Journal of Vegetation Science **26** (2015) 839–847

Let the concept of indicator species be functional!

Carlo Ricotta, Marta Carboni & Alicia T.R. Acosta

Keywords
coastal dunes; diagnostic species; ecological value; fidelity; indicator value; plant functional traits

Nomenclature
Conti et al. (2005)

Received 29 July 2014
Accepted 8 February 2015
Co-ordinating Editor: Norman Mason

Ricotta, C. (carlo.ricotta@uniroma1.it)[1],
Carboni, M. (corresponding author, marta.carboni@gmx.net)[2],
Acosta, A.T.R. (acosta@uniroma3.it)[3]

[1]Department of Environmental Biology, University of Rome 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy;
[2]Laboratoire d'Ecologie Alpine, Unité Mixte de Recherche 5533 CNRS–Université Joseph Fourier, Grenoble, France;
[3]Department of Sciences, University 'Roma Tre', Viale Marconi 446, 00146 Roma, Italy

Abstract

Aims: The identification of diagnostic or indicator species with high fidelity to a given group of sites is an important step for the ecological characterization of habitats or community types. The determination of the degree of fidelity to a target group is traditionally performed by analysing the concentration of species occurrences or abundances in different groups of sites. Surprisingly, although one of the main purposes of indicator species analysis is to give ecological meaning to groups of sites, none of the methods proposed to date take into account the functional ecology of diagnostic species. Therefore, the question we address here is: can we use functional traits of species to improve the diagnostic value of indicator species?

Location: Sand dune communities in central Italy.

Methods: In this paper we propose a two-step procedure for incorporating the functional traits of a given species in the evaluation of its diagnostic value. For a given set of plots that are classified into different groups, first the indicator species that best characterize each group of plots are identified with the usual statistical tools based on species occurrences. Next, the functional association between the indicator species and the target groups of plots is tested by measuring the functional distance between the indicator species and the centroids of all plots in the target group. A species is positively associated with a group if its mean functional distance from all plot centroids in the group is significantly lower than expected.

Results: In this example, we show that the functional association of the indicator species with a given habitat type is represented by less species than the association highlighted solely through species occurrences and/or abundances. This subset of species appears to better characterize the functional ecology of coastal dune plant assemblages and shows a higher diagnostic value in comparison with those obtained through the traditional indicator analysis.

Conclusions: As functional traits are the main ecological attributes by which different species influence ecosystem processes, we believe that the methodology proposed here provides a relevant tool for ecological applications as distinct as vegetation science, conservation biology or landscape management.

Introduction

An important step for the ecological characterization of community types is the identification of diagnostic or indicator species with high fidelity to a given group of sites (or plots, reléves, sampling units, etc.). Determination of indicator species is usually performed by analysing the concentration of species occurrences or abundances in distinct groups of sites (Dufrêne & Legendre 1997; McGeoch 1998; Chytrý et al. 2002). Species with a high fidelity to a given group of sites relative to the other groups of sites are considered as indicator species of that group (Tichý & Chytrý 2006; De Cáceres et al. 2010). The classification of sites into groups may have been obtained from any hierarchical or non-hierarchical classification of the species composition within sites, or from the environmental conditions of sites. The groups may also reflect other qualitative characteristics as diverse as successional stages, land-use types, repeated surveys of permanent plots or different levels of controlled experimental designs (Dufrêne & Legendre

Doi: 10.1111/jvs.12291

1997; De Cáceres et al. 2012). While indicator species have been extensively used for characterizing plant communities belonging to different hierarchical levels of a vegetation classification (Barkman 1989), the concept of indicator species is not exclusive to vegetation science and can be applied in all branches of community ecology for the characterization of plant or animal assemblages (Tichý & Chytrý 2006).

Several statistical measures exist for the identification of indicator species (e.g. Jancey 1979; Dufrêne & Legendre 1997; Bruelheide 2000; Chytrý et al. 2002; Podani & Csányi 2010; Chazdon et al. 2011; Urban et al. 2012). Among them, the most frequently used measures of the association between species and groups of sites involve correlation or indicator value indices (De Cáceres et al. 2012). For a review, see Chytrý et al. (2002) and De Cáceres & Legendre (2009). Correlation indices assess the (positive or negative) preference of a given species for a given group of sites, compared with the complementary portion of the set of sites. The simplest correlation index is the phi coefficient of association for species presence/absence data (r_{Φ}). The phi coefficient measures the Pearson correlation between two binary (0/1) vectors: a vector of presences and absences of the target species within each site, and a vector of membership/non-membership of each site to the target group (Chytrý et al. 2002).

As an alternative, Dufrêne & Legendre (1997) introduced a composite index called IndVal (Indicator Value) to measure to what extent the sites of a target group match the sites where a given species is found (De Cáceres & Legendre 2009). The index IndVal is the product of two terms, specificity and fidelity (Dufrêne & Legendre 1997). Specificity, which can be calculated from either presence–absence or abundance data, measures the probability that a given plot j belongs to a target group G in case species *i* has been encountered (De Cáceres et al. 2012; Wildi & Feldmeyer-Christe 2013). This quantity is also called the 'positive predictive value' of species *i*. The second part of the index, called fidelity, is the proportion of plots in which species *i* is present within group G, thus expressing the chance that the species can be detected in the field in terms of presences and absences within group G (Podani & Csányi 2010).

