Linking patterns in phylogeny, traits, abiotic variables and space: a novel approach to linking environmental filtering and plant community assembly

Sandrine Pavoine[1,2*], **Errol Vela**[3], **Sophie Gachet**[4], **Gérard de Bélair**[5] **and Michael B. Bonsall**[1,6]

[1]*Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK;* [2]*Muséum national d'Histoire naturelle, Département Ecologie et Gestion de la Biodiversité, UMR 7204 MNHN-CNRS-UPMC, 61 rue Buffon, 75005 Paris, France;* [3]*Université Montpellier 2, UMR AMAP, TA A51/PS2, 34398 Montpellier Cedex 5, France;* [4]*Institut Méditerranéen d'Ecologie et de Paléoécologie, UMR CNRS-IRD 6116, Université Paul Cézanne, 13397 Marseille Cedex 20, France;* [5]*B.P. 533, 23000 Annaba, Algeria; and* [6]*St. Peter's College, Oxford OX1 2DL, UK*

Summary

1. We introduce a novel method that analyses environmental filtering of plant species in a geographic and phylogenetic context. By connecting species traits with phylogeny, traits with environment, and environment with geography, this comprehensive approach partitions the ecological and evolutionary processes that influence community assembly.

2. Our analysis extends **RLQ** ordination, which connects site attributes in matrix **R** (here environmental variables and spatial positions) with species attributes in matrix **Q** (here biological traits and phylogenetic positions), through the composition of sites in terms of species presences or abundances (matrix **L**). This methodology, which explores and identifies environmental filters that organize communities, was developed to answer four questions: which combinations of trait states are filtered by the environment, which lineages are affected by these filters, which environmental variables contribute to the assemblage of local communities and where do these filters act?

3. At La Mafragh in north-eastern Algeria, our approach shows that plant species traits were distributed according to environmental filters associated with a salinity gradient. Traits associated with the salinity gradient were convergent among Juncaceae, Cyperaceae and Amaranthaceae. The observed phylogenetic and trait patterns were related to how species survived the xeric season. Juncaceae and Cyperaceae, being perennials and anemogamous, tolerate the xeric hot season by restricting their range to the humid centre of the study area (where conditions are close to a subtropical climate). Several Amaranthaceae species co-occur with the Juncaceae and Cyperaceae in two areas with the highest salinity. Most dicots were observed at higher elevations (up to 7 m a.s.l.), had hairy structures that can retain water and reflect solar radiation and were mostly annual or biennial, completing their life cycle before the onset of the xeric season.

4. *Synthesis.* Our methodology describes environmental filters in terms of identified combinations of traits and environmental factors. It allows spatial and phylogenetic signals to be determined by identifying convergent and conserved patterns in the evolution of traits and spatial scales that structured the environment. Our statistical framework is generic and can be readily extended to a wide range of exciting issues, such as host-parasite, plant-pollinator and predator–prey interactions.

Key-words: coastal marsh plain, community ecology, determinants of plant community diversity and structure, functional traits, phylogenetic clustering, phylogenetic signal, **RLQ** analysis, spatial analysis, species assemblage, statistical graphics

---

Introduction

One of the oldest questions raised in ecology is how species assemblages are formed and maintained at local through to global scales. To explain these patterns in species compositions alternative models of community assembly have been proposed (Chase *et al.* 2005). These include neutral models, where species within a trophic level are identical in their competitive ability, movement and fitness (Hubbell 2006). On the contrary, niche-based models, where deterministic mechanisms apply, have identified two main processes that affect the composition of communities: environmental filtering and limiting similarity. Limiting similarity assumes that biotic forces (e.g. competition, mutualism, facilitation) tend to keep coexisting species from being too similar. In contrast, environmental filtering assumes that abiotic forces act to constrain certain traits within limits.

