement of the Inequality of Incomes. *Econ. J.* 30, 348–361.
 Hardy, G.H., Littlewood, J.E., Pólya, G., 1952. *Inequalities*. Second edition. Cambridge University Press, Cambridge.
 Leinster, T., Cobbold, C.A., 2012. Measuring diversity: the importance of species similarity. *Ecology* 93, 477–489.

Appendix B. The value of evenness and uniqueness indices should not change if the composition of a community is replicated

1. Evenness indices

Let t_T be a set of species and $A = \{A_i\}_{i \in t_T}$ their distribution of abundance. Consider that there are S species ($\text{card}(t_T) = S$). Then, if applied to species, the parametric indices used in the main text are

$${}^q \text{div}_{HCDT}(A) = \left[1 - \sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right] / [q-1]$$

$${}^q \text{div}_{Hill}(A) = \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{(1-q)}}$$

$${}^q \text{div}_{Rényi}(A) = \frac{1}{1-q} \log \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]$$

Corresponding evenness indices could thus be (ratio of the diversity indices to their maximum):

$${}^q \text{eve}_{HCDT}(A) = \left[1 - \sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right] / [1 - S^{1-q}]$$

$${}^q \text{eve}_{Hill}(A) = \frac{1}{S} \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{(1-q)}}$$

$${}^q \text{eve}_{Rényi}(A) = \frac{1}{\log(S)} \frac{1}{1-q} \log \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]$$

Consider that we replicate the species composition N times. This will lead to $N \times S$ species. Let A_N be the new distribution of abundance for the $N \times S$ species. The following equations show that only for the Hill index, the evenness of A is equal to the evenness of A_N :

$${}^q eve_{HCDT}(A_N) = \left[1 - N \sum_{i \in t_T} \left(\frac{A_i}{N \sum_{i \in t_T} A_i} \right)^q \right] / \left[1 - (NS)^{1-q} \right]$$

$${}^q eve_{HCDT}(A_N) = \left[1 - N^{1-q} \sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right] / \left[1 - (NS)^{1-q} \right]$$

$${}^q eve_{Hill}(A_N) = \frac{1}{NS} \left[N \sum_{i \in t_T} \left(\frac{A_i}{N \sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}} = \frac{1}{NS} \left[N^{1-q} \sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}}$$

$${}^q eve_{Hill}(A_N) = \frac{N^{1-q}}{NS} \left[\sum_{i \in t_T} \left(\frac{A_i}{N \sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}} = \frac{1}{S} \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}} = {}^q eve_{Hill}(A)$$

$${}^q eve_{Rényi}(A_N) = \frac{1}{\log(NS)} \frac{1}{1-q} \log \left[N \sum_{i \in t_T} \left(\frac{A_i}{N \sum_{i \in t_T} A_i} \right)^q \right] = \frac{1}{\log(NS)} \frac{1}{1-q} \log \left[N^{1-q} \sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]$$

$${}^q eve_{Rényi}(A_N) = \frac{1}{\log(NS)} \log \left\{ N \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}} \right\} = \frac{\log(N) + \log \left\{ \left[\sum_{i \in t_T} \left(\frac{A_i}{\sum_{i \in t_T} A_i} \right)^q \right]^{\frac{1}{1-q}} \right\}}{\log(N) + \log(S)}$$

1. Uniqueness indices

Consider a rooted tree T with S species as tips and K branches. Let t_T be the set of species and $A = \{A_i\}_{i \in t_T}$ their distribution of abundance. As in the main text, t_k is the set of species

descending from branch k ; b_T is the set of branches in the tree T ; L_k is the length of branch k ; a_k is the sum of abundances for all species descending from branch k ($a_k = \sum_{i \in t_k} A_i$).

Consider that the species and their abundance are replicated N times, leading to the distribution of abundance named A_N . Consider that their associated tree is also replicated N times and the N resulting trees are connected by their root (see the example in Fig. 3 of the main text), leading to a new tree named T_N . Then the parametric uniqueness indices given in Table 4 of the main text are modified as follows:

$$\begin{aligned}
{}^q f_{uni_{HCDT}}(T_N, A_N) &= \frac{\left[1 - N \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} N \times L_k a_k \right)^q \right]}{\left[1 - N \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} N \times h_i A_i \right)^q \right]} \\
{}^q f_{uni_{HCDT}}(T_N, A_N) &= \frac{\left[1 - N^{1-q} \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q \right]}{\left[1 - N^{1-q} \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q \right]} \neq {}^q f_{uni_{HCDT}}(T, A) \\
{}^q f_{uni_{Hill}}(T_N, A_N) &= \frac{\left[\frac{N \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} N \times L_k a_k \right)^q}{N \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} N \times h_i A_i \right)^q} \right]^{\frac{1}{(1-q)}}}{\left[\frac{N^{1-q} \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q}{N^{1-q} \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q} \right]^{\frac{1}{(1-q)}}} \\
{}^q f_{uni_{Hill}}(T_N, A_N) &= \frac{\left[\frac{\sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q}{\sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q} \right]^{\frac{1}{(1-q)}}}{\left[\frac{\sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q}{\sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q} \right]^{\frac{1}{(1-q)}}} = {}^q f_{uni_{Hill}}(T, A) \\
{}^q f_{uni_{Rényi}}(T, A) &= \frac{\log \left[\sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q \right]}{\log \left[\sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q \right]}
\end{aligned}$$

$${}^q \text{funi}_{\text{Rényi}}(T_N, A_N) = \frac{\log \left[N \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} N \times L_k a_k \right)^q \right]}{\log \left[N \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} N \times h_i A_i \right)^q \right]} = \frac{\log \left[N^{1-q} \times \sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q \right]}{\log \left[N^{1-q} \times \sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q \right]}$$

$${}^q \text{funi}_{\text{Rényi}}(T_N, A_N) = \frac{\log \left[\sum_{k \in b_T} L_k \left(a_k / \sum_{k \in b_T} L_k a_k \right)^q \right] + (1-q) \log(N)}{\log \left[\sum_{i \in t_T} h_i \left(A_i / \sum_{i \in t_T} h_i A_i \right)^q \right] + (1-q) \log(N)} \neq {}^q \text{funi}_{\text{Rényi}}(T, A)$$

Let d_{ij} be the dissimilarity between species i and j , bounded between 0 and 1. The ratio of Rao index to Simpson index has the following formula:

$$U = \frac{\sum_{i \in t_T} \sum_{j \in t_T} \frac{A_i}{\sum_{k \in t_T} A_k} \frac{A_j}{\sum_{k \in t_T} A_k} d_{ij}}{1 - \sum_{j \in t_T} \left(\frac{A_j}{\sum_{k \in t_T} A_k} \right)^2}$$

Consider that the species and their abundance are replicated N times. Consider that the replicated species are maximally dissimilar: if s' , s'' , s''' are replicated species of s , then $d_{ss'} = d_{ss''} = d_{ss'''} = 1$. The ratio of Rao index to Simpson index for this replicated community is:

$$U_R = \frac{N \sum_{i \in t_T} \sum_{j \in t_T} \frac{A_i}{\sum_{k \in t_T} A_k} \frac{A_j}{\sum_{k \in t_T} A_k} d_{ij} + N(N-1) \sum_{i \in t_T} \sum_{j \in t_T} \frac{A_i}{\sum_{k \in t_T} A_k} \frac{A_j}{\sum_{k \in t_T} A_k}}{1 - N \sum_{j \in t_T} \left(\frac{A_j}{\sum_{k \in t_T} A_k} \right)^2}$$

$$U_R = \frac{\frac{1}{N} \sum_{i \in t_T} \sum_{j \in t_T} \frac{A_i}{\sum_{k \in t_T} A_k} \frac{A_j}{\sum_{k \in t_T} A_k} d_{ij} + \frac{N-1}{N}}{1 - \frac{1}{N} \sum_{j \in t_T} \left(\frac{A_j}{\sum_{k \in t_T} A_k} \right)^2}$$

$$U_R = \frac{\sum_{i \in t_T} \sum_{j \in t_T} \frac{A_i}{\sum_{k \in t_T} A_k} \frac{A_j}{\sum_{k \in t_T} A_k} d_{ij} + N-1}{N - \sum_{j \in t_T} \left(\frac{A_j}{\sum_{k \in t_T} A_k} \right)^2}$$

Our modified uniqueness index is

$$U^* = \frac{\sum_{j \in t_r} \left(\frac{A_j}{\sum_{k \in t_r} A_k} \right)^2}{1 - \sum_{i \in t_r} \sum_{j \in t_r} \frac{A_i}{\sum_{k \in t_r} A_k} \frac{A_j}{\sum_{k \in t_r} A_k} d_{ij}}$$

Its value for the replicated community is

$$U_R^* = \frac{\frac{1}{N} \sum_{j \in t_r} \left(\frac{A_j}{\sum_{k \in t_r} A_k} \right)^2}{1 - \frac{1}{N} \sum_{i \in t_r} \sum_{j \in t_r} \frac{A_i}{\sum_{k \in t_r} A_k} \frac{A_j}{\sum_{k \in t_r} A_k} d_{ij} - \frac{N-1}{N}}$$

$$U_R^* = \frac{\frac{1}{N} \sum_{j \in t_r} \left(\frac{A_j}{\sum_{k \in t_r} A_k} \right)^2}{\frac{1}{N} - \frac{1}{N} \sum_{i \in t_r} \sum_{j \in t_r} \frac{A_i}{\sum_{k \in t_r} A_k} \frac{A_j}{\sum_{k \in t_r} A_k} d_{ij}}$$

$$U_R^* = \frac{\sum_{j \in t_r} \left(\frac{A_j}{\sum_{k \in t_r} A_k} \right)^2}{1 - \sum_{i \in t_r} \sum_{j \in t_r} \frac{A_i}{\sum_{k \in t_r} A_k} \frac{A_j}{\sum_{k \in t_r} A_k} d_{ij}} = U^*$$

Appendix C. Additional figures on the analysis of the case study.

In the graphs below, blue is for the early successional stage, black for the mid successional stage and red for the late successional stage

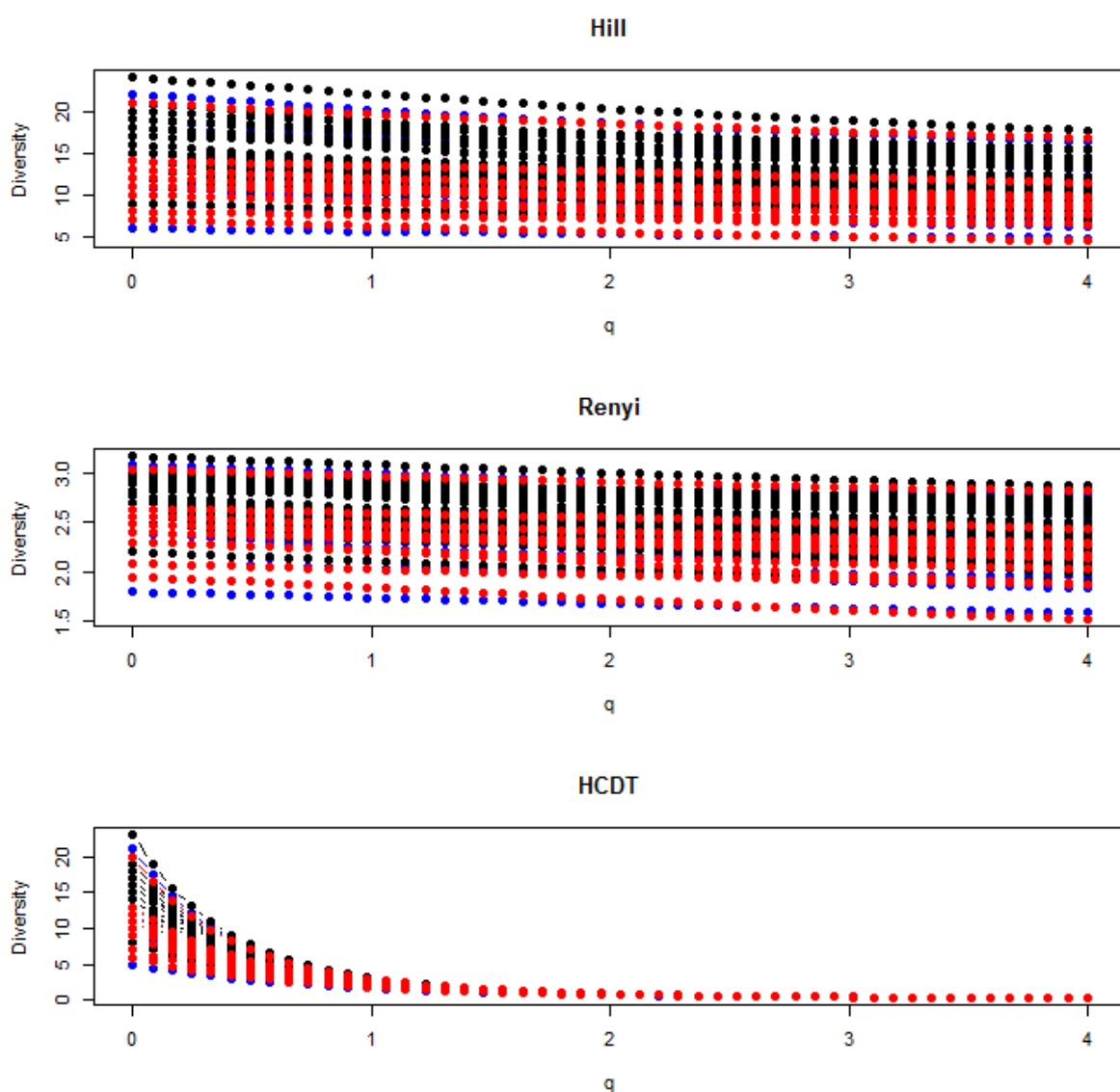


Fig. C-1. Species diversity profile for each of the 59 plots-Note that here the scale changes from a graph to another.

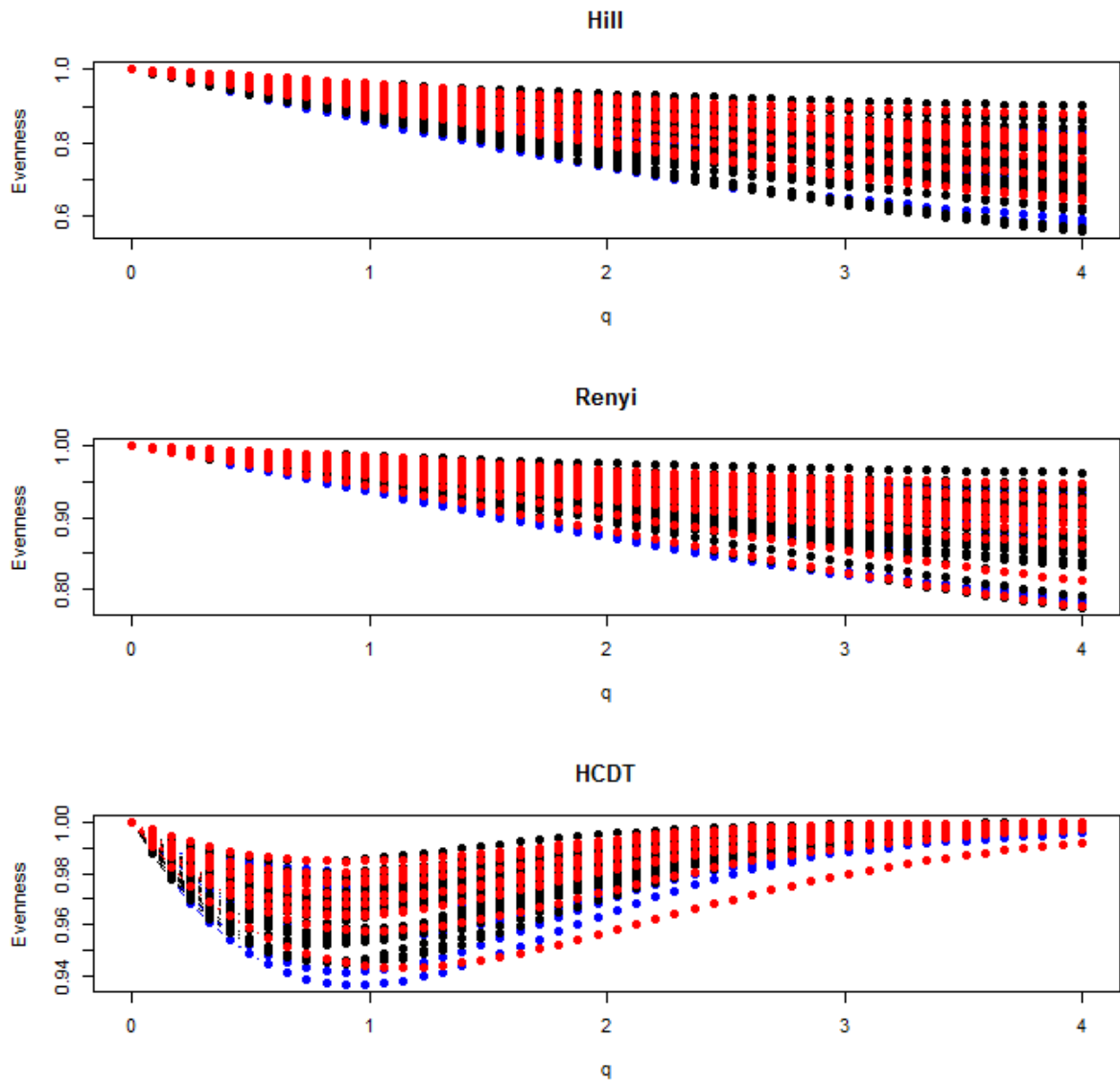


Fig. C-2. Species evenness profile for each of the 59 plots -Note that here the scale changes from a graph to another.