In order to determine if a species is significantly associated with a target group of plots we must compare the observed index value with a distribution obtained under the null hypothesis that there is no such relationship. In traditional indicator value analysis, the statistical significance of species-group associations is usually tested using a permutation test that randomly re-orders the species occurrence or abundance data among all plots, such that the fact that the species is found in a given plot is due to chance alone (De Cáceres et al. 2012).

Surprisingly, although the aim of indicator species analysis is to "give ecological meaning to groups of sites" (Dufrêne & Legendre 1997: p. 345), to the best of our knowledge, none of the methods proposed to date take into account the functional traits of individual species (but see Moretti et al. 2010). As functional traits are the main ecological attributes by which different species influence ecosystem processes (de Bello et al. 2010), the aim of this paper is to propose a two-step procedure for incorporating the functional traits of a given species in the evaluation of its diagnostic value (Fig. 1). For a given set of plots that are classified into different groups, the first step consists in identifying the indicator species that best characterize each group of plots with the usual statistical tools. Next, the functional association between the indicator species and the target groups of plots is tested by measuring the functional distance between the indicator species and the centroids of all plots in the target group. The proposed method is illustrated using a dedicated example from a pre-existing data set on sand dune communities in central Italy.

Data

We analysed plant communities on sand dunes along the Tyrrhenian coast of Italy (ca. 250 km in the Latium region;

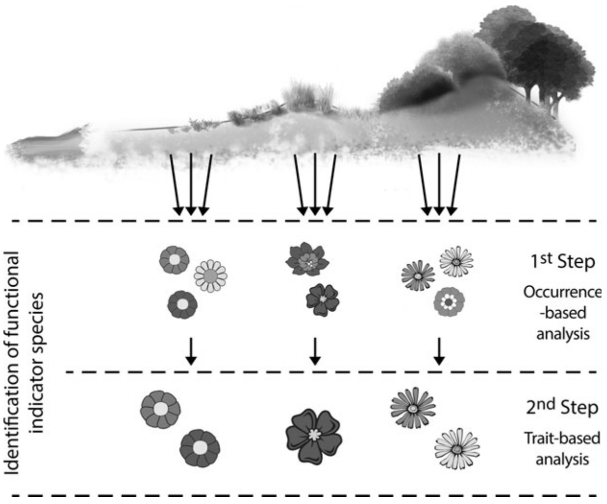

Fig. 1. Conceptual figure illustrating the two-step procedure for incorporating the functional traits of species in the evaluation of their diagnostic value. In the first step, the indicator species of three hypothetical habitat types along a sea–inland gradient are identified based on their occurrence within plots of the habitats as in traditional indicator species analysis. In the second step, a subset of the species identified in the first step is selected as functional indicator species based on their traits and on the mean functional distance to the plots in the habitat. The trait-based step allows screening out species that are not sufficiently functionally representative of the group (e.g. the white flower in the habitat closest to the sea, 1st step line).

Doi: 10.1111/jvs.12291

Fig. 2). These communities occur along a clearly defined stress gradient that drives considerable turnover in species composition and physiognomy over short distances. The vegetation follows a compressed zonation along the sea–inland environmental gradient: from the pioneer communities of the upper beach to the woody communities of the inland fixed dunes (Acosta et al. 2003). We used a georeferenced vegetation database including most of the best-conserved remnant dune systems of the region (Carboni et al. 2013). The database contains 564 randomly sampled plots of 2 m × 2 m in size collected in spring (April–June) between 2004 and 2009. In each plot, all vascular plant species were recorded and the cover of all species was visually estimated using a 10% interval scale (for further details on sampling and nomenclature see Carboni et al. 2011).

The plots were hierarchically classified using the abundance data of all species through cluster analysis using PC-ORD (MjM Software, Gleneden Beach, OR), with group average as linkage method and relative Euclidean as distance measure (see Santoro et al. 2012 for details). Next, according to the guidelines of Biondi et al. (2009), the resulting clusters were assigned to the coastal habitat types included in the Habitats Directive of the Council of European Communities 92/43/EEC (EEC 1992). For this study we selected 225 plots covering three of the most typical coastal dune habitats (Table 1): (i) embryo dunes (Habitat code: 2110 – Embryonic shifting dunes; 37 plots), pioneer halophilous assemblages with dune-forming plants; (ii) mobile dunes (Habitat code: 2120 – Shifting dunes along the shoreline with *Ammophila arenaria*; 120 plots), perennial herb assemblages dominated by rhizomatous tussock grasses, and (iii) transition dunes (Habitat codes: 2210 – *Crucianellion maritimae* fixed beach dunes and 2230 *Malcolmietalia* dune grasslands; 68 plots), perennial herb assemblages, partially sheltered from winds, dominated by chamaephytic species but often intermingled with annual grasses. Both communities usually grow in a mosaic pattern on transition dunes, so we decided to consider them as a single group. Functional traits were measured for a subset of 46 dominant species present in the analysed plots (see Appendix S1). Collectively, these selected species account for ~80% of the standing live biomass in each dune habitat. This threshold has been shown to ensure a satisfactory description of overall community properties (Pakeman & Quested 2007). Seven quantitative life-history traits were selected based on previous studies according to their relevance to the functional ecology of plants in coastal dune environments and which reflect as best as possible the environmental differences between the different habitats (Ricotta et al. 2012; Carboni et al. 2013; Jucker et al. 2013): plant height (cm), specific leaf area (SLA; $\mathrm{mm^2 \cdot mg^{-1}}$), leaf dry mass content (LDMC; $\mathrm{mg \cdot g^{-1}}$), leaf size ($\mathrm{cm^2}$), leaf thickness (mm), seed mass (mg) and seed shape (variance of the three main dimensions according to Thompson et al. 1993). These traits were measured in the field or laboratory by taking the mean of ten different individuals for each species (for details see Carboni et al. 2013). SLA and LDMC are related to plant investment in leaf defences and to leaf life span, leaf thickness is related to species' strategies of resource acquisition and use, plant height is a good measure of the plant's