Here we develop a new statistical framework that aims to analyse environmental filtering in an explicit geographic and phylogenetic context. There has been growing interest in how information about phylogenetic relationships between co-occurring species aids our understanding of community assembly (Webb *et al.* 2002; Pavoine, Love & Bonsall 2009; Pavoine, Baguette & Bonsall 2010). These recent analyses build on biogeographic approaches, the merging of evolutionary and ecological approaches (Brooks 1985; Brooks & McLennan 1993; Losos 1996) and the long history of using taxonomy (especially through the use of genus:species ratios) to understand how communities assemble (e.g. Elton 1946; Simberloff 1970). However, many approaches have conflated phylogenetic information with trait values (particularly where trait information is unavailable), relying on the underlying hypothesis that closely related species are more likely to have similar traits than distantly related species. Studies that have combined the analyses of traits with phylogenies, in a context of community assembly, have revealed that convergence in trait states can occur among unrelated species (Cavender-Bares *et al.* 2004; Silvertown *et al.* 2006). For instance, Cavender-Bares *et al.* (2004) found that high phylogenetic diversity within local oak tree communities was explained by convergence in the traits. This suggests that these tree communities are structured through habitat preferences (including preference for pH and soil moisture). In fact ecological traits have been found to display various phylogenetic signals from convergence to conservatism (Losos 2008).

In parallel with the effects of evolutionary factors, communities are organized in space (Fortin & Dale 2005). Environmental variables are expected to have an effect on the spatial distribution of species while functional traits are expected to be indirectly correlated with space through their relation with the environment (Fortin & Dale 2005). If functional traits are phylogenetically conserved, then the phylogenetic composition of communities is expected to be influenced by these (indirect) spatial effects.

Obviously, analyses on environmental filtering, phylogenetic signal and spatial signal are not novel. However, the combined analysis of these three processes is rare. The interests of combining traits with phylogeny and environmental variables with space are fourfold. First, the association between traits and phylogeny allows us to determine whether the traits involved in environmental filters were evolutionarily conserved or convergent, and thus whether the turnover of species trait states across habitats explains the distribution of specific lineages in space. Second, at coarse spatial scales, the association of phylogeny with space, combined with an absent (or distinct) association between traits and environment, might indicate that historical processes (such as the history of the colonization) predominate, or interact with environmental factors, in explaining the composition of communities. Third, the absence of clear organization of traits despite clear organization of species phylogeny over the study area might indicate that some important traits have been omitted (or not measured) in the analysis. Finally, if traits and/or phylogeny are correlated with spatial variables but not environmental variables, then some key environmental processes might also have been omitted from the analysis.

Very few methods have been developed so far that have taken an integrative approach to the analysis of traits, phylogeny, environment and space. Several methods have been developed to evaluate the relative effects of the phylogeny and the environment (Desdevises *et al.* 2003; Diniz-Filho & Bini 2008; Jetz, Sekercioglu & Böhning-Gaese 2008) or the space (Freckleton & Jetz 2009) on the diversity of species traits. However, these analyses do not aim to evaluate the contributions of specific lineages, environmental variables or spatial areas that are specifically involved in trait diversity. To address this question, Mayfield, Boni & Ackerly (2009) explored the correlations between specified traits, specific clades and specific habitat types. This approach is restricted to nominal traits, discrete habitats and clades (non-overlapping groups of related species). A drawback of this approach is that, although a large set of traits might be included, each trait was analysed separately and this results in a large number of statistical tests being performed (with the obvious statistical complications). As far as we are aware, only two methods combine the four aspects (traits, phylogeny, environment and space). The first aims to find spatial patterns in the components of trait diversity attributable to phylogenetic effects and/or environmental effects (Diniz-Filho *et al.* 2007). The second approach removes any phylogenetic information in traits and any spatial signal in environmental variables, to associate 'phylogenetic-free' traits with 'space-free' environment (Kühn, Nobis & Durka 2009).

To obtain an overall description of environmental filters and traits in an explicitly phylogenetic and geographic context, we extended the use of a modern ordination method (the **RLQ** approach). Historically, the **RLQ** analysis was developed to study environmental filtering in ecological communities (Dolédec *et al.* 1996) by elucidating combinations of traits that have the highest covariances with combinations of environmental characteristics. For instance, recent applications of this analysis have focused on plant traits and environmental conditions over the last 15 000 calendar years in coastal British Columbia, Canada (Lacourse 2009), plant and animal trait responses to

regular winter fires in chestnut forests of southern Switzerland (Moretti & Legg 2009) and the association between plant and spider traits with environmental variables in highway verges in France (LeViol *et al.* 2008).