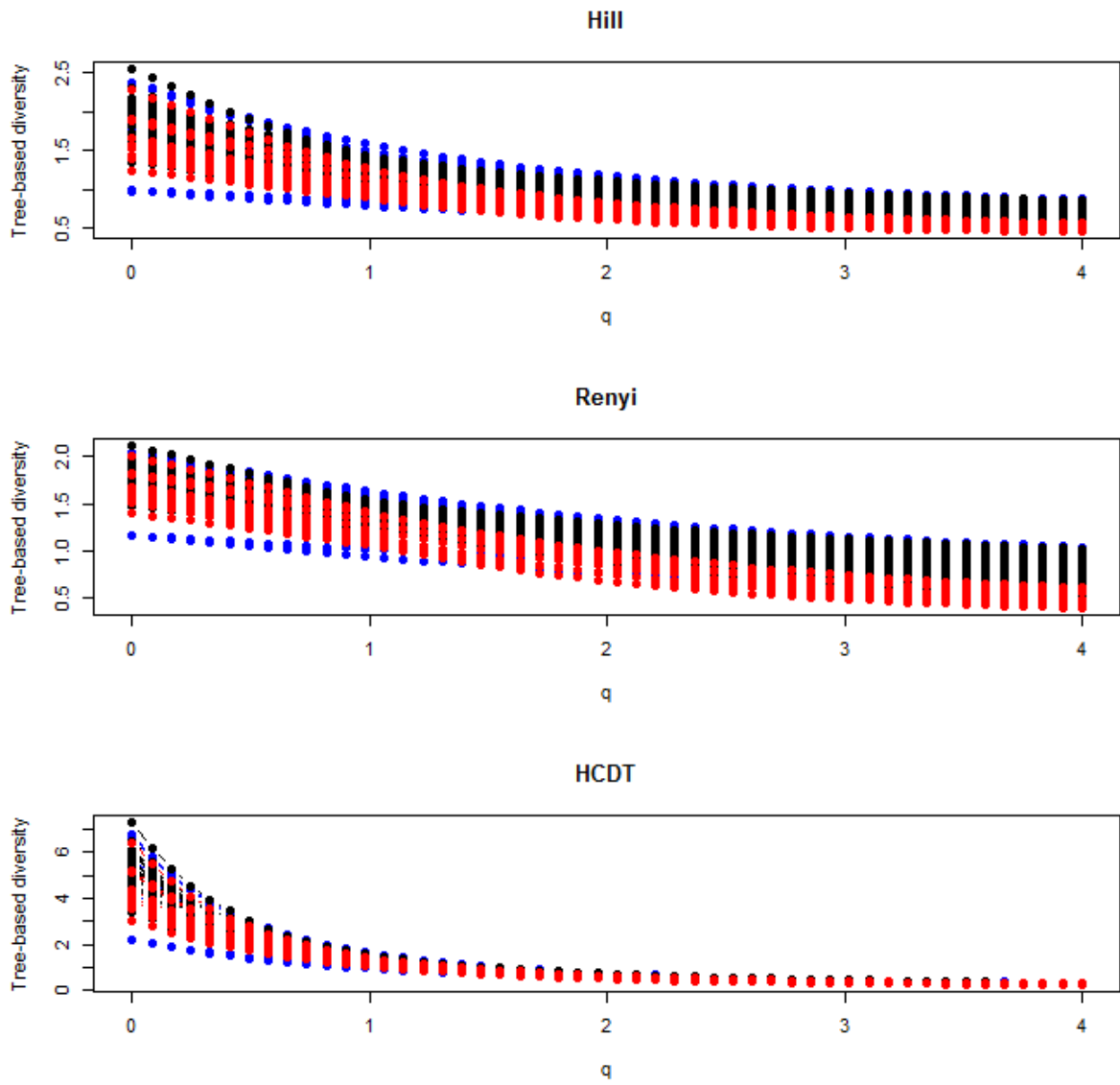


Fig. C-3. Functional feature diversity profile for each of the 59 plots. -Note that here the scale changes from a graph to another.

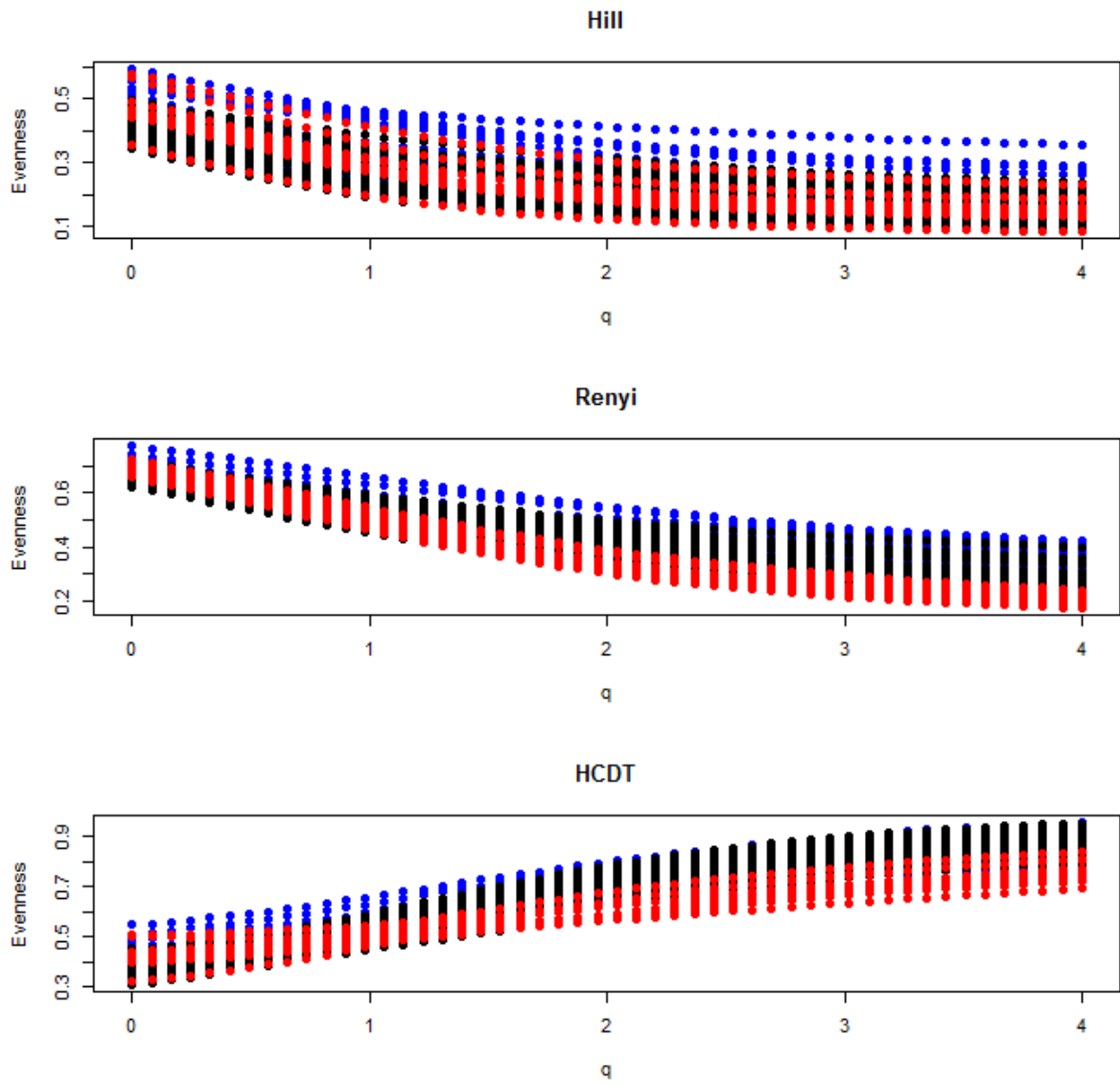


Fig. C-4. Functional feature evenness profile for each of the 59 plots -Note that here the scale changes from a graph to another.

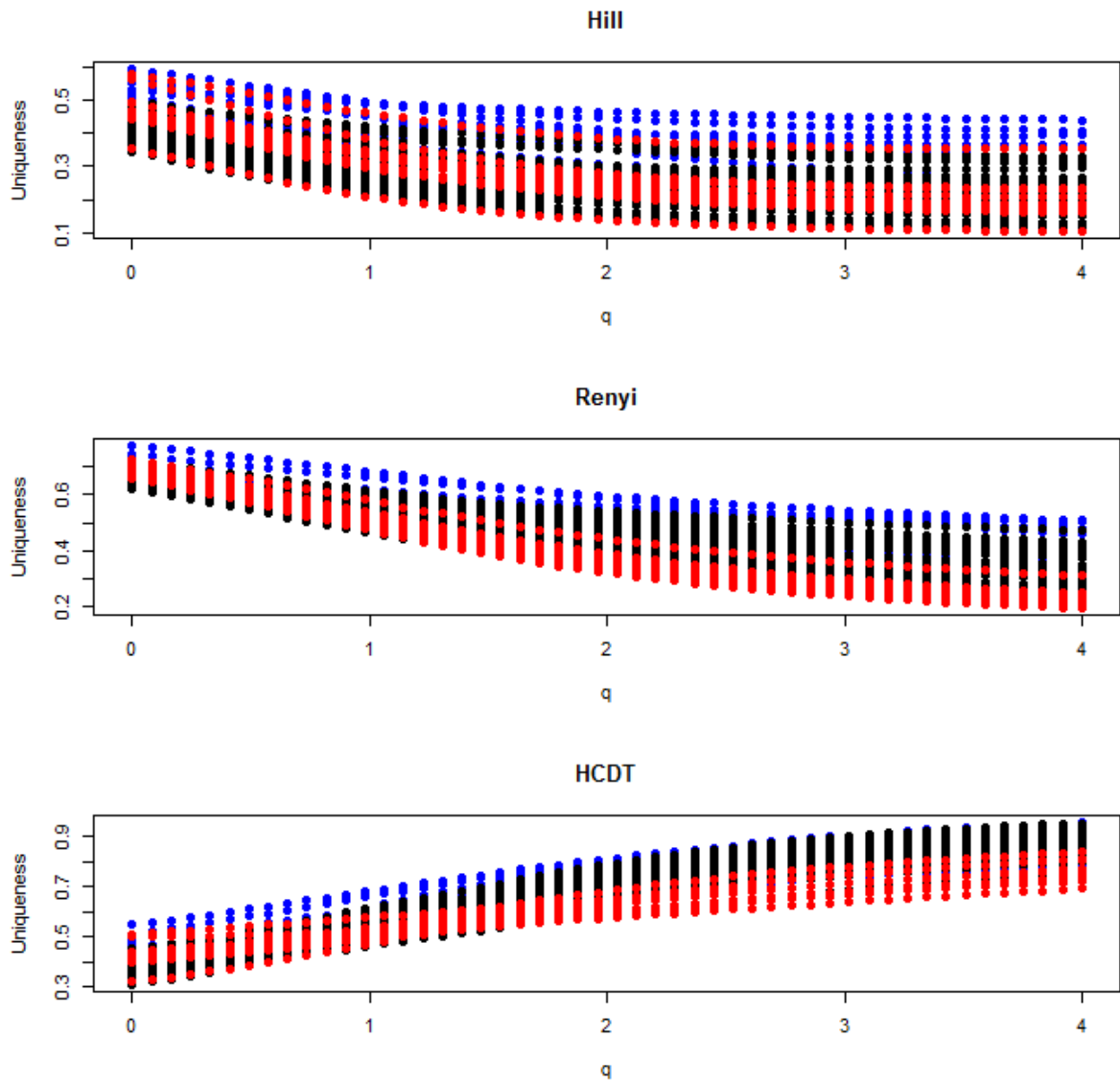


Fig. C-5. Functional feature uniqueness profile for each of the 59 plots -Note that here the scale changes from a graph to another.

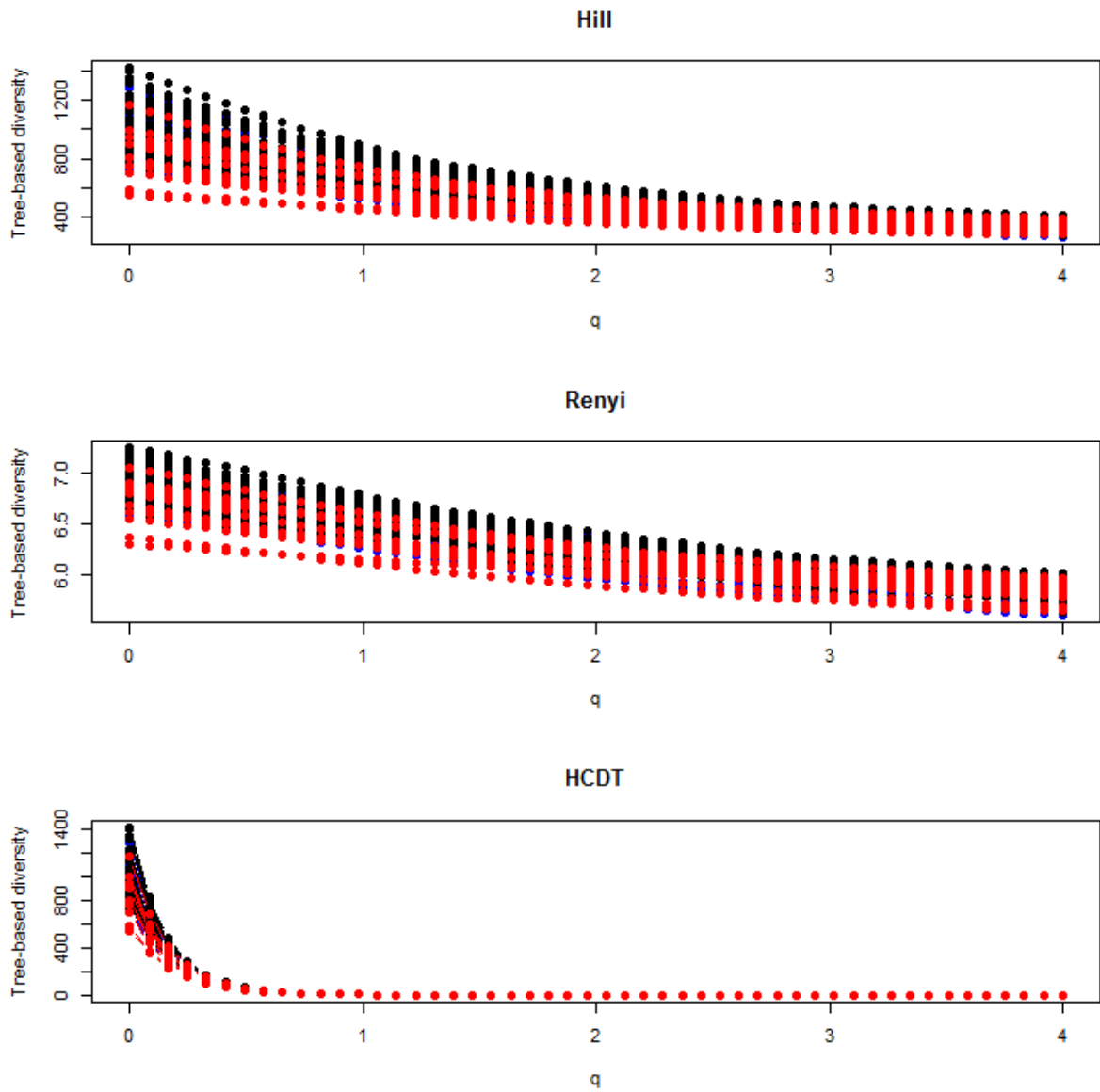


Fig. C-6. Phylogenetic feature diversity profile for each of the 59 plots -Note that here the scale changes from a graph to another.