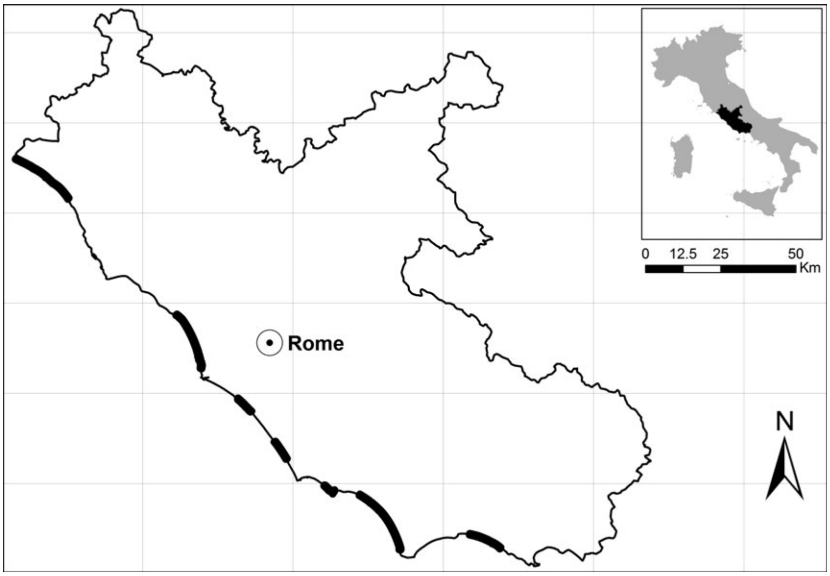

Fig. 2. Location of the dune habitats used in this study (Latium Region, Central Italy).

Table 1. Vegetation types used in this study: position along the coastal dune profile, vegetation structure and habitat code according to the Directive of the Council of European Communities 92/43/EEC on the conservation of natural habitats (EEC 1992).

Distribution along the coastal dune zonation	Vegetation structure	Habitat code according to the Directive 92/43/EEC
Embryo dunes	Pioneer perennial assemblages dominated by dune-forming plants	2110 – Embryonic shifting dunes
Mobile dunes	Perennial herbs and grasses dominated by rhizomatous species	2120 – Shifting dunes along the shoreline with *Ammophila arenaria*
Transition dunes	Perennial herb assemblages, dominated by chamaephytic species forming a mosaic with annual grasses	2210 – *Crucianellion maritimae* fixed beach dunes 2230 – *Malcolmietalia* dune grasslands

competitive strength as it is an indirect proxy of the ability to compete for light, while seed traits correlate with the plant's dispersal ability (Weiher et al. 1999; Cornelissen et al. 2003; Vile et al. 2005). Because of environmental sorting of species according to habitat preferences, we expect that dominant species with particular suites of traits (e.g. short annual species with succulent leaves on the upper beach) should gradually be replaced along the gradient by other dominant species with different traits (e.g. tall perennial rhizomatous species on the main dune ridge; see for example Carboni et al. 2013).

Methods

First, we measured the functional distinctness of the three groups of plots corresponding to the habitats of embryo dunes, mobile dunes and transition dunes. We performed a permutational multivariate ANOVA on the Bray-Curtis dissimilarity matrix between the community-weighted mean trait values (CWMs – see below for details on the calculation of the CWMs) of each pair of plots (999 randomizations, one-tailed test; see Anderson 2001).

Next, we tested, using the 46 species with trait data, whether they were significantly associated with any of the three pre-defined groups of plots. As in traditional indicator species analysis, the first step consists in identifying the indicator species that best characterize each group of plots (Fig. 1). Here, indicator species are defined as those species that are more common in a given group of plots than expected from a random null model in which all plots have equal probability to host each species, irrespective of the species environmental preferences. Accordingly, for species presence and absence data, to determine if a given species is significantly associated with a target group of plots, the number of presences of that species in each group of plots was compared with a null distribution obtained from 999 permutations, in which the species occurrence values are randomly reassigned to all plots. The null hypothesis is that the presence of a species within a plot belonging to a given target group is due to chance alone (De Cáceres & Legendre 2009). The P-value of the permutation test of positive species association with a target group of plots was then calculated as the proportion of permutation-derived values of species occurrences that were as high or higher than the actual value (De Cáceres et al. 2010).

Note that the reason for using the number of occurrences of a given species within the target group for summarizing its diagnostic value is because of the observation that for species presence and absence data, the only quantity that is affected by the permutation procedure is the number of species occurrences within the target group of plots. This renders the number of species occurrences equivalent to more complex statistics for testing whether the species is more tightly associated to the target group of plots than to other plots (De Cáceres & Legendre 2009).

Compared to the analysis of species presence and absence data, there are many more possible algorithms and constraints that can be used to randomize the abundances of a given species (Ulrich & Gotelli 2010). Among the many possible options, we used a so-called 'floating-zero algorithm' for permuting species abundances among plots, allowing for the placement of non-zero abundance values in plots in which the target species was originally absent. P-values of positive species association with a target group of plots was then calculated as the proportion of permutation-derived values of mean species abundance within the target group that were as high or higher than the actual value (999 permutations, one-tailed test).