Here we propose to extend the **RLQ** approach so that the traits and the phylogeny of the species are correlated with the geographic locations and the environment where they occur. As Mayfield, Boni & Ackerly (2009) highlighted, most studies of environmental filtering have aimed to detect this process across entire communities although it might be more appropriate to focus on species groups within communities with distinct evolutionary history and ecology. To paraphrase Hubbell (2005), the real question is 'how did these niche differences evolve, how are they maintained ecologically, and what niche differences, if any, matter to the assembly of ecological communities?'.

Accordingly, within a geographic area our approach can identify which trait states are associated with which environmental factors and which parts of the phylogenetic tree are involved. It can include any number and/or type of trait and environmental factors (e.g. binary, circular, fuzzy, nominal, ordinal, ratio-scale). It has the added advantage of determining combinations of traits affected by combinations of environmental variables rather than treating each factor separately. The previous approaches cited above used summarized data, for instance by considering the centre of the geographical range of a species instead of all locations of a species (Freckleton & Jetz 2009), or by reducing the environmental factors to the average conditions where species are located. Our approach utilizes the full raw data and, as far as we are aware, combines for the first time the variability associated with species traits and phylogenetic compositions within sites and, the variability associated with the environments and locations where a species occurs.

To illustrate the strength and potential of our approach, we analyse the floristic inventories from La Mafragh (Mekhada), a site in Algeria which has strong abiotic gradients. Our main hypothesis is that plant communities in La Mafragh are structured by environmental filtering. To interpret the spatial and phylogenetic patterns in La Mafragh we add two additional hypotheses: (1) traits have phylogenetic signals (that is to say closely related species are expected to have similar trait states, while distantly related species are likely to have distinct trait states); (2) environmental variables have spatial signals (they are not distributed randomly in space). Now, the novelty of our approach is that we can explore this environmental filtering process and address the following questions:

- Which combinations of trait states are filtered by the environment?
- Which lineages are affected by these filters?
- Which and by how much do environmental variables contribute to the assembly of the local plant community?
- Where do these filters act?

We discuss our findings in light of recent advances in phylogenetic community ecology.

## Materials and methods

MATHEMATICAL METHODS

The **RLQ** approach

Our application of the **RLQ** analysis (Dolédec *et al.* 1996) explores the relationship between environmental and spatial variables (columns of a matrix **R**) and trait and phylogenetic variables (columns of a matrix **Q**). Matrix **R** has sites as rows and environmental and spatial variables as columns. Matrix **Q** has species as rows and trait and phylogenetic variables as columns. These matrices are linked by a third matrix **L** whose rows are the sites, whose columns are the species and whose entries are abundance, or presence/absence. In the **RLQ** approach, each matrix is first analysed through a factorial analysis. Matrix **L** is treated by correspondence analysis (Greenacre 1984). However, it is necessary to be clear on how matrix **R** is defined and treated (by combining information of environment and space), and how the new matrix **Q** is defined and treated (by combining information of the traits and phylogeny).

The method starts with four matrices (Fig. 1): a matrix **E** for the environment in sites, a matrix **S** for the geographical space, a matrix **T** for the traits of the species and a matrix **P** for the phylogeny (note that our approach might be applied to any phylogenetic tree, e.g. molecular trees, time-calibrated trees). The definition of each matrix depends on the kind of data available for the analysis. For instance, if the traits are all numeric then, matrix **T** might be a species × trait matrix with the trait state for a given species as entry. However, if the traits are a mix of distinct statistical types (for instance numeric, nominal, ordinal, binary, circular) then matrix **T** will be a species × species matrix of distances. Then each matrix is analysed by a factorial method. Table 1 provides a guideline of the possible definition for each matrix and its associated factorial analysis.

To maintain the structure associated with each matrix, the Cartesian coordinates of the sites are retained in the factorial analysis associated with the environmental variables ($\mathbf{X}_E$) and geographic space ($\mathbf{X}_S$) analysis, and the coordinates of the species in the factorial analyses of the traits ($\mathbf{X}_T$) and the phylogeny ($\mathbf{X}_P$). In order to ensure that each matrix is comparable at the same scale then all matrices are standardized. Here we use the square root of the first eigenvalue of each analysis as our standardization. The new matrices are thus: $\mathbf{X}_E^* = \mathbf{X}_E/\sqrt{\lambda_{E,1}}$, $\mathbf{X}_S^* = \mathbf{X}_S/\sqrt{\lambda_{S,1}}$, $\mathbf{X}_T^* = \mathbf{X}_T/\sqrt{\lambda_{T,1}}$, $\mathbf{X}_P^* = \mathbf{X}_P/\sqrt{\lambda_{P,1}}$, where $\lambda_{E,1}$, $\lambda_{S,1}$, $\lambda_{T,1}$ and $\lambda_{P,1}$ are the first eigenvalue of the environmental, spatial, trait and phylogenetic factorial analyses, respectively. Other weights might be chosen and alternatives are proposed in Appendix S1 in Supporting Information.