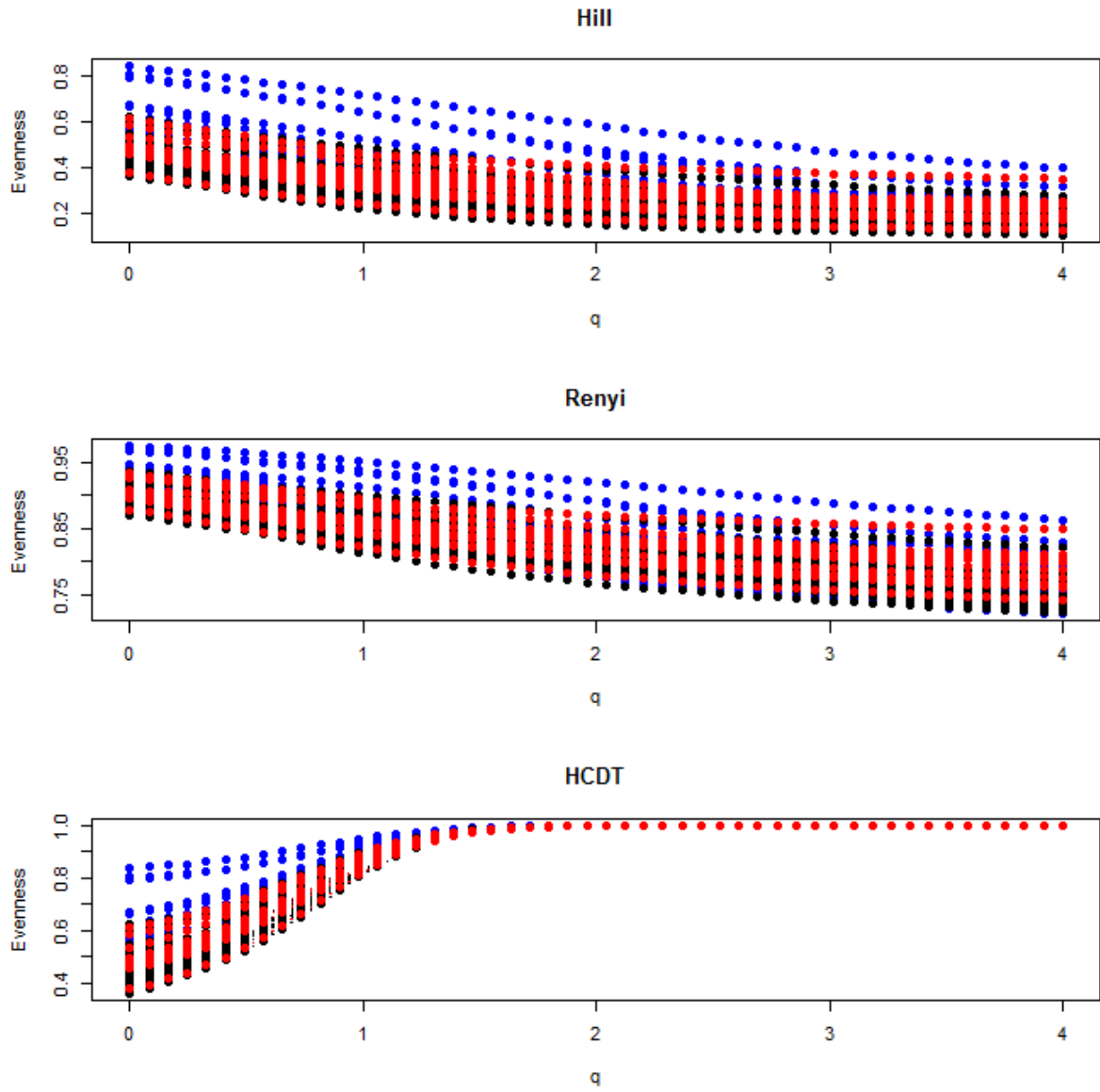


Fig. C-7. Phylogenetic feature evenness profile for each of the 59 plots -Note that here the scale changes from a graph to another.

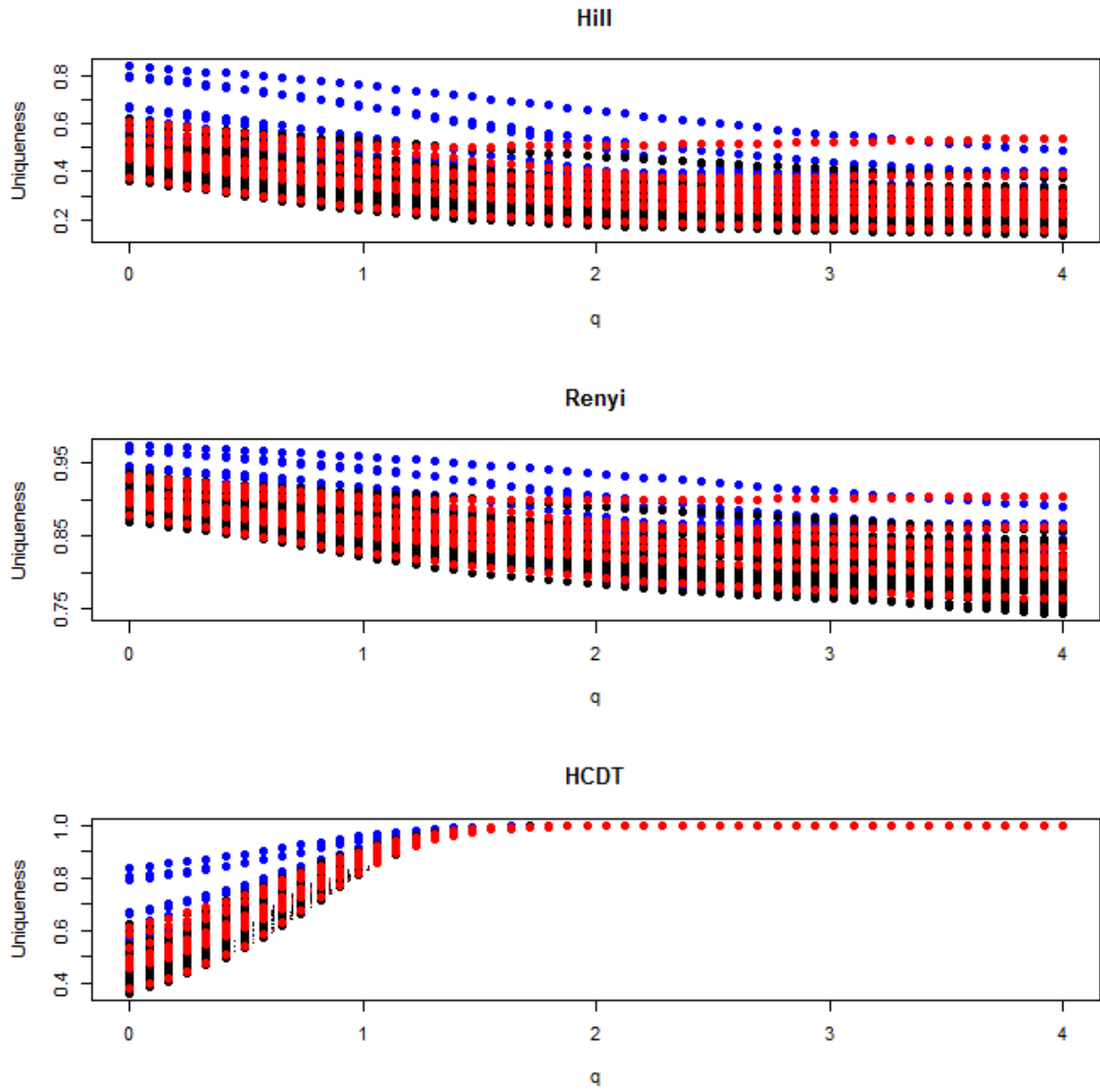


Fig. C-8. Phylogenetic feature uniqueness profile for each of the 59 plots -Note that here the scale changes from a graph to another.

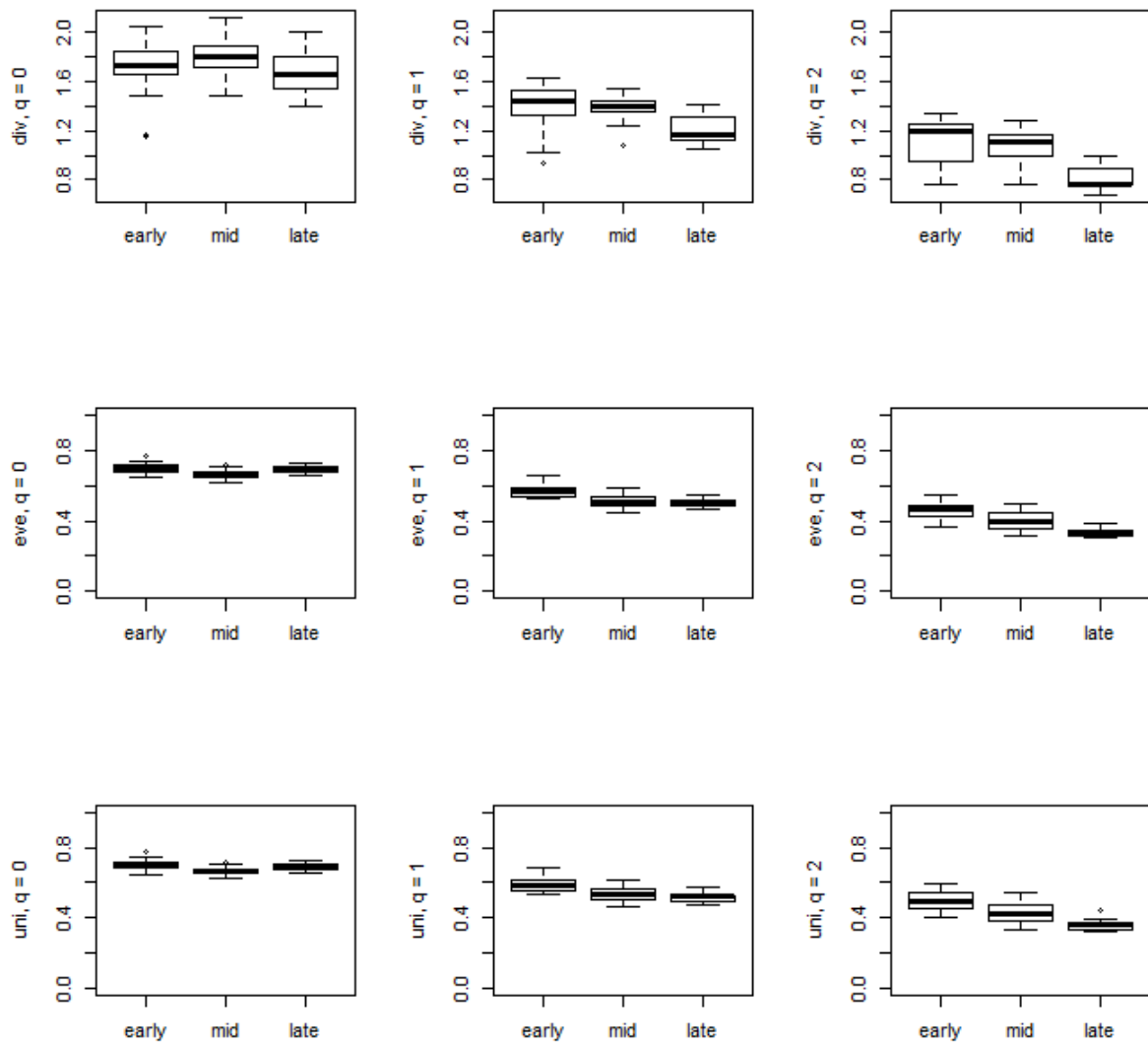


Fig. C-9. Functional data, Renyi index, div=diversity, eve=evenness, uni=uniqueness. Here we used the same scale for a given category of indices (e.g. uniqueness) but different values for q . In the next figure, the scales are different.

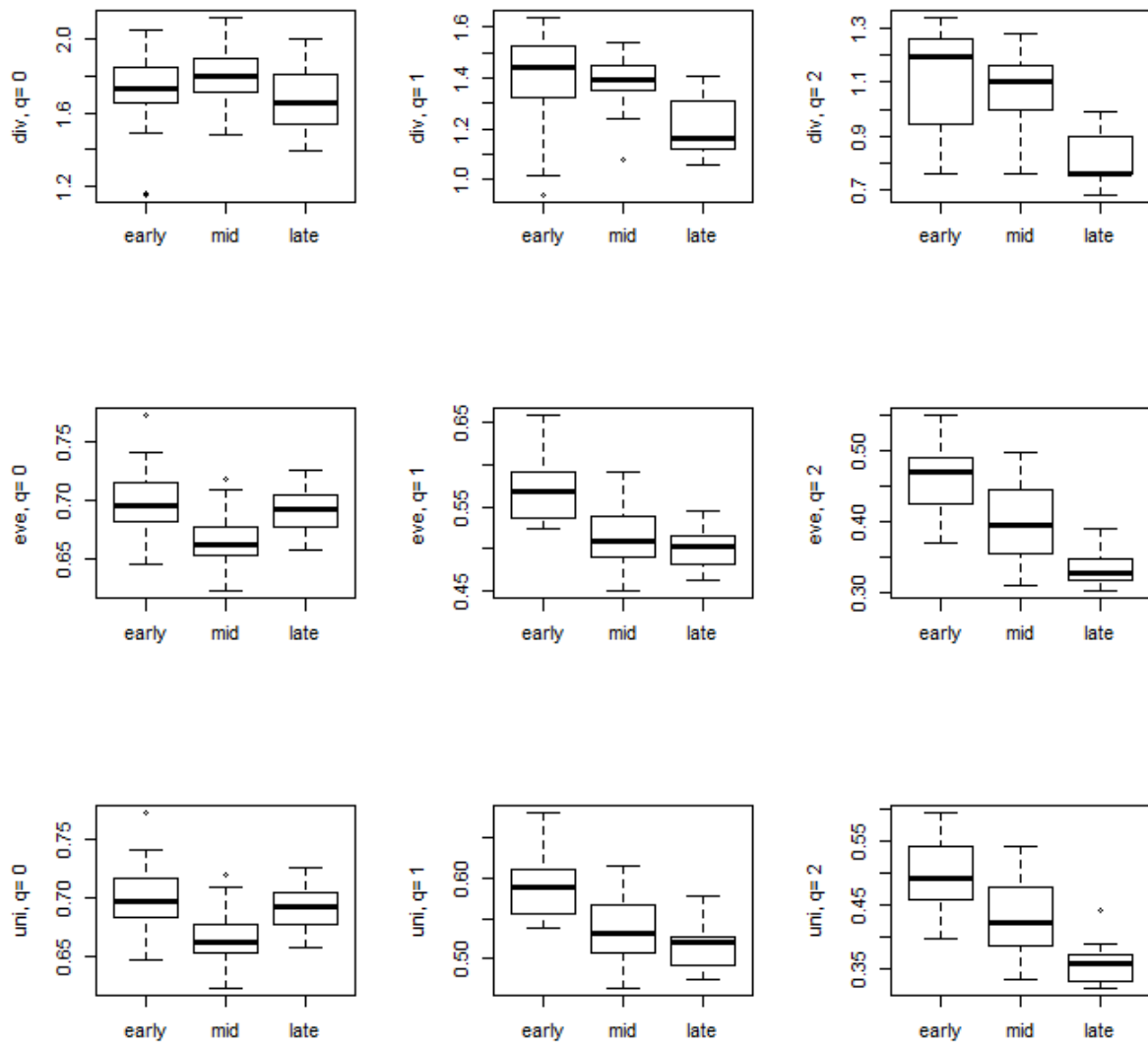


Fig. C-10. Same legend as Fig. C-9, but the scales here are different from a panel to another.

The trends observed are a decrease in functional diversity, evenness and uniqueness from early successional stage to late successional stage with $q=1$ and $q=2$ and U-shaped variations with $q=0$.