After having identified the indicator species that best characterize each group of plots based either on presence and absence scores or on abundance data, the second step consists in testing the functional association between the indicator species and the target group of plots (Fig. 1). Here, we considered a species as positively associated to a target group of plots if its mean functional distance from all plot centroids in the target group is significantly lower than expected by chance. Hence, we measured the functional distance between each indicator species and the centroids of all 225 plots. For all functional traits we calculated the community-weighted mean trait values (CWMs) at each plot, defined as the mean of all trait values present in a given plot weighted by the relative abundance of the species having each value (Garnier et al. 2004):

$$\mathrm{CWM} = \sum_{i=1}^{S} p_i \times \tau_i$$

where τ_i is the value of trait τ for species i ($i = 1, 2, \ldots, S$) and p_i is the relative abundance of species i. Before calculations, traits were log-transformed when necessary and scaled to unit length. We then calculated the Bray-Curtis dissimilarity between the trait values of the indicator species and the corresponding CWMs at each plot (i.e. the multivariate centroid of each plot). Finally, we used a randomization test to assess whether the mean functional distance between each indicator species and the corresponding target group of plots was significantly lower than expected by chance alone. For each indicator species we permuted the functional distances from plot centroids among all 225 plots, which is equivalent to permuting all plots among the three target groups, or habitat types. *P*-values of positive functional association between a given indicator species and the corresponding target group of plots were then calculated as the proportion of permutation-derived mean distance values that are as low or lower than the actual value (999 permutations, one-tailed test). All calculations were performed in the R statistical framework (R Foundation for Statistical Computing, Vienna, AT). A new R function, called 'FuncVal', to perform these calculations is available in the electronic supplementary material (Appendix S2).

Results

The permutational ANOVA highlighted the functional uniqueness of the three groups of plots ($F = 33.54$, $P < 0.01$), meaning that the habitat classification used for analysing the coastal vegetation of Latium is functionally sustained.

The indicator species that were significantly associated with the three groups of plots (habitat types) sampled on the coast of Latium are shown in Table 2. Among the 46 coastal dune species, 28 showed a significant association with one group of plots ($P < 0.01$) based either on presence and absence data, on species abundances or both. In this view, looking at the species abundance values, the results obtained are in very good agreement with those obtained from species presence and absence data, meaning that, on average, the species with a significantly higher than expected species occurrence within a target group of plots also tend to have higher than expected abundances within the same group. This is not really surprising, as species abundances are not completely independent of species occurrences, especially if there is a strong difference in the occurrence of a given species among groups.

In contrast, looking at the species mean functional distances from the plot centroids of each group, only 17 species showed significant functional association ($P < 0.01$) with the three habitat types. *Cakile maritima, Chamaesyce peplis* and *Salsola kali* showed a significant preference for the embryo dune vegetation; *Ammophila arenaria, Echinophora spinosa, Elymus farctus* and *Eryngium maritimum* showed a significant preference for the mobile dunes; while the remaining ten species were significantly associated with the transition dunes. Thus, it seems that these habitat types may be functionally characterized by a reduced number of species, which adequately reflect the functional ecology of species assemblages in coastal dune environments (Feola et al. 2011).

Discussion

While indicator or diagnostic species are widely used in plant ecology, both for classifying the vegetation (e.g. by the cocktail approach; Bruelheide 2000; Tichý 2002; Roleček 2007), and for the *a-posteriori* ecological characterization of a given classification, our proposal is unequivocally related only to the second case. In this framework, indicator species are usually considered diagnostic of particular habitats, communities or environmental conditions. In this paper we suggested improving their diagnostic value by taking into account not only their abundances, but also their mean functional distances from the target groups of plots. In our worked example, the functional association of the indicator species with a given habitat type is always represented by less species than the association highlighted solely through species occurrences and/or abundances. This subset of species appeared to better characterize the functional ecology of coastal dune plant assemblages. For example, most of the species that showed a significant functional species–habitat association, such as *Cakile maritima, Chamaesyce peplis, Salsola kali, Ammophila arenaria, Echinophora spinosa, Elymus farctus, Eryngium maritimum, Crucianella maritima* and *Vulpia fasciculata*, are also considered diagnostic species for the selected habitats *sensu* the Habitat Directive 92/43/EEC, according to the Italian interpretation manual (see Biondi et al. 2009).

Note that in this study, the transition dunes habitat contained a mosaic of two structurally and floristically different plant communities, probably biasing the functional species–habitat association. Note also that in our study the plots have been classified based on their species composition. Therefore, the abundance-based indicator species analysis relies to some extent on the same data used for the classification of plots, and *P*-values must be interpreted with care. In this case, the indicator value statistics will be larger than the values expected under the null hypothesis of independence, leading to inflated type I errors (De

Table 2. Expected vs observed values of the number of occurrences, mean abundance values and mean functional distances from plot centroids in the three groups of plots sampled on the Holocenic dunes of the Tyrrhenian coast of Latium (central Italy). Only the indicator species that are significantly associated with at least one target group of plots based either on presence and absence scores or on abundance data are shown. NS = not significant at $P = 0.01$ (999 permutations, one-tailed test).