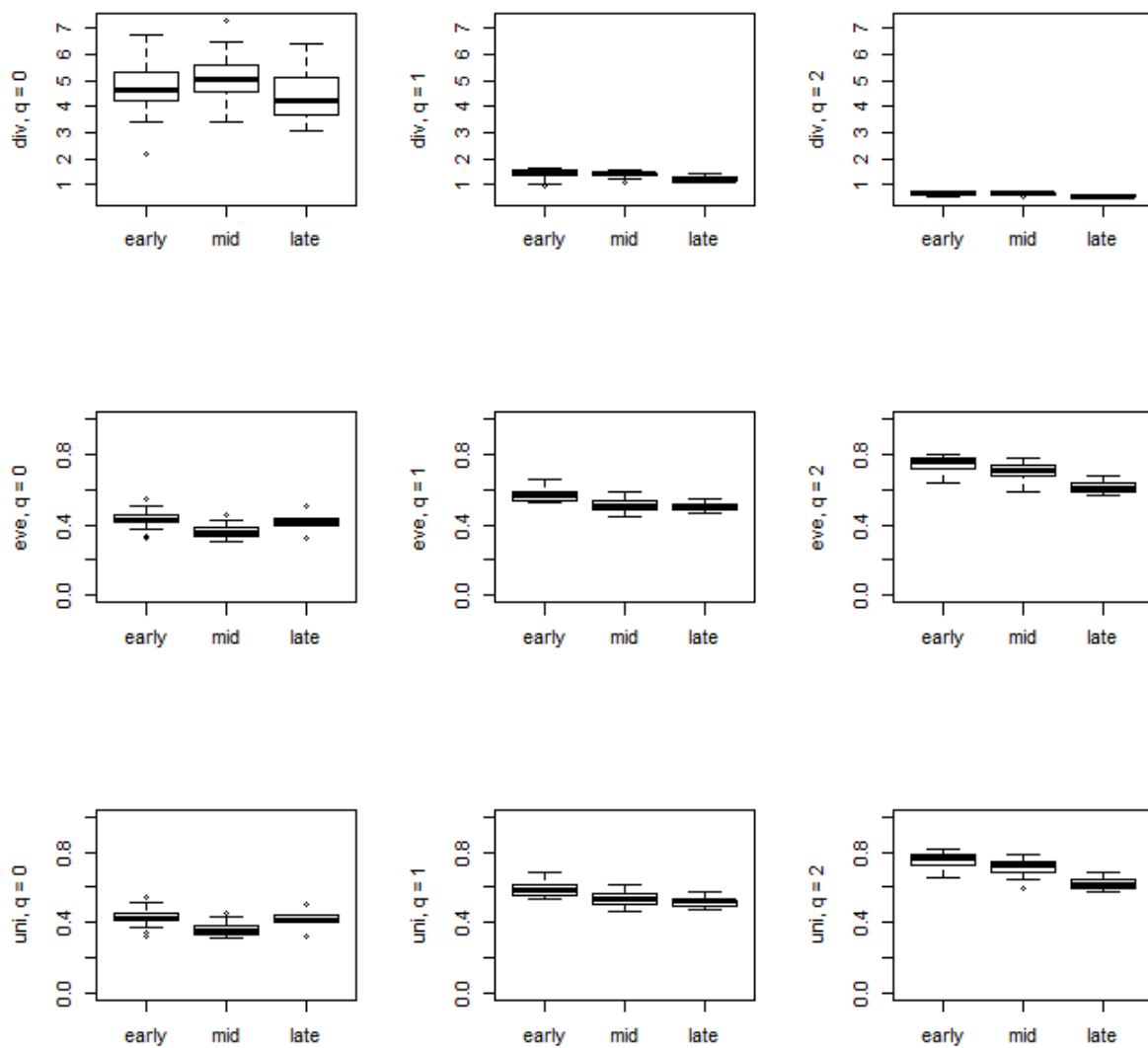


Fig. C-11. Functional data, HCDT index, div=diversity, eve=evenness, uni=uniqueness

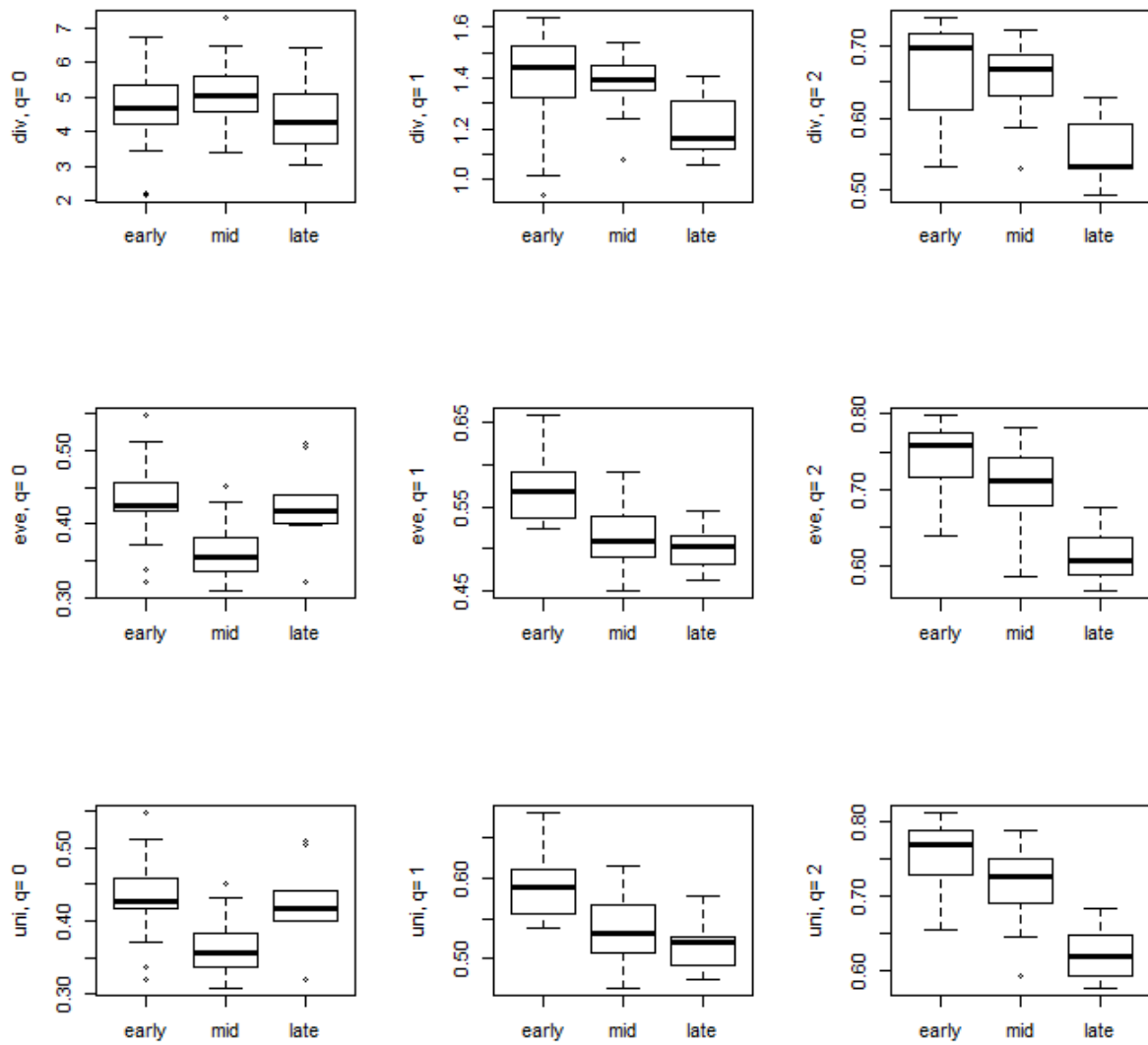


Fig. C-12. Same legend as Fig. C-11, but the scales here are different from a panel to another. The trends observed are close to those observed with the Rényi index: a decrease in functional diversity, evenness and uniqueness from early successional stage to late successional stage with $q=1$ and $q=2$ and U-shaped variations with $q=0$.

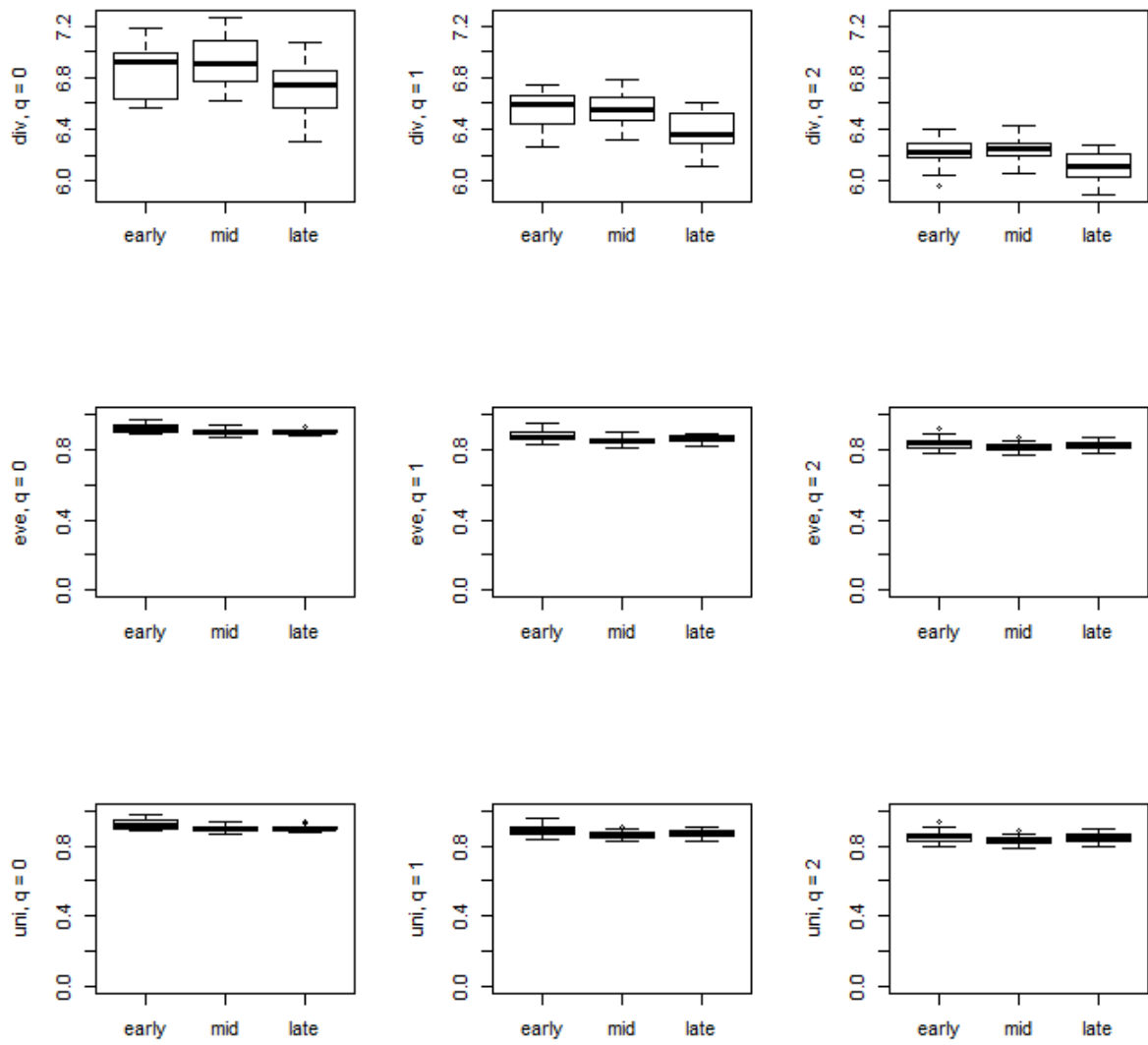


Fig. C-13. Phylogenetic data, Renyi index, div=diversity, eve=evenness, uni=uniqueness

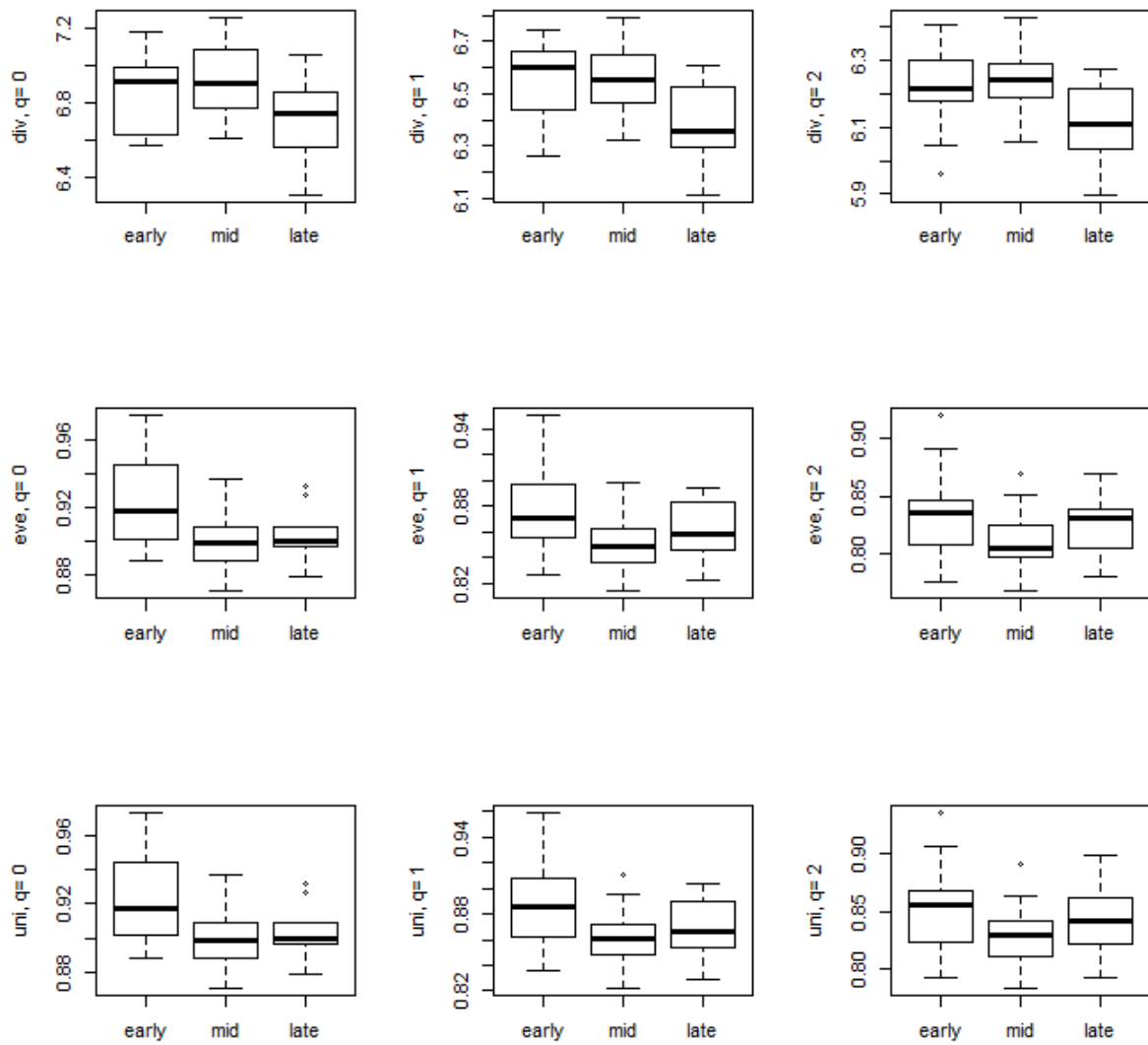


Fig. C-14. Same legend as Fig. C-13, but the scales here are different from a panel to another.

The trends observed are: a decrease in phylogenetic diversity from early/mid successional stage to late stage; a decrease in phylogenetic evenness and uniqueness from early to mid/late successional stages with $q=0$ and U-shaped variations with $q=1$ and $q=2$. In any cases, as shown by Fig. C-13, measured variations between successional stages are very small.

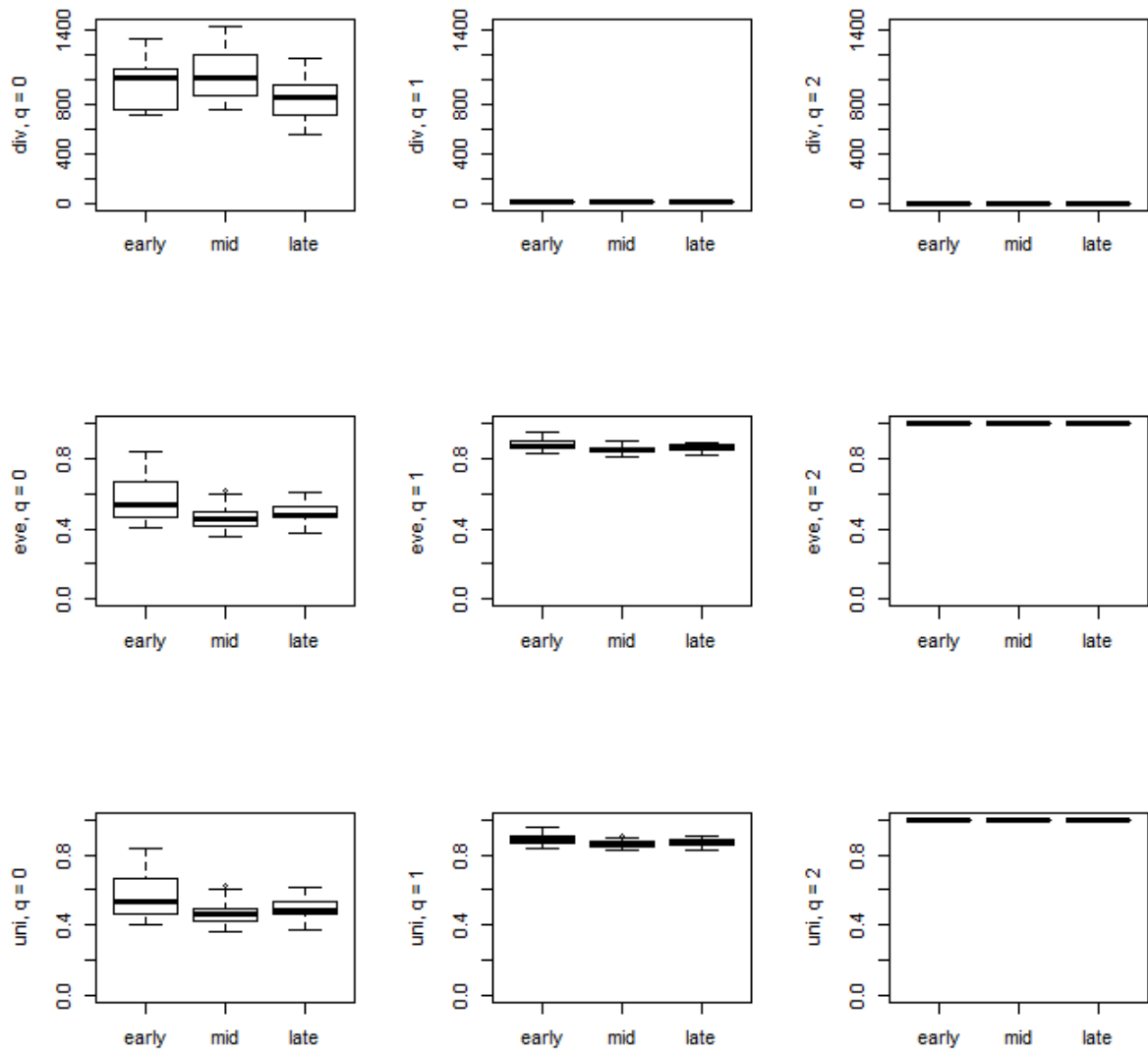


Fig. C-15. Phylogenetic data, HCDT index, div=diversity, eve=evenness, uni=uniqueness