Species	Number of occurrences		Mean abundance value		Mean functional distance from plot centroids	
	Expected	Observed	Expected	Observed	Expected	Observed
Embryo dunes (37 plots)						
Cakile maritima	10.152	32	0.391	1.351	0.139	0.126
Chamaesyce peplis	6.424	26	0.216	0.838	0.302	0.285
Otanthus maritimus	1.664	6	0.098	0.432	0.140	0.150$^{\mathrm{NS}}$
Polygonum maritimum	2.300	9	0.077	0.324	0.142	0.132$^{\mathrm{NS}}$
Salsola kali	10.947	33	0.397	1.378	0.326	0.267
Mobile dunes (120 plots)						
Ammophila arenaria	25.488	44	0.793	1.450	0.222	0.196
Anthemis maritima	73.593	94	1.127	1.633	0.152	0.155$^{\mathrm{NS}}$
Echinophora spinosa	22.448	30	0.239	0.350	0.311	0.290
Elymus farctus	77.450	87	1.198	1.508	0.173	0.156
Eryngium maritimum	17.563	27	0.161	0.250	0.222	0.202
Medicago marina	22.499	32	0.371	0.583	0.175	0.191$^{\mathrm{NS}}$
Ononis variegata	32.081	39$^{\mathrm{NS}}$	0.589	0.867	0.262	0.282$^{\mathrm{NS}}$
Transition dunes (68 plots)						
Bromus diandrus	7.143	20	0.183	0.500	0.195	0.172
Centaurea sphaerocephala	4.940	10	0.123	0.235$^{\mathrm{NS}}$	0.132	0.142$^{\mathrm{NS}}$
Clematis flammula	4.216	11	0.114	0.279	0.185	0.187$^{\mathrm{NS}}$
Crucianella maritima	11.862	23	0.529	1.250	0.217	0.166
Cutandia maritima	22.339	36	0.678	1.456	0.151	0.124
Euphorbia terracina	6.545	19	0.150	0.441	0.264	0.219
Helichrysum stoechas	3.952	9	0.156	0.426	0.195	0.154
Juniperus oxycedrus	1.471	5	0.035	0.118	0.320	0.320$^{\mathrm{NS}}$
Lagurus ovatus	4.139	11	0.120	0.338	0.384	0.372
Lotus cytisoides	9.855	27	0.311	0.971	0.206	0.158
Medicago littoralis	12.746	27	0.518	1.265	0.239	0.194
Pistacia lentiscus	1.788	5$^{\mathrm{NS}}$	0.040	0.118	0.262	0.279$^{\mathrm{NS}}$
Pycnocomon rutifolium	7.162	21	0.223	0.662	0.102	0.120$^{\mathrm{NS}}$
Silene canescens	25.550	42	0.797	1.382	0.196	0.168
Smilax aspera	4.217	10	0.066	0.147	0.205	0.214$^{\mathrm{NS}}$
Vulpia fasciculata	12.957	33	0.401	1.132	0.247	0.234

Cáceres & Legendre 2009; De Cáceres et al. 2010). On the contrary, unless the species functional characters are explicitly used for classifying the plots, the functional indicator analysis is less affected by the non-independence of the target groups from species data, thus leading to a less inflated rate of rejection in inferential tests.

Irrespective of the inflated type I error rate, indicator species based on mean functional distances seem to show a higher diagnostic value in comparison with those obtained through the traditional indicator analysis. In fact, the proposed two-step procedure allowed us to discriminate a set of species, such as *Clematis flammula*, *Juniperus oxycedrus*, *Pistacia lentiscus* and *Smilax aspera*, that have their main ecological optimum outside this data set, being typically associated to more interior habitats on fixed dunes. Since they are also occasionally present on the adjacent transition dunes, these species were identified as significantly associated to this habitat type, based on species occurrences and/or abundances, but not on their functional characters (Table 2).

Note also that the algorithm used for testing the functional concordance between a given species and the target groups of plots implies that the functional distances between the species and the plot centroids are also calculated for those plots in which the target species was originally absent. This is because an indicator species is considered diagnostic of the functioning of all plots in the target group, and not only of the plots in which the species is present. A species can be absent because the sampling method may be unable to detect it or because it was actually not there. Also, as species are never completely in equilibrium with their environment (Pulliam 2000), a species can be absent from suitable habitats for accidental reasons or due to dispersal limitations (Holt 2003).

On the other hand, one might argue that there is no reason to limit the functional indicator analysis to those species that were determined to be diagnostic of a particular habitat type based on classical indicator species analysis. However, we see two main reasons for a restriction based on abundance concentration: (i) species rarity and (ii) incomplete functional information.

First, indicator species analysis is used at the same time to provide information on the target species and on the target site group. While rare species are usually not homogeneous in terms of their ecological requirements (Grime 1998; Mouillot et al. 2013), some of them could be well adapted to the environmental conditions of a given group of plots (i.e. they could be functionally very close to the plot centroids of the target site group) without being diagnostic of that habitat type, based either on species occurrences or abundances. That is, the species may be very close functionally to a target site group but cannot be used to 'identify' the group. Nevertheless, we believe that to consider a species as 'functionally diagnostic' of a particular habitat, the species should possess a reasonable chance of being detected in the field. Therefore, its indicator value should be best analysed in terms of both the functional association with the target group of plots *and* its abundance concentration within the plots (as in the two-step procedure presented here).