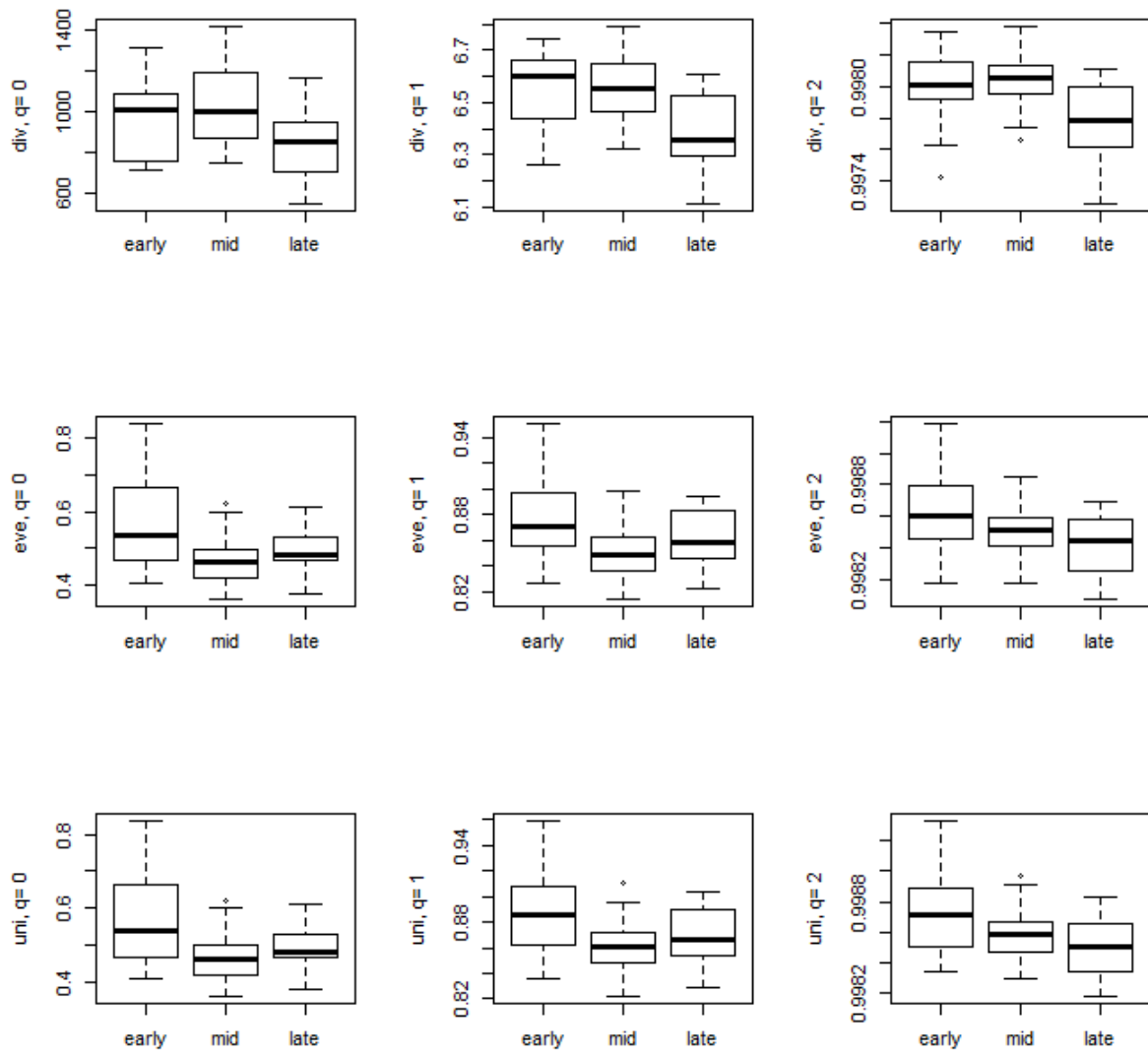


Fig. C-16. Same legend as Fig. C-15, but the scales here are different from a panel to another. The trends observed are: a decrease in phylogenetic diversity from early/mid successional stages to late stages; a decrease in phylogenetic evenness and uniqueness from early to mid/late successional stages with $q=2$ and U-shaped variations with $q=0$ and $q=1$. In any cases, as shown by Fig. C-15, measured variations between successional stages are very small.

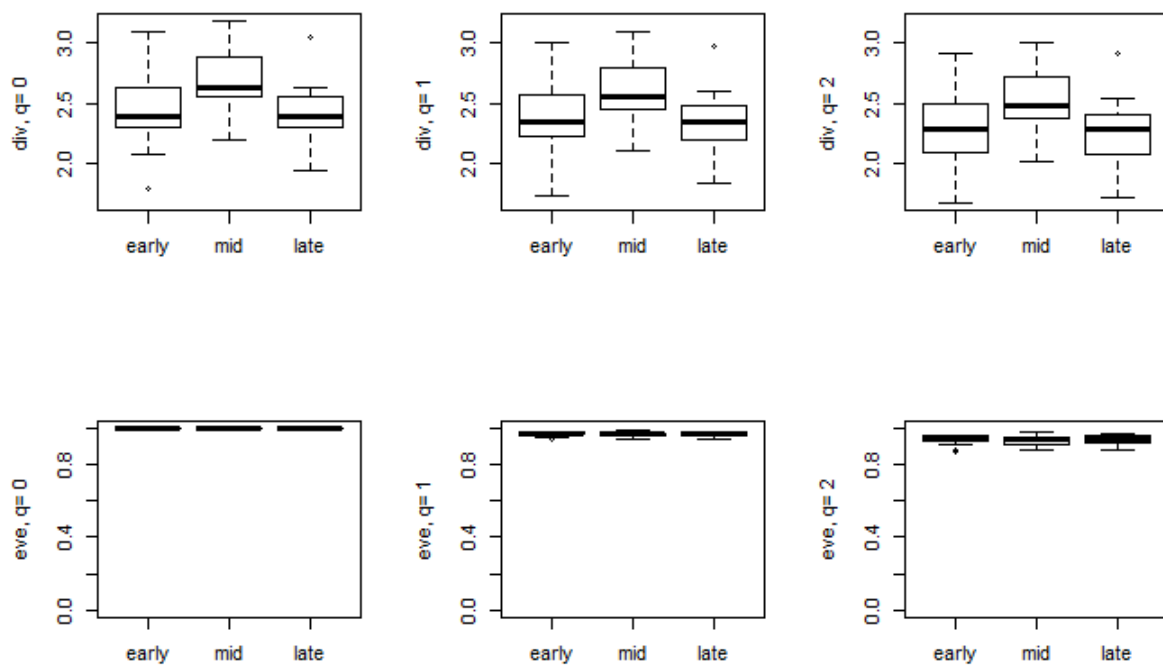


Fig. C-17. Species abundance data, HCDT index, div=diversity, eve=evenness

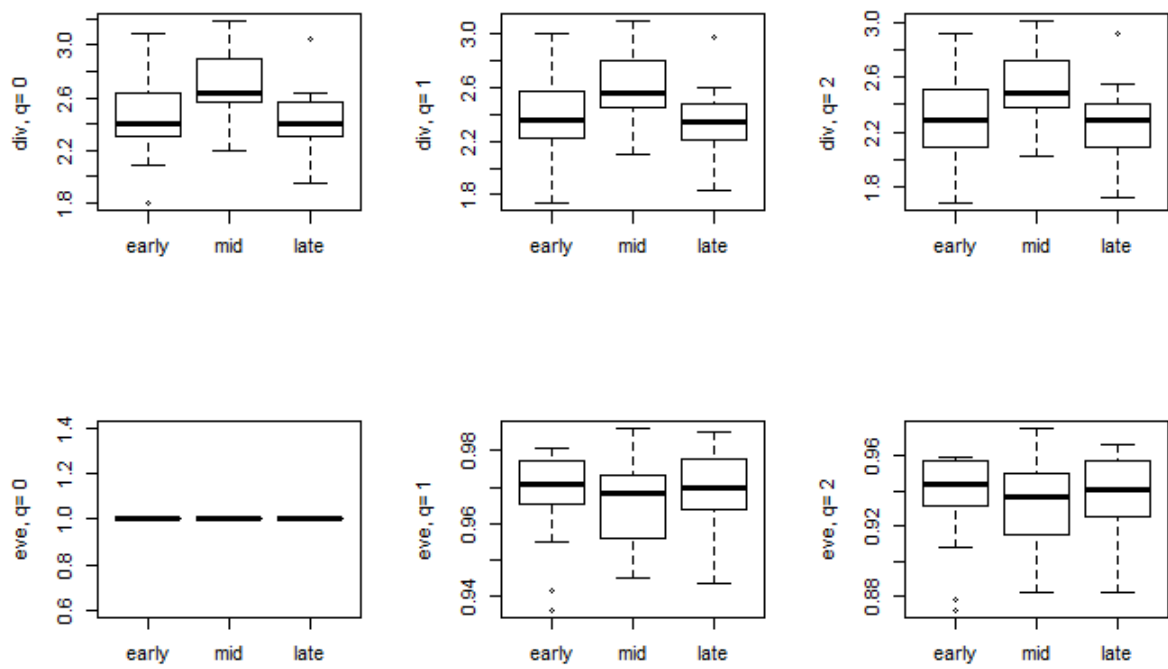


Fig. C-18. Same legend as Fig. C-17, but the scales here are different from a panel to another.

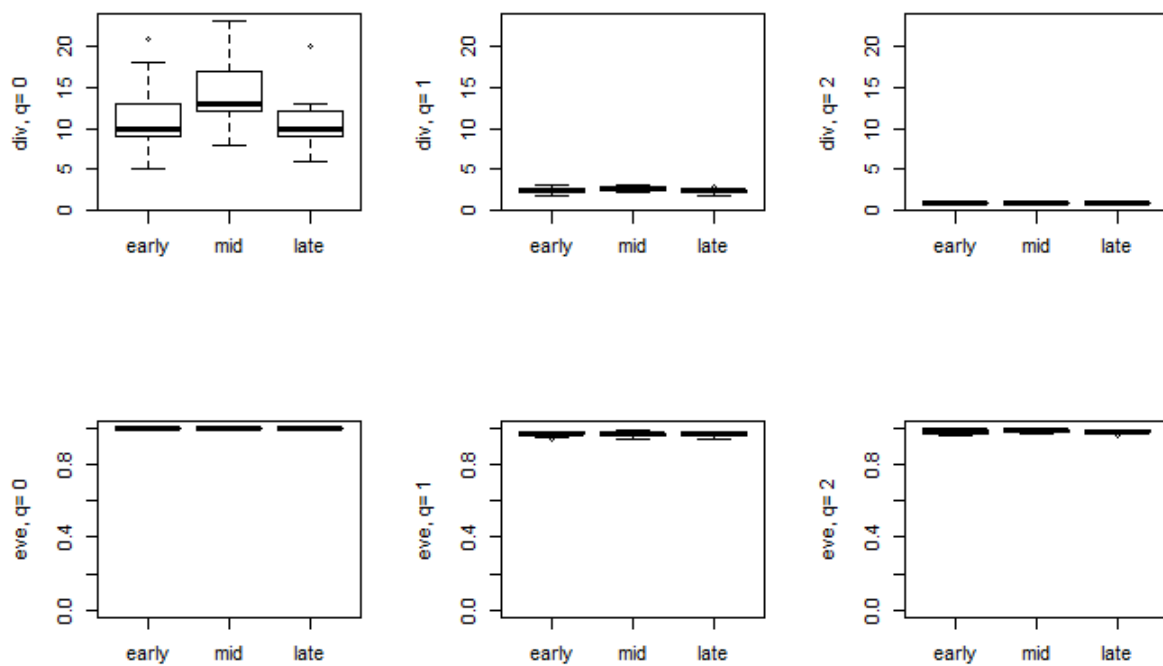


Fig. C-19. Species abundance data, HCDT index, div=diversity, eve=evenness

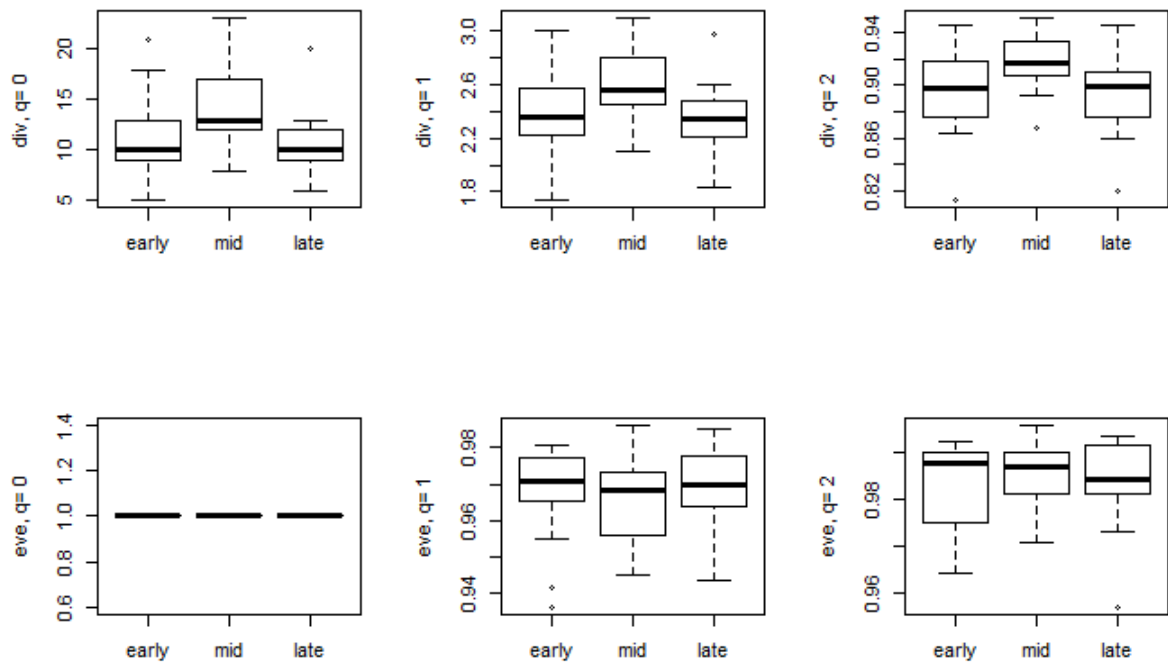


Fig. C-20. Same legend as Fig. C-19, but the scales here are different from a panel to another.

Appendix D. Illustrative examples associated with Table 5

Most of our illustrative examples below rely on the theoretical data set introduced in Fig. 3a of the main text. We provide this data set again in Fig. D.1 attributing names to species (tips of the tree).

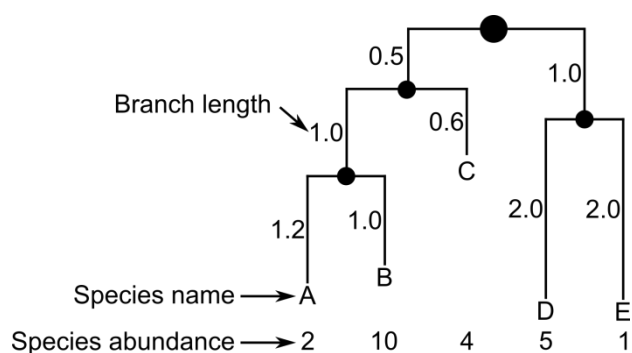


Fig. D.1. Data set derived from Fig. 3a of the main text

Properties D1 and E3: the ${}^q fdiv$ indices discussed in the main text, and ${}^q feve_{Hill}$ and ${}^q feve_{Rényi}$ are non-increasing functions of q , while ${}^q feve_{HCDT}$ can increase with q

Examples can be found in Appendix C in connection with our case study on plant diversity on the foreland of the Rutor glacier in Italy.

We provide also in Figs. D.2 and D.3 another illustrative example using the data from Fig. D.1.

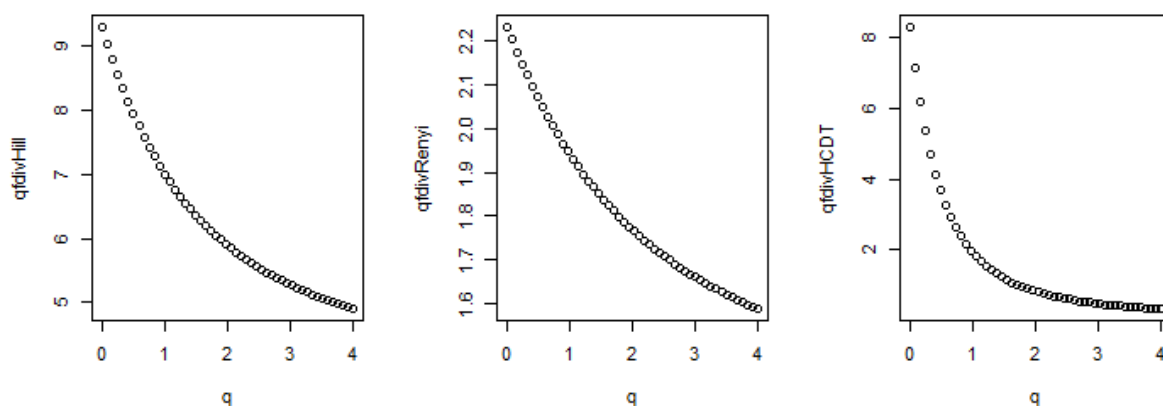


Fig. D.2. ${}^q fdiv$ indices applied to the data set given in Fig. D.1

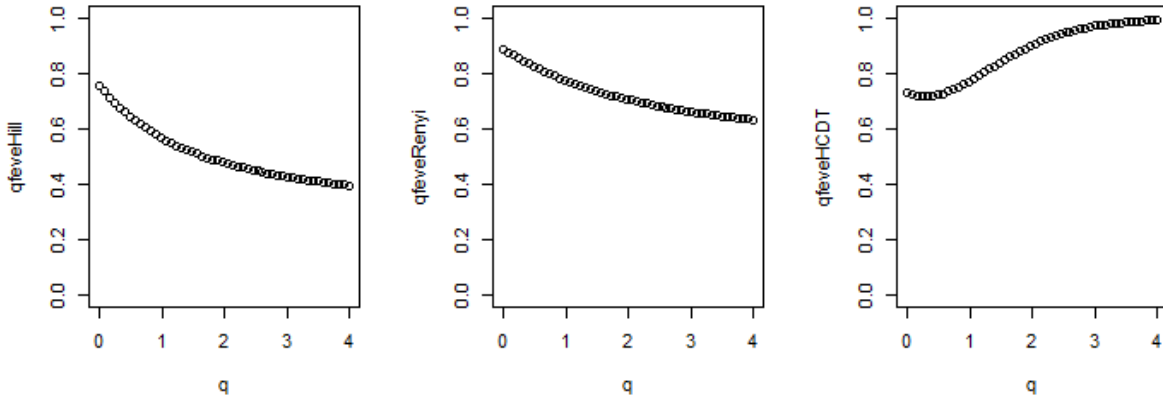


Fig. D.3. ${}^q\text{feve}$ indices applied to the data set given in Fig. D.1

Property D2: the ${}^q\text{fdiv}$ indices discussed in the main text are unaffected if a species is split into two similar subspecies supported by zero-length branches

We split the third species in Fig. 3a into two subspecies of abundance 1 and 3, respectively. This leads to the following modified data set (Fig. D.4):

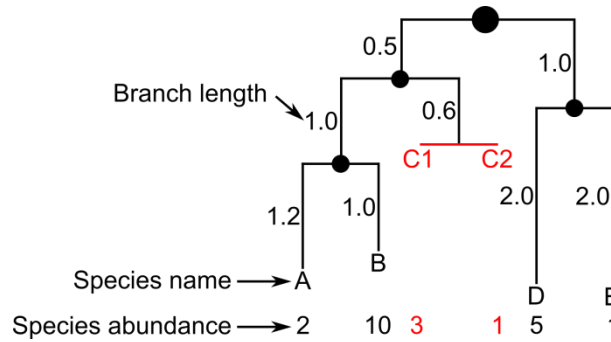


Fig. D.4. Theoretical data set modified from Fig. D.1. Here species C has been split into two subspecies C1 and C2 supported by zero-length branches.

The values of the ${}^q\text{fdiv}$ indices discussed in the main text are unaffected by this change as shown in Fig. D.5.

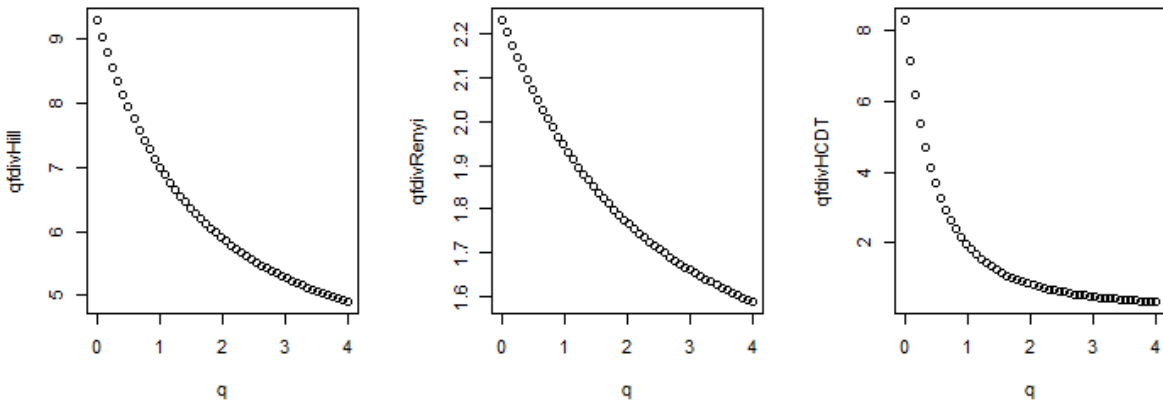


Fig. D.5. ${}^q\text{fdiv}$ indices applied to the data set given in Fig. D.4.

Property D3: the ${}^q fdiv$ indices discussed in the main text are Schur-concave

Starting again with the data set given in Fig. D.1, we added a branch of length 1.5 to root of the tree. We decreased the abundance of species D from 5 to 4.5 and attributed an abundance of 1 to the new species F. This led to the following modified data set (Fig. D.6):

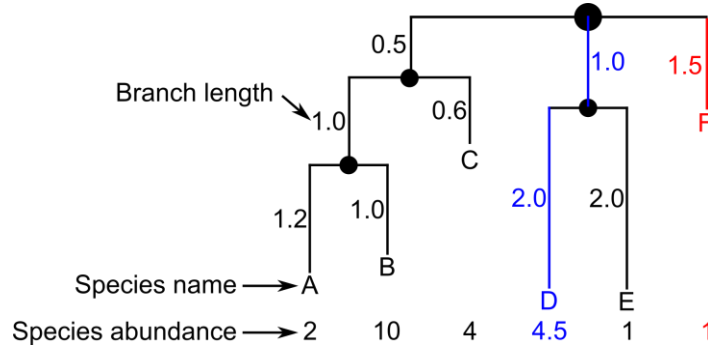


Fig. D.6. Theoretical data set modified from Fig. D.1. Here a new species F has been added with abundance 1 and the abundance of species D was decreased from 5 to 4.5. That way, the abundance gained by all features supported by the red branch ($1.5[\text{branch length}] * 1[\text{gain in species F abundance}] = 1.5[\text{gain in red feature abundance}]$) are compensated by the loss in abundance that affected all features supported by the blue branches ($3[\text{branch length}] * 0.5[\text{loss in species D abundance}] = 1.5[\text{loss in blue feature abundance}]$).

The values of the ${}^q fdiv$ indices are increased by this changed as shown in Fig. D.7.

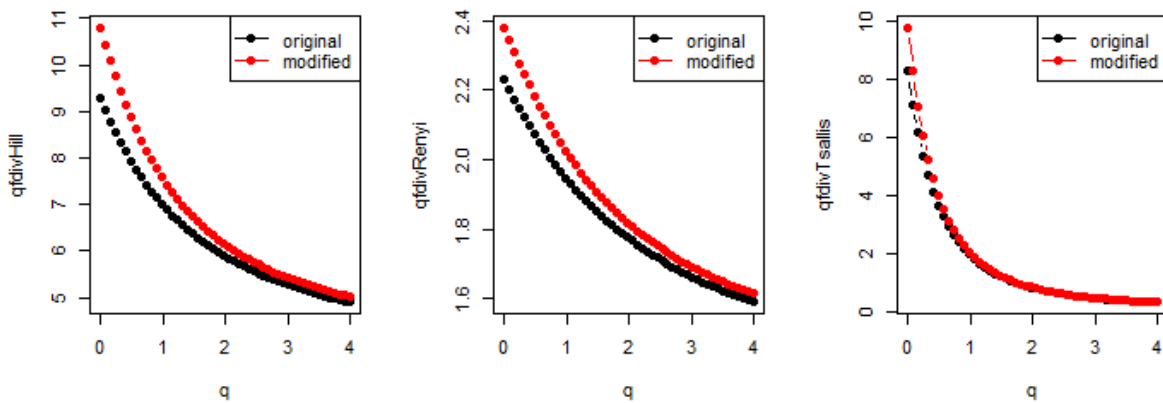


Fig. D.7. ${}^q fdiv$ indices applied to the data sets given in Fig. D.1 (black line) and D.6 (red line)

Property D4 [the ${}^q fdiv$ indices discussed in the main text are dependent of the measurement unit of branch length] and Property D8 [${}^q fdiv_{Hill}$ is multiplied by λ if the branch lengths are multiplied by λ ($\lambda > 0$) while ${}^q fdiv_{Renyi}$ and ${}^q fdiv_{HCDT}$ are not]

We modified the data set given in Fig. D.1, multiplying all branch lengths by 10. This led to the following modified data set (Fig. D.8):

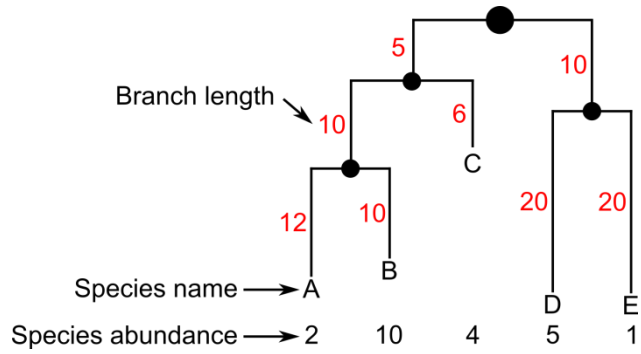


Fig. D.8. Theoretical data set modified from Fig. D.1. Here branch lengths have been multiplied by 10.

The values of the ${}^q fdiv$ indices are modified by this change in branch lengths as shown in Figs. D.9 ad D.10.

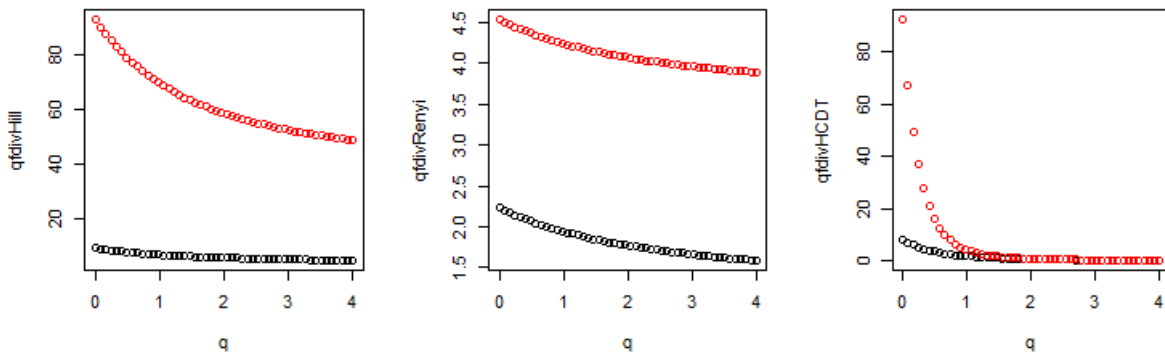


Fig. D.9. ${}^q fdiv$ indices applied to the data sets given in Fig. D.1 (black line) and D.8 (red line)

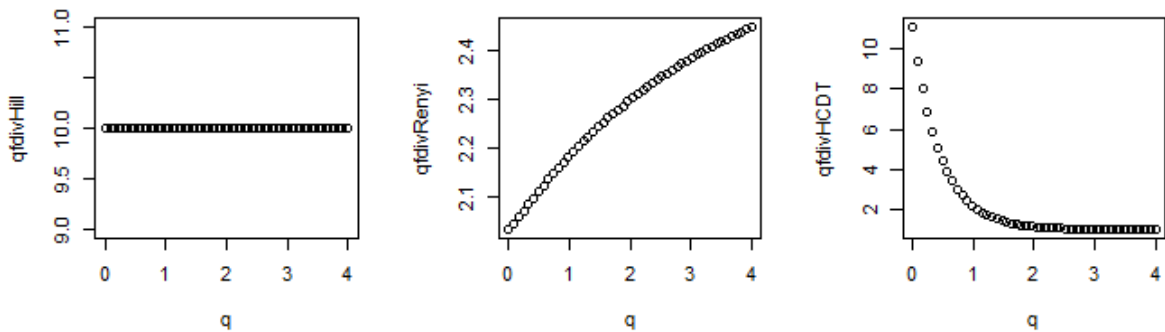


Fig. D.10. Ratio, for each ${}^q fdiv$ index, of its value for the modified data set as in Fig. D.8 to its value for the original data set (Fig. D.1).

Property D5: ${}^q fdiv_{Hill}$ and ${}^q fdiv_{Rényi}$ are not dependent of q at their maximum, while ${}^q fdiv_{HCDT}$ is.

We modified the data set given in Fig. D.1 so that all species are independent and have equal abundance. This led to the following modified data set (Fig. D.11):

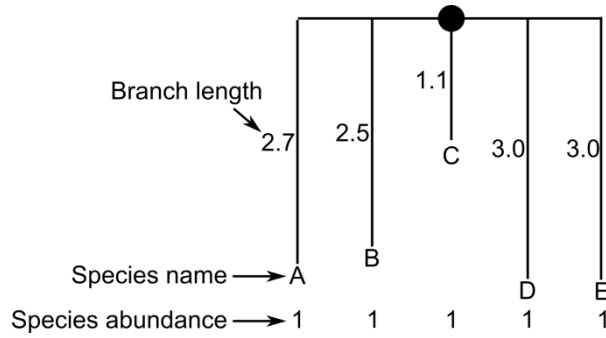


Fig. D.11. Theoretical data set modified from Fig. D.1 so that all species are independent and have equal abundance.

The ${}^q\text{div}$ values obtained with this modified data set are given in Fig. D.12.

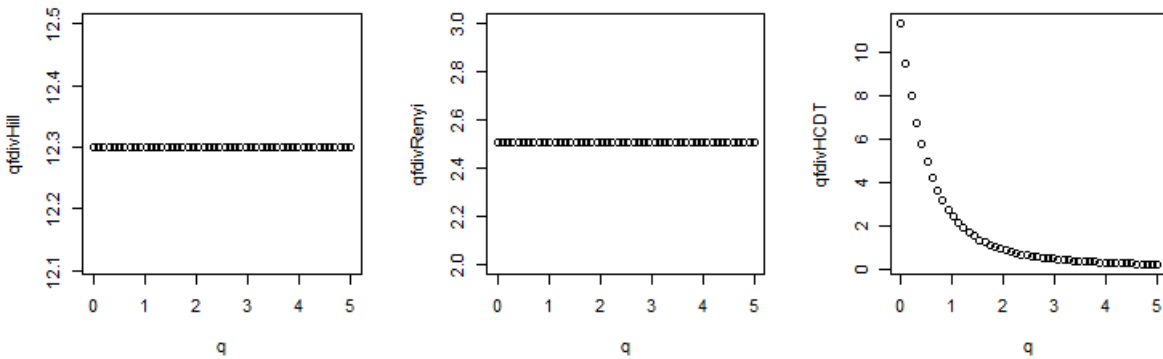


Fig. D.12. ${}^q\text{div}$ indices applied to the data sets given in D.11.

Property D6: ${}^q\text{div}_{Hill}$ is an effective number of features while ${}^q\text{div}_{Renyi}$ and ${}^q\text{div}_{HCDT}$ are not

We illustrate below this property for index ${}^q\text{div}_{Hill}$.

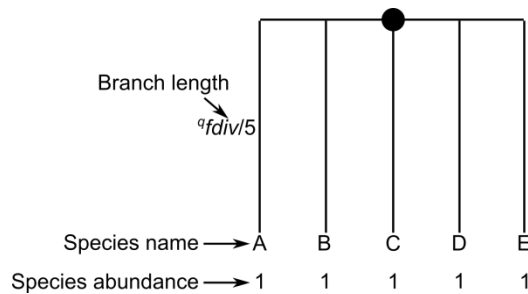


Fig. D.13. Theoretical data set modified from Fig. D.1. Here the tree is like a star with each species connected directly to the root. The length of each branch will be equal to the fifth of the value taken by ${}^q\text{div}_{Hill}$.

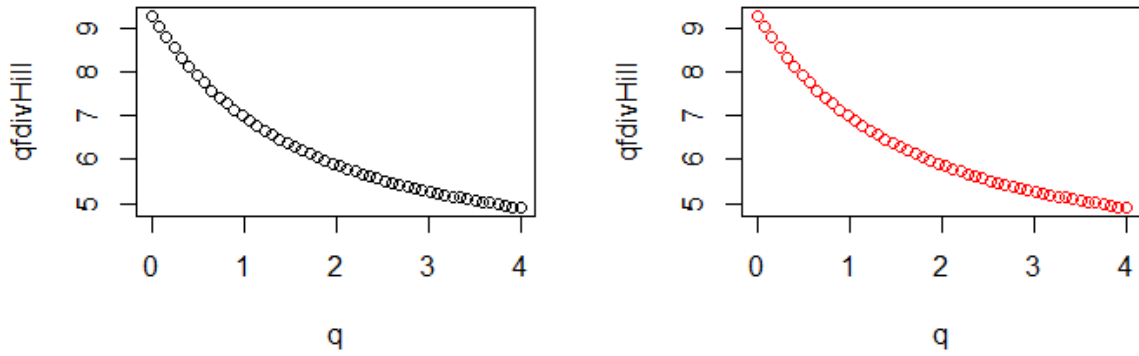


Fig. D.14. ${}^q fdiv_{Hill}$ index applied to the data sets given in Fig. D.1 (black line) and Fig. D.13 (red line). The two graphs are equal because ${}^q fdiv_{Hill}$ is an effective number of features. By definition, ${}^q fdiv_{Rényi}$ is the log of an effective number of features. ${}^q fdiv_{HCDT}$ can take values lower than 1 and is not expressed as an effective number of features.