Second, and more concerning, the functional traits that are commonly available in the literature or sampled in the field (i.e. 'soft traits' *sensu* Hodgson et al. 1999; Weiher et al. 1999) may not necessarily capture all facets of the ecological functioning of a given vegetation type. If this is the case, we are likely to only partially discriminate plant strategies with the available functional data, which may result in spurious functional similarities between some species and a target group of plots. In other words, they may appear similar, but only because we did not capture all functional differentiations. For example, shrubby sclerophyllous plants on one hand and grasses on the other may be similar in terms of LDMC but we will clearly differentiate them functionally if we also account for height. Similarly, even though we accounted for a number of traits, in our worked example, some of the species typically found on embryo dunes, such as *Chamaesyce peplis* and *Polygonum maritimum* (see Table 1), turn out to be functionally close to the transition dune assemblages (see Appendix S3). This example helps in clarifying the role of our proposal as a method to complement traditional indicator species analysis with an analysis focusing on functional aspects, rather than as a stand-alone method. To avoid such unexpected associations, the set of functional traits used would have to be even more comprehensive and carefully selected in order to reflect the environmental differences between the groups of plots and to maximize the between-group functional differences. This is very difficult for most vegetation types, and often the 'hard traits' allowing fine differentiation among distinct plant strategies are not easily available. Hence, we advocate that the functional association to groups of sites should always be taken into account only after that diagnostic species have been preselected through a traditional indicator species analysis.

More generally, as in the traditional indicator analysis based on abundance data, the functional approach should take into account that species vary in their niche breadths. This implies that the species may be diagnostic of more or less ample environmental conditions. In vegetation science, the varying breadth of environmental conditions is usually reflected in the different levels of a hierarchical classification. Accordingly, if the target groups of plots used in indicator analysis are obtained from a hierarchical classification, it might be helpful to perform indicator species analysis for the different hierarchies of the (vegetation) classification (see De Cáceres et al. 2010).

From a more technical viewpoint, the method used for calculating the plot centroids works only for quantitative variables or for ordinal variables transformed to ranks. This means that before calculation, the information of all nominal variables should be coded by as many independent binary variables as the number of categories. Alternatively, the non-exclusive categories of a nominal variable for which a species may be characterized by the simultaneous presence of two or more categories, such as dietary habits in animals or Grime (1974) CSR (Competitor–Stress tolerator–Ruderal) strategies in plants may be fuzzy-coded (Chevenet et al. 1994). In this case, each category receives a score in the range of 0–1 that represents the species affinity for that category. To avoid recoding the nominal variables, a different line of attack may consist in calculating a dissimilarity matrix between pairs of species using a dissimilarity coefficient for mixed data (see e.g. Gower 1971; Podani 1999; Pavoine et al. 2009). Then, the centroids of each plot can be obtained by applying a principal coordinates analysis to the dissimilarity matrix, as suggested in Legendre & Anderson (1999) and Anderson (2006).

Overall, we believe that the extended methodology proposed here for the identification of indicator species from a joint abundance-based and functional perspective can be used for a large range of plant traits and can be successfully applied for other vegetation types, from forest to grasslands, that show at least a moderate functional response to underlying environmental gradients. If the necessary precautions are taken by potential users (e.g. selection of traits accurately reflecting the environmental gradients of interest), we believe it provides a relevant tool for many ecological applications as different as vegetation science, conservation biology, landscape management or restoration ecology.

Acknowledgements

We thank Miquel De Cáceres and Florian Jansen for useful comments on a previous version of our manuscript. We also thank Norman Mason for helpful suggestions. They all contributed greatly to improve our paper.

References

Acosta, A., Stanisci, A., Ercole, S. & Blasi, C. 2003. Sandy coastal landscape of the Lazio region (Central Italy). *Phytocoenologia* 33: 715–726.

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

Anderson, M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62: 245–253.

Barkman, J.J. 1989. Fidelity and character-species, a critical evaluation. *Vegetatio* 85: 105–116.

de Bello, F., Lavergne, S., Meynard, C.N., Lěps, J. & Thuiller, W. 2010. The partitioning of diversity: showing Theseus a way out of the labyrinth. *Journal of Vegetation Science* 21: 992–1000.

Biondi, E., Blasi, C., Burrascano, S., Casavecchia, S., Copiz, R., Del Vico, E., Galdenzi, D., Gigante, D., Lasen, C., Spampinato, G., Venanzoni, R. & Zivkovic, L. 2009. *Manuale Italiano di interpretazione degli habitat della Direttiva 92/43/CEE (Italian Interpretation Manual of the 92/43/EEC Habitats Directive)*. URL: http://vnr.unipg.it/habitat/index.jsp2.

Bruelheide, H. 2000. A new measure of fidelity and its application to defining species groups. *Journal of Vegetation Science* 11: 167–178.

Carboni, M., Santoro, R. & Acosta, A.T.R. 2011. Dealing with scarce data to understand how environmental gradients and propagule pressure shape fine-scale alien distribution patterns on coastal dunes. *Journal of Vegetation Science* 22: 751–765.

Carboni, M., Acosta, A.T.R. & Ricotta, C. 2013. Are differences in functional diversity among plant communities on Mediterranean coastal dunes driven by their phylogenetic history? *Journal of Vegetation Science* 24: 932–941.

Chazdon, R.L., Chao, A., Colwell, R.K., Lin, S.Y., Norden, N., Letcher, S.G., Clark, D.B., Finegan, B. & Arroyo, J.P. 2011. A novel statistical method for classifying habitat generalists and specialists. *Ecology* 92: 1332–1343.

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

Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13: 79–90.

Cornelissen, J.H.C., Lavorel, S., Garnier, E., Diaz, S., Buchmann, N., Gurvich, D.E., Reich, P.B., ter Steege, H., Morgan, H.D., van der Heijden, M.G.A., Pausas, J.G. & Poorter, H. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany* 51: 335–380.

De Cáceres, M. & Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* 90: 3566–3574.

De Cáceres, M., Legendre, P. & Moretti, M. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* 119: 1674–1684.

De Cáceres, M., Legendre, P., Wiser, S.K. & Brotons, L. 2012. Using species combinations in indicator value analyses. *Methods in Ecology and Evolution* 3: 973–982.

Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.

EEC 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Community, Series L*. 206: 7–49.

Feola, S., Carranza, M.L., Schaminée, J.H.J., Janssen, J.A.M. & Acosta, A.T.R. 2011. EU habitats of interest: an insight into Atlantic and Mediterranean beach and foredunes. *Biodiversity and Conservation* 20: 1457–1468.

Garnier, E., Cortez, J., Billes, G., Navas, M.L., Roumet, C., Debussche, M., Laurent, G., Blanchard, A., Aubry, D., Bellmann, A., Neill, C. & Toussaint, J.P. 2004. Plant functional markers capture ecosystem properties during secondary succession. *Ecology* 85: 2630–2637.

Gower, J.C. 1971. General coefficient of similarity and some of its properties. *Biometrics* 27: 857.

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

Grime, J.P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology* 86: 902–910.

Hodgson, J., Wilson, P.J., Hunt, R., Grime, J.P. & Thompson, K. 1999. Allocating C-S-R functional types: a soft approach to a hard problem. *Oikos* 85: 282–294.

Holt, R.D. 2003. On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research* 5: 159–178.

Jancey, R.C. 1979. Species ordering on a variance criterion. *Vegetatio* 39: 59–63.

Jucker, T., Carboni, M. & Acosta, A.T.R. 2013. Going beyond taxonomic diversity: deconstructing biodiversity patterns reveals the true cost of iceplant invasion. *Diversity and Distributions* 19: 1566–1577.

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

McGeoch, M.A. 1998. The selection, testing and application of terrestrial insects as bioindicators. *Biological Reviews* 73: 181–201.

Doi: 10.1111/jvs.12291

Moretti, M., De Cáceres, M., Pradella, C., Obrist, M.K., Wermelinger, B., Legendre, P. & Duelli, P. 2010. Fire-induced taxonomic and functional changes in saproxylic beetle communities in fire sensitive regions. *Ecography* 33: 760–771.

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J. & Thuiller, W. 2013. Rare species support vulnerable functions in high-diversity ecosystems. *PloS Biology* 11: e1001569.

Pakeman, R.J. & Quested, H.M. 2007. Sampling plant functional traits: what proportion of the species need to be measured? *Applied Vegetation Science* 10: 91–96.

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

Podani, J. 1999. Extending Gower's general coefficient of similarity to ordinal characters. *Taxon* 48: 331–340.

Podani, J. & Csányi, B. 2010. Detecting indicator species: some extensions of the IndVal measure. *Ecological Indicators* 10: 1119–1124.

Pulliam, H.R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3: 349–361.

Ricotta, C., Pavoine, S., Bacaro, G. & Acosta, A.T.R. 2012. Functional rarefaction for species abundance data. *Methods in Ecology and Evolution* 3: 519–525.

Roleček, J. 2007. Formalized classification of thermophilous oak forests in the Czech Republic: what brings the Cocktail method? *Preslia* 79: 1–21.

Santoro, R., Jucker, T., Carboni, M. & Acosta, A.T.R. 2012. Patterns of plant community assembly in invaded and non–invaded communities along a natural environmental gradient. *Journal of Vegetation Science* 23: 483–494.

Thompson, K., Band, S.R. & Hodgson, J.G. 1993. Seed size and shape predict persistence in soil. *Functional Ecology* 7: 236–241.

Tichý, L. 2002. JUICE, software for vegetation classification. *Journal of Vegetation Science* 13: 451–453.

Tichý, L. & Chytrý, M. 2006. Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science* 17: 809–818.

Ulrich, W. & Gotelli, N.J. 2010. Null model analysis of species associations using abundance data. *Ecology* 91: 3384–3397.

Urban, N.A., Swihart, R.K., Malloy, M.C. & Dunning, J.B. 2012. Improving selection of indicator species when detection is imperfect. *Ecological Indicators* 15: 188–197.

Vile, D., Garnier, E., Midgley, G., Poorter, H., Rutherford, M.C., Wilson, P.J., Wright, I.J., Shipley, B., Laurent, G., Navas, M.L., Roumet, C., Lavorel, S., Díaz, S., Hodgson, J.G. & Lloret, F. 2005. Specific leaf area and dry matter content estimate thickness in laminar leaves. *Annals of Botany* 96: 1129–1136.

Weiher, E., van der Werf, A., Thompson, K., Roderick, M., Garnier, E. & Eriksson, O. 1999. Challenging Theophrastus: a common core list of plant traits for functional ecology. *Journal of Vegetation Science* 10: 609–620.

Wildi, O. & Feldmeyer-Christe, E. 2013. Indicator values (IndVal) mimic ranking by F-ratio in real-world vegetation data. *Community Ecology* 14: 139–143.

Supporting Information

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

Appendix S1. List of species used in this study.

Appendix S2. R function 'FuncVal' for the calculation of functional indicator species.

Appendix S3. List of species that are functionally associated to at least one target group of plots ($P < 0.01$) without being diagnostic of that group based on occurrence or abundance data.

Table S1. List of species used in this study. Nomenclature follows Conti et al. (2005).