Property D7: ${}^q fdiv_{Hill}$ is multiplied by N if the composition of a community is replicated N times as in Fig. 3 while ${}^q fdiv_{Rényi}$ and ${}^q fdiv_{HCDT}$ are not

We provide below diversity profiles for Fig. 3b (Fig. D.15) and the ratio of the diversity profiles for Fig. D.1 to the diversity profiles for Fig. 3b (Fig. D.16).

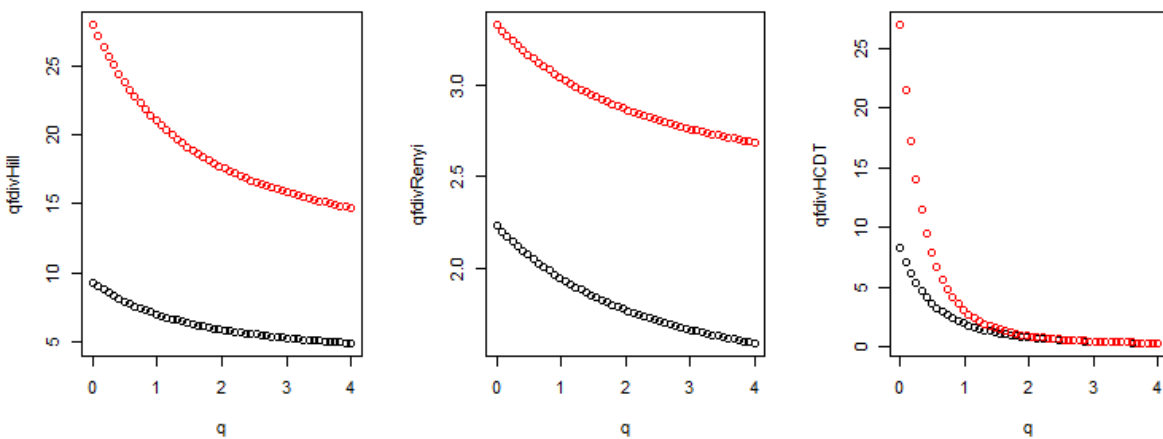


Fig. D.15. ${}^q fdiv$ indices applied to the data sets given in Fig. D.1 (black line) and Fig. 3b (red line)

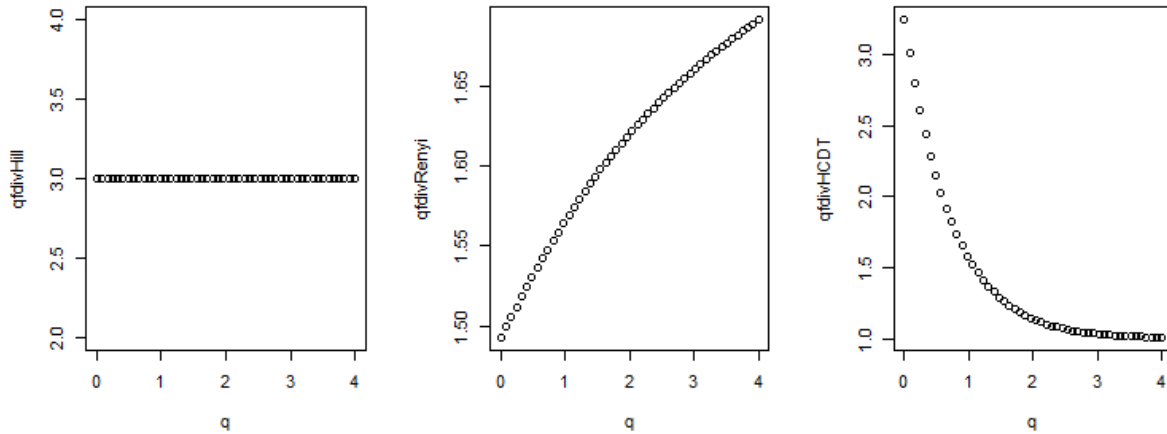


Fig. D.16. Ratio, for each ${}^q fdiv$ index, of its value for the replicated data set (Fig. 3b) to its value for the original data set (Fig. D.1)

Properties E1 and U1: the ${}^q feve$ and ${}^q funi$ indices discussed in the main text are constrained to a 0-1 range for convenience

We provide illustrations of this property for evenness in Fig. D.17 and for uniqueness in Fig. D.18.

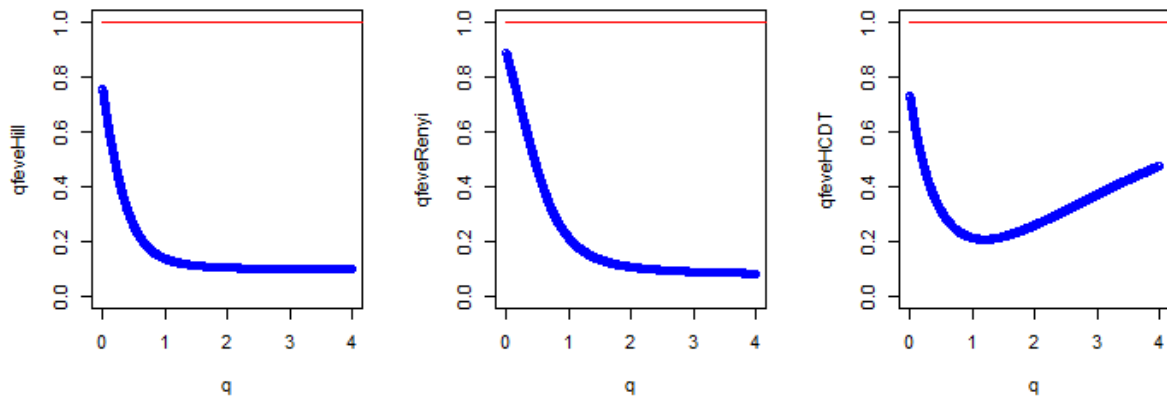


Fig. D.17. ${}^q feve$ profiles in two extreme cases: for the blue curve, we provided an abundance of 400 to species C in Fig. D.1. as an example of low evenness; for the red curve, we considered the data set in Fig. D.11 as an example of maximum evenness.

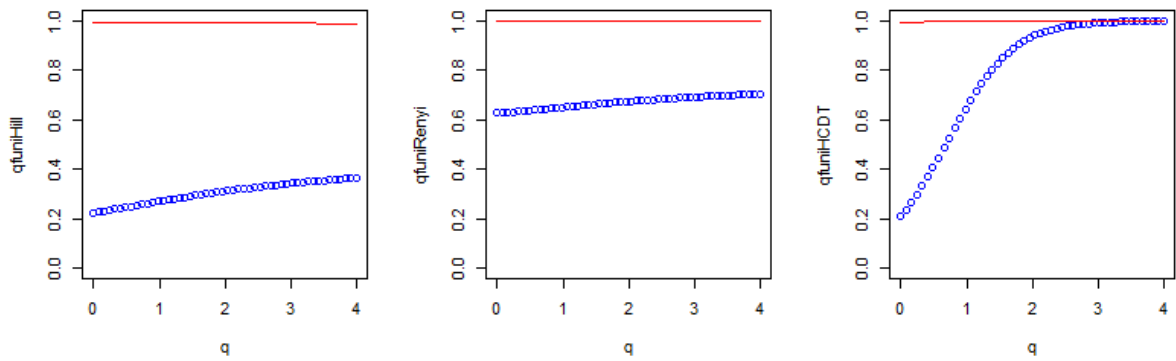


Fig. D.18. ${}^q\text{uni}$ profiles in two extreme cases: species are very similar (blue curve based on the data set in Fig. D.19) or very dissimilar (blue curve based on the data set in Fig. D.20).

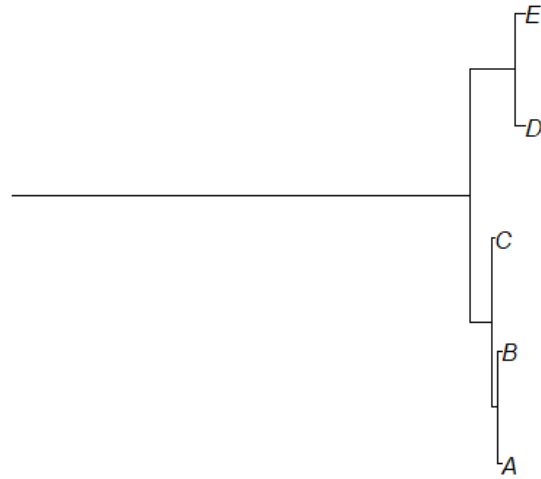


Fig. D.19. Theoretical data set modified from Fig. D.1 so that species are all very similar (related).

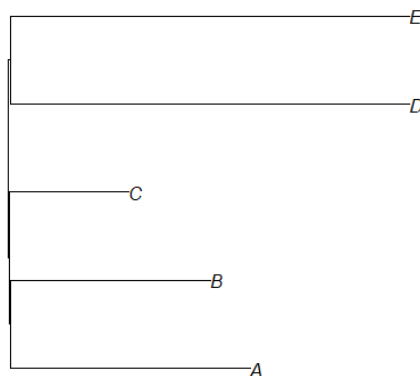


Fig. D.20. Theoretical data set modified from Fig. D.1 so that species are all very dissimilar (unrelated).

Property E2: ${}^q feve_{Hill}$ and ${}^q feve_{Rényi}$ reflect the evenness in the abundances of features, while ${}^q feve_{HCDT}$ doesn't

Fig. D.17 illustrates this statement as the profile curve for ${}^q feve_{HCDT}$ has a U-shape instead of decreasing with q .

Property U2: ${}^q feve_{Hill}$ and ${}^q feve_{Rényi}$ always increase with the amount of differences between species and reach their maximum only when species are maximally dissimilar, while ${}^q feve_{HCDT}$ doesn't

Fig. D.18 illustrates this statement as ${}^q feve_{HCDT}$ rapidly reaches its maximum when applied to the data set in Fig. D.20 although species in this data set are very similar.

Properties E4 and U3: ${}^q feve_{Hill}$ and ${}^q funi_{Hill}$ are not changed if the composition of a community is replicated N times as in Fig. 3, while ${}^q feve_{Rényi}$, ${}^q funi_{Rényi}$, ${}^q feve_{HCDT}$ and ${}^q funi_{HCDT}$ are

We provide illustrations of these properties below:

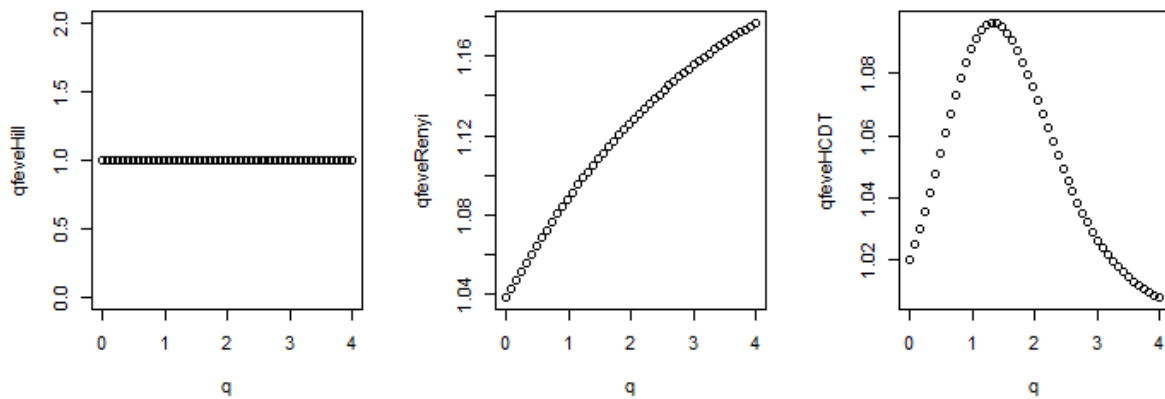


Fig. D.21. Ratio, for each ${}^q feve$ index, of its value for the replicated data set (Fig. 3b) to its value for the original data set (Fig. D.1)

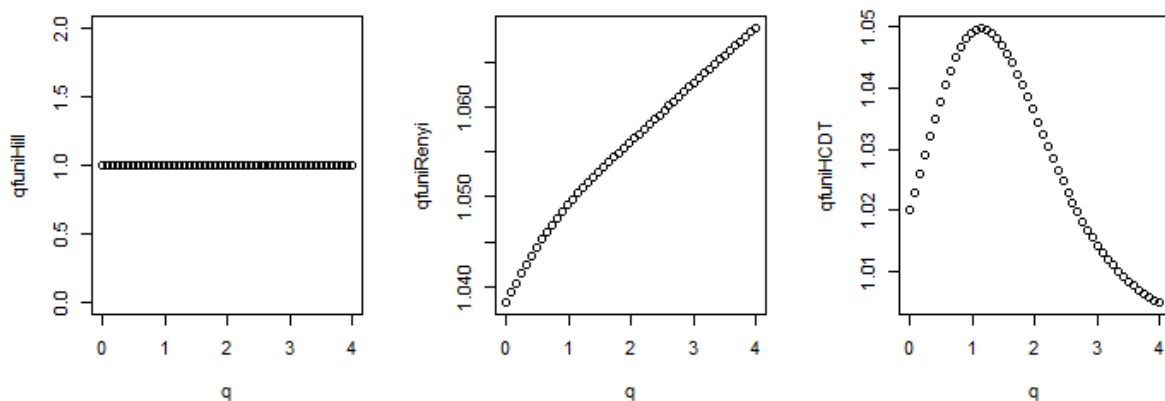


Fig. D.22. Ratio, for each ${}^q funi$ index, of its value for the replicated data set (Fig. 3b) to its value for the original data set (Fig. D.1)

"A simple translation from indices of species diversity to indices of phylogenetic diversity" by

Sandrine Pavoine & Carlo Ricotta

Appendix E. R scripts

We provide in Appendix F the R scripts of four functions named `evoveparam`, `plot.evoveparam`, `evouniparam`, `plot.evouniparam` + an internal function named `.checkphyloarg`. Updated versions of these functions will be integrated, after the publication of our paper, in the R package `adiv` (Pavoine, 2018). This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License <http://www.gnu.org/licenses/>.

Disclaimer: users of this code are cautioned that, while due care has been taken and it is believed accurate, it has not been rigorously tested and its use and results are solely the responsibilities of the user.

D.1. Functions `evoveparam` and `plot.evoveparam`

Function "evoveparam" calculates evenness in the abundance of features (branch units on a phylogenetic [or functional] tree) using parametric indices ${}^q feve_{HCDT}$, ${}^q feve_{Hill}$, ${}^q feve_{Renyi}$. Function "plot.evoveparam" plots the results of function "evoveparam".

usage:

```
evoveparam(phyl, comm, method = c("hill", "HCDT", "renyi"), q = 2, option = 1:3, H = NULL, tol = 1e-8)
```

```
plot.evoveparam(x, legend = TRUE, legendposi = "topright", type="b", col = if(is.numeric(x)) NULL else sample(colors(distinct = TRUE), nrow(x$eve)), lty = if(is.numeric(x)) NULL else rep(1, nrow(x$eve)), pch = if(is.numeric(x)) NULL else 1:nrow(x$eve), ...)
```

```
function(x, legend = TRUE, legendposi = "topright", type="b", col = if(is.numeric(x)) NULL else sample(colors(distinct = TRUE), nrow(x$eve)), lty = if(is.numeric(x)) NULL else rep(1, nrow(x$eve)), pch = if(is.numeric(x)) NULL else rep(19, nrow(x$eve)), ...)
```

dependence: `adiv`, `ape`, `phylobase`

arguments:

- `phyl`: an object inheriting the class "phylo" (see package `ape`), "phylo4" (see package `phylobase`), or "hclust".
- `comm`: a data frame or a matrix typically with communities as rows, species as columns and an index of abundance as entries. Species should be labeled as in the tree (object `phyl`) where they are the tips.
- `method`: a character with one of the following codes: "hill" (for ${}^q feve_{Hill}$), "HCDT" (for ${}^q feve_{HCDT}$), or "renyi" (for ${}^q feve_{Renyi}$). If several value are given, only the first one is considered.