Ammophila arenaria (L.) Link
Anthemis maritima L.
Asparagus acutifolius L.
Bromus diandrus Roth
Cakile maritima Scop.
Calystegia soldanella (L.) Roem. & Schult.
Centaurea sphaerocephala L.
Chamaesyce peplis (L.) Prokh.
Cistus creticus L.
Clematis flammula L.
Crucianella maritima L.
Cutandia maritima (L.) Barbey
Cyperus capitatus Vand.
Daphne gnidium L.
Echinophora spinosa L.
Elymus farctus (Viv.) Runemark ex Melderis
Eryngium maritimum L.
Euphorbia terracina L.
Helichrysum stoechas (L.) Moench
Juniperus oxycedrus L.
Lagurus ovatus L.
Lonicera implexa Aiton
Lotus cytisoides L.
Medicago littoralis Loisel.
Medicago marina L.
Ononis variegata L.
Otanthus maritimus (L.) Hoffmanns. & Link
Pancratium maritimum L.
Phillyrea angustifolia L.
Pistacia lentiscus L.
Plantago coronopus L.
Polygonum maritimum L.
Prasium majus L.
Pseudorlaya pumila (L.) Grande
Pycnocomon rutifolium (Vahl) Hoffmanns. & Link
Quercus ilex L.
Rhamnus alaternus L.
Rostraria litorea (All.) Holub
Rubia peregrina L.
Salsola kali L.
Silene colorata Poir.
Smilax aspera L.
Sonchus bulbosus (L.) N. Kilian & Greuter
Sporobolus virginicus Kunth
Teucrium flavum L.
Vulpia fasciculata (Forssk.) Fritsch

References

Conti, F., Abbate, G., Alessandrini, A., Blasi, C. 2005. An annotated checklist of the Italian vascular flora. Palombi Editore, Roma.

```
#########################################################################
#######
#Function to calculate Functional Indicator Value of species for groups
of plots.
#
#########################################################################
#######

#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.□□#Description:
#given a matrix of N species abundances × M plots, together with a factor
with group belongings for M plots, and either a matrix of T traits x N
species or a matrix of T Community weighted Means x M plots, this
function calculates the Functional Indicator Value of species for groups
of plots as described in the main text, and assess the significance of
the association with the groups based on randomizations. □□#Dependencies:
#FD, vegan, picante

#Usage:
# FuncVal (samp, groups, traits, cwms=NULL, nrep=99, dist.method="bray",
null.model = "richness")

#Arguments:
#samp: Community data matrix
#groups: factor with group belonging for each plot in samp
#traits: Traits x Species matrix
#cwms: community weighted means matrix (optional)
#null.model: Null model to use (see ?randomizeMatrix). Default =
"richness" (maybe the only one which makes sense)
#dist.method= Distance method to calculate species-to-plot multivariate
functional distance (see ?vegdist). Default= "bray"
#nrep: Number of randomizations

#Returns:
#A 3-dimensional array in which rows are species, columns relate to the
mean functional distance from plot centroids (column1=observed values,
column 2=mean of randomized values, column3=p-value), on the third
dimension are the different groups of plots.

#########################################################################
########
require(FD)           #functcomp
require(vegan)        #vegdist
require(picante) #randomize matrix


FuncVal <- function(samp, groups, traits, cwms=NULL, nrep=99,
dist.method="bray", null.model = "richness"){

    #calculate cwms if not given
    if(is.null(cwms)) cwms<-functcomp(traits, samp)

    #calculate species to plot multivariate distances
    all<-rbind(traits,cwms)
    all.dist<-as.matrix(vegdist(all, method=dist.method))  #method="bray"
default
    spxplot.dist<-all.dist[row.names(traits),row.names(cwms)]
```

```
    #observed means per group
    spxgroup.dist<-t(apply(spxplot.dist,1,tapply,groups,mean))

    simus <- array(NA,
dim=c(nrow(spxgroup.dist),ncol(spxgroup.dist),nrep+1),
dimnames=list(rownames(spxgroup.dist),colnames(spxgroup.dist), c("obs",
1:nrep)))
    simus[,,1] <- spxgroup.dist

    #simulated means per group
    for (n in 2:nrep){
      tmp<-randomizeMatrix(spxplot.dist, null.model = null.model)
      simus[,,n]<-t(apply(tmp,1,tapply,groups,mean))

    }

    #calculate mean of simulated values, p-values
    rand.means <- apply(X = simus[,,-1], MARGIN = 1:2, FUN = mean, na.rm
= TRUE)
     p.values <- apply(X = simus, MARGIN = 1:2, function(o)
mean(o[names(o)=="obs"] > o[!names(o)=="obs"], na.rm = TRUE))


    #put it together
    out.tmp<-
abind(obs=spxgroup.dist,rand.means=rand.means,p.values=p.values,along=3)
    out<-aperm(out.tmp, c(1,3,2))

    return(out)
}
```

Appendix S3. List of species that are functionally associated to at least one target group of plots ($p < 0.01$) without being diagnostic of that group based on occurrence or abundance data.

Embryo Dunes

Anthemis maritima, *Calystegia soldanella*, *Plantago coronopus*, *Pseudorlaya pumila*.

Mobile dunes

Centaurea sphaerocephala, *Clematis flammula*, *Cyperus kalli*, *Pancratium maritimum*, *Pistacia lentiscus*, *Plantago coronopus*, *Pseudorlaya pumila*, *Pycnocomon rutifolium*, *Quercus ilex*, *Smilax aspera*.

Transition dunes

Asparagus acutifolius, *Chamaesyce peplis*, *Cistus incanus*, *Daphne gnidium*, *Lonicera implexa*, *Lophocloa pubescens*, *Medicago marina*, *Ononis variegata*, *Polygonum maritimum*, *Prasium majus*, *Rubia peregrina*, *Sonchus bulbosus*, *Sporobolus virginicus*.