- q : a vector with nonnegative value(s) for parameter " q " of functions ${}^q\text{feve}_{\text{HCDT}}$, ${}^q\text{feve}_{\text{Hill}}$, and ${}^q\text{feve}_{\text{Renyi}}$.
- option: an integer (either 1, 2 or 3). If 1, the (Hill, HCDT or Renyi) diversity index is divided by the value it would have if species had same abundance and were independent, if 2, the diversity is divided by the value it would have if species had same abundance, were independent, and at the maximum observed distance from tip to root, if 3, the diversity is divided by the value it would have if species had same abundance, were independent, and at a distance equal to H from the root of the tree. Options 1 and 2 are equivalent in case of an ultrametric tree.
- H : a numeric; H must be higher than the largest observed distance from tip to root.
- tol: numeric tolerance threshold: values between $-tol$ and tol are considered equal to zero.
- x : an object of class "evoeveparam" obtained with function "evoeveparam".
- legend: a logical. If TRUE a legend is given with the colour, the type of line (etc.) used to define the evenness curve of each community.
- legendposi: a character that gives the position of the legend to be passed to function "legend" of the base of R.
- type: a character to be passed to the graphic parameter "type" of functions "plot" and "lines" used to draw the evenness curve of each community.
- col: vector of colours to be passed to the graphic parameter "col" of functions "plot" and "lines" to define the colour of the evenness curve of each community.
- lty: vector of type of line (plain, broken etc.) to be passed to the graphic parameter "lty" of functions "plot" and "lines" used to draw the evenness curve of each community.
- pch: type of point (open circle, close circle, square etc.) to be passed to the graphic parameter "pch" of functions "plot" and "lines" used to draw the evenness level of each community.
- ...: other parameters can be added and passed to the functions "plot" and "lines" used to draw the graphic.

value

If only one value of q is given, the function "evoeveparam" returns a vector with communities' evenness values.

If more than one value of q is given, a list of two objects is returned:

- q : the vector of values for q ;
- eve: a data frame with the evenness of each community calculated for all values of q .

The function "plot.evoeveparam" returns a graphic.

example:

```
library(adv)
library(ape)
library(phylobase)
data(batcomm)
phy <- read.tree(text=batcomm$tre)
ab <- batcomm$ab[, phy$tip.label]
plot(evoeveparam(phy, ab))
plot(evoeveparam(phy, ab, q=seq(0, 10, length=20)))
```

D.2. Functions `evouniparam` and `plot.evouniparam`

Function "evouniparam" calculates feature uniqueness (features = branch units on a phylogenetic [or functional] tree) using parametric indices ${}^qf_{uni_{HCDT}}$, ${}^qf_{uni_{Hill}}$, ${}^qf_{uni_{Renyi}}$. Function "plot.evouniparam" plots the results of function "evouniparam".

usage:

```
evouniparam(phy, comm, method = c("hill", "HCDT", "renyi"), q = 2, tol = 1e-08)
```

```
plot.evouniparam(x, legend = TRUE, legendposi = "topright", type="b", col =  
if(is.numeric(x)) NULL else sample(colors(distinct = TRUE), nrow(x$uni)), lty =  
if(is.numeric(x)) NULL else rep(1, nrow(x$uni)), pch = if(is.numeric(x)) NULL else  
1:nrow(x$uni), ...)
```

dependence: `adiv`, `ape`, `phytools`, `phylobase`

arguments:

- `phy`: an object inheriting the class "phylo" (see package `ape`), "phylo4" (see package `phylobase`), or "hclust".
- `comm`: a data frame or a matrix typically with communities as rows, species as columns and an index of abundance as entries. Species should be labeled as in the tree (object `phy`) where they are the tips.
- `method`: a character with one of the following codes: "hill" (for ${}^qf_{uni_{Hill}}$), "HCDT" (for ${}^qf_{uni_{HCDT}}$), or "renyi" (for ${}^qf_{uni_{Renyi}}$). If several value are given, only the first one is considered.
- `q`: a vector with nonnegative value(s) for parameter "q" of functions ${}^qf_{uni_{Hill}}$, ${}^qf_{uni_{HCDT}}$, or ${}^qf_{uni_{Renyi}}$.
- `tol`: numeric tolerance threshold: values between `-tol` and `tol` are considered equal to zero.
- `x`: an object of class "evouniparam" obtained with function "evouniparam".
- `legend`: a logical. If TRUE a legend is given with the colour, the type of line (etc.) used to define the uniqueness curve of each community.
- `legendposi`: a character that gives the position of the legend to be passed to function "legend" of the base of R.
- `type`: a character to be passed to the graphic parameter "type" of functions "plot" and "lines" used to draw the uniqueness curve of each community.
- `col`: vector of colours to be passed to the graphic parameter "col" of functions "plot" and "lines" to define the colour of the uniqueness curve of each community.
- `lty`: vector of type of line (plain, broken etc.) to be passed to the graphic parameter "lty" of functions "plot" and "lines" used to draw the uniqueness curve of each community.
- `pch`: type of point (open circle, close circle, square etc.) to be passed to the graphic parameter "pch" of functions "plot" and "lines" used to draw the uniqueness level of each community.
- ...: other parameters can be added and passed to the functions "plot" and "lines" used to draw the graphic.

value

If only one value of q is given, the function "evouniparam" returns a vector with communities' uniqueness values.

If more than one value of q is given, a list of two objects is returned:

- q : the vector of values for q ;
- uni: a data frame with the uniqueness of each community calculated for all values of q .

The function "plot.evouniparam" returns a graphic.

example:

```
library(adv)
library(ape)
library(phytools)
library(phylobase)
data(batcomm)
phy <- read.tree(text=batcomm$tre)
ab <- batcomm$ab[, phy$tip.label]
plot(evouniparam(phy, ab))
plot(evouniparam(phy, ab, q=seq(0, 10, length=20)))
```

References

Pavoine, S., 2018. adv: Analysis of Diversity. R package, version 1.2. <https://CRAN.R-project.org/package=adv>


```

evolveparam <-
function(phyl, comm, method = c("hill", "HCDT", "renyi"),
  q = 2, option = 1:3, H = NULL, tol = 1e-8){

tre <- .checkphyloarg(phyl)
PHYTEST <- tre$phyl.phylo
A <- write.tree(PHYTEST)
if(!substr(A, nchar(A)-1, nchar(A))==";");")
  PHYTEST <- read.tree(text=paste0("(", substr(A,1,nchar(A)-1), ");"))
tre4 <- as(PHYTEST, "phylo4")
m <- comm
method <- method[1]
option <- option[1]
if(!option%in%(1:3)) stop("Unavailable option")
if(!method%in%c("HCDT", "hill", "renyi")) stop("Unavailable method")
if(!(is.numeric(q) | is.integer(q))) stop("Incorrect definition for q")
if(any(q < -tol)) stop("q must be nonnegative")
if(is.null(colnames(m))) stop("m must have names for column")
if(any(colSums(m)==0)){
  nsp <- ncol(m)
  nspreal <-length((1:nsp)[colSums(m) > 0])
  if(nspreal>1)
    m <- m[, colSums(m) > 0, drop = FALSE]
}
ncom <- nrow(m)
if(any(m<0)) stop("m should contain nonnegative values")
if(any(rowSums(m)==0)) stop("empty communities should be discarded")
if(any(!colnames(m)%in%tipLabels(tre4))) stop("m contains tip names that are not available in phyl")

if(!hasEdgeLength(tre4)){
  treape <- as(tre4, "phylo")
  tre4 <- as(compute.brlen(treape, 1), "phylo4")
}
if(!isRooted(tre4)){
  treape <- as(tre4, "phylo")
  treape$root.edge <- 0
  tre4 <- as(treape, "phylo4")
}

if(!hasNodeLabels(tre4)) nodeLabels(tre4) <- names(nodeLabels(tre4))
else{
  e <- nodeLabels(tre4)
  e[is.na(e)] <- names(e[is.na(e)])
  nodeLabels(tre4) <- e
}

reduceTree <- function(tree){
  C <- vcv(tree, model = "Brownian")
  if(min(diag(C))<1){
    warning("The phylogenetic tree was re-scaled so that the shortest distance from tip to root is equal to 1")
    tree$edge.length <- tree$edge.length/min(diag(C))
  }
  return(tree)
}
if(method!="hill")
  tre4 <- as(reduceTree(as(tre4, "phylo")), "phylo4")
a <- edgeLength(tre4)

```

```

b <- a[getEdge(tre4, rootNode(tre4))]
if(is.na(b)){
ab <- a
ab[getEdge(tre4, rootNode(tre4))] <- 0
edgeLength(tre4) <- ab
}
tre4 <- subset(tre4, tips.exclude=tipLabels(tre4)[!tipLabels(tre4)%in%colnames(m)])

des <- lapply(as.vector(nodeLabels(tre4)), function(x) names(descendants(tre4, x, type="tips")))
des <- lapply(des, function(x) x[x%in%colnames(m)])
fun <- function(namestips){
  return(rowSums(m[, namestips]))
}
abundancesnodes <- cbind.data.frame(lapply(des, fun))
mBabtot <- cbind(abundancesnodes, m)
colnames(mBabtot) <- c(nodeLabels(tre4), colnames(m))

branchlengths <- getEdge(tre4, colnames(mBabtot), missing = "OK")

branchlengths <- edgeLength(tre4)[branchlengths]

if(any(is.na(branchlengths))) stop("the lengths of some branches are missing in the phylogenetic tree; note that
lengths of zero are allowed")

tab1 <- mBabtot
tab2 <- (t(t(mBabtot)*branchlengths))

composition <- as.data.frame(sweep(tab1, 1, rowSums(tab2), "/"))
tre.phylo <- as(tre4, "phylo")
V <- diag(vcv(tre.phylo, model = "Brownian"))
names(V) <- tre.phylo$tip.label
mV <- m[, names(V)]
if(option==1)
  CC <- apply(mV, 1, function(x) sum(V[x>0]))
if(option==2)
  CC <- apply(mV, 1, function(x) length(x[x>0])*max(V[x>0]))
if(option==3){
  if(is.null(H)){
    warning("Parameter H, a numeric, must be specified if option = 3; option 2 was used here")
    CC <- apply(mV, 1, function(x) length(x[x>0])*max(V[x>0]))
  }
  else
    CC <- apply(mV, 1, function(x) length(x[x>0])*H)
}
HCDT <- function(x, q){
  funHCDT <- function(y, q) {
    b <- branchlengths[y>0]
    y <- y[y>0]
    if(abs(q-1) < tol){
      resi <- -sum(b*y*log(y))
    }
    else{
      resi <- (1-sum(b*y^q))/(q-1)
    }
    return(resi)
  }
}
res <- sapply(as.data.frame(t(x)), funHCDT, q)

```

```

    return(res)
  }
  hill <- function(x, q){
    funhill <- function(y, q) {
      b <- branchlengths[y>0]
      y <- y[y>0]
      if(abs(q-1) < tol){
        resi <- exp(-sum(b*y*log(y)))
      }
      else{
        resi <- (sum(b*y^q))^(1/(1-q))
      }
      return(resi)
    }
    res <- sapply(as.data.frame(t(x)), funhill, q)
    return(res)
  }
  renyi <- function(x, q){
    funrenyi <- function(y, q) {
      b <- branchlengths[y>0]
      y <- y[y>0]
      if(abs(q-1) < tol){
        resi <- -sum(b*y*log(y))
      }
      else{
        resi <- log((sum(b*y^q))^(1/(1-q)))
      }
      return(resi)
    }
    res <- sapply(as.data.frame(t(x)), funrenyi, q)
    return(res)
  }
  funq <- function(q){
    if(method == "HCDT"){
      T <- HCDT(composition, q)
      if(abs(q-1) < tol){
        vres <- T/log(CC)
      }
      else{
        vres <- T/(1-CC^(1-q))*(q-1)
      }
      vres[T < tol] <- 0
      return(vres)
    }
    if(method == "hill"){
      vres <- hill(composition, q)/CC
      return(vres)
    }
    if(method == "renyi"){
      R <- renyi(composition, q)
      vres <- R/log(CC)
      vres[R < tol] <- 0
      return(vres)
    }
  }
  if(length(q)==1){
    v <- funq(q)
  }

```

```

class(v) <- "evoeveparam"
return(v)
}
if ( length(q) > 1){
  calcul1 <- sapply(q, funq)
  tab1 <- cbind.data.frame(calcul1)
  listtotale <- list()
  listtotale$q <- q
  listtotale$eve <- tab1
  class(listtotale) <- "evoeveparam"
  return(listtotale)
}
}

```

evouniparam <-

```

function(phy1, comm, method = c("hill", "HCDT", "renyi"),
  q = 2, tol = 1e-8){

```

```

  tre <- .checkphyloarg(phy1)
  phy11 <- tre$phyl.phylo
  A <- write.tree(phy11)
  if(!substr(A, nchar(A)-1, nchar(A))=="");
  phy11 <- read.tree(text=paste0("(", substr(A,1,nchar(A)-1), ");"))
  method <- method[1]
  if(method=="HCDT") method <- "tsallis"
  reduceTree <- function(tree){
    C <- vcv(tree, model = "Brownian")
    if(min(diag(C))<1){
      warning("The phylogenetic tree was re-scaled so that the shortest distance from tip to root is equal to 1")
      tree$edge.length <- tree$edge.length/min(diag(C))
    }
    return(tree)
  }
  if(method!="hill")
  phy11 <- reduceTree(phy11)

```

```

  hi <- diag(vcv(phy11, model = "Brownian"))

```

```

  phylstar <- starTree(phy11$tip.label, hi)

```

```

  if(length(q)==1){

```

```

    vnum <- evodivparam(phy11, comm, method, q, tol)

```

```

    vden <- evodivparam(phylstar, comm, method, q, tol)

```

```

    v <- vnum/vden

```

```

    v[vnum < tol] <- 0

```

```

    class(v) <- "evouniparam"

```

```

    return(v)
  }

```

```

  if ( length(q) > 1){

```

```

    tabnum <- evodivparam(phy11, comm, method, q, tol)$div

```

```

    tabden <- evodivparam(phylstar, comm, method, q, tol)$div

```

```

    tab1 <- as.matrix(tabnum)/as.matrix(tabden)

```

```

    tab1[as.matrix(tabnum) < tol] <- 0

```

```

    listtotale <- list()

```

```

    listtotale$q <- q

```

```

    listtotale$uni <- as.data.frame(tab1)

```

```

    class(listtotale) <- "evouniparam"

```

```

    return(listtotale)
  }

```

```
}  
}
```

```
plot.evoeveparam <-  
function(x, legend = TRUE, legendposi = "topright", type="b", col = if(is.numeric(x)) NULL else  
sample(colors(distinct = TRUE), nrow(x$eve)), lty = if(is.numeric(x)) NULL else rep(1, nrow(x$eve)), pch =  
if(is.numeric(x)) NULL else rep(19, nrow(x$eve)), ...)  
{
```

```
  if(is.numeric(x)){  
    y <- as.vector(x)  
    names(y) <- names(x)  
    dotchart(y, xlab = "Evenness", ...)  
  }
```

```
  if(is.list(x)){  
    if(length(col)==1) col <- rep(col, nrow(x$eve))  
    if(length(pch)==1) pch <- rep(pch, nrow(x$eve))  
    plot(x$q, x$eve[1, ], type = type, col = col[1], ylim = c(min(x$eve), max(x$eve)), pch = pch[1], ,  
ylab="Evenness", xlab="q", ...)  
    for(i in 1:nrow(x$eve)){  
      lines(x$q, x$eve[i, ], type = type, col = col[i], pch = pch[i], ...)  
    }  
    if(legend[1]){  
      legend(legendposi, legend = rownames(x$eve), col = col, lty = lty, pch = pch, ...)  
    }  
  }  
}
```

```
}
```

```
plot.evouniparam <-  
function(x, legend = TRUE, legendposi = "topright", type="b", col = if(is.numeric(x)) NULL else  
sample(colors(distinct = TRUE), nrow(x$uni)), lty = if(is.numeric(x)) NULL else rep(1, nrow(x$uni)), pch =  
if(is.numeric(x)) NULL else rep(19, nrow(x$uni)), ...)  
{
```

```
  if(is.numeric(x)){  
    y <- as.vector(x)  
    names(y) <- names(x)  
    dotchart(y, xlab = "Uniqueness", ...)  
  }
```

```
  if(is.list(x)){  
    if(length(col)==1) col <- rep(col, nrow(x$uni))  
    if(length(pch)==1) pch <- rep(pch, nrow(x$uni))  
    plot(x$q, x$uni[1, ], type = type, col = col[1], ylim = c(min(x$uni), max(x$uni)), pch = pch[1], ,  
ylab="Uniqueness", xlab="q", ...)  
    for(i in 1:nrow(x$uni)){  
      lines(x$q, x$uni[i, ], type = type, col = col[i], pch = pch[i], ...)  
    }  
    if(legend[1]){  
      legend(legendposi, legend = rownames(x$uni), col = col, lty = lty, pch = pch, ...)  
    }  
  }  
}
```

```
}
```

```
.checkphyloarg <- function(phy){  
  if(inherits(phy,"phylo4d"))
```

```

  phyl <- extractTree(phyl)
if(inherits(phyl,"phylo4"))
  phyl.phylo <- as(phyl,"phylo")
if(inherits(phyl,"phylo")){
  phyl.phylo <- phyl
  phyl <- as(phyl.phylo,"phylo4")
}
if(inherits(phyl, "hclust")){
  phyl.phylo <- as.phylo(phyl)
  phyl <- as(phyl.phylo,"phylo4")
}
if(!exists("phyl.phylo"))
  stop("unconvenient phyl: must be a hclust, phylo, phylog, phylo4 or phylo4d object")

if(!isRooted(phyl)){
  phyl.phylo$root.edge <- 0
  phyl <- as(phyl.phylo, "phylo4")
}

phyl <- reorder(phyl)
phyl.phylo <- as(phyl, "phylo")

res <- list(phyl = phyl, phyl.phylo = phyl.phylo)

}

```