Matrix **R** is then defined as $[\mathbf{X}_E^*|\mathbf{X}_S^*]$, where matrices $\mathbf{X}_E^*$ and $\mathbf{X}_S^*$ are simply juxtaposed. Matrix **Q** is then defined as $[\mathbf{X}_T^*|\mathbf{X}_P^*]$. These two matrices are analysed in the **RLQ** framework with centred PCA (Fig. 2). The methods used to combine matrices **E** and **S**, and **T** and **P**, follow a multiple factorial analysis (Escofier & Pagès 1994) adapted to deal with the different types of statistical variables.
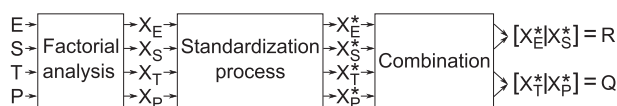
Fig. 1. Linking matrices **E** (environment) and **S** (space) and matrices **T** (traits) and **P** (phylogeny) – summarizing scheme.

Table 1. Examples of factorial analysis appropriate for our analysis of environment, space, traits and phylogeny

Data type	Matrix type	Factorial analysis*
Environmental (**E**) and trait (**T**) matrices		
Numeric	Species × variable	PCA
Nominal and numeric	Species × variable	Hill & Smith (1976) PCA
Mix of unusual types	Species × species†	PCoA
Missing data	Species × variable or Species × species†	NIPALS PCoA
Spatial (**S**) matrix		
Latitude and longitude	Species × variable‡	PCA
Neighbour graph	Species × variable§	PCA
Phylogenetic (**P**) matrix		
Phylogenetic tree	Species × variable¶	PCA
	Species × species**	PCoA

These methods are available, for instance, in the ade4 package of R (Dray & Dufour 2007).
*PCA = principal component analysis; PCoA = principal coordinate analysis (Gower 1966); NIPALS = non-linear iterative partial least squares (Wold, Esbensen & Geladi 1987)
†Distance matrix defined for instance from Gower (1971) or Pavoine *et al.* (2009) (missing data handled)
‡Latitude and longitude might be treated by polynomial transforms (Legendre & Legendre 1998)
§See the Materials and Methods section for details and Appendix S1 for alternatives
¶Variables are defined by orthonormal transforms (Giannini 2003; Ollier, Couteron & Chessel 2006)
**Pairwise distances among species defined as the squared root of the sum of branch length (or number of nodes) on the smallest path that connects two species (see the Materials and Methods section for justifications).

CASE STUDY: PLANT COMMUNITY STRUCTURE IN A COASTAL MARSH PLAIN, LA MAFRAGH (NE ALGERIA)

La Mafragh (36°48′ N–008°00′ E) is used here to refer to a coastal marsh plain (Mekhada) in the east of Annaba in Algeria, bounded by dunes with a narrow connection (Oued Mafragh) with the Mediterranean sea in the north, by Numidian clay-sandstone mountains in the south, by a river (Oued El Kebir) in the east, and by an irrigated agricultural zone in the west (Fig. 3). This region is located in a subhumid bioclimate with warm winters (Emberger 1955, 1966). Plant species in La Mafragh have various origins. The effects of permanent water in the marsh all year long (even in summer) and the warm winters (freezing absent) make this area a rare ecosystem, particularly in the Mediterranean Basin. These rare conditions at La Mafragh allow the unusual coexistence of subtropical and Euro-Siberian plant species. The fact that species with diverse origins co-occur in the plain raises questions on the ecological and evolutionary processes that enable species to co-occur. La Mafragh does not contain any endemic species.

The area is punctuated by the partial effects of anthropogenic developments: drainage, river control, abandoned and active rice fields, extensive exploitation of natural fields as fodder and pasture (cattle breeding), an abandoned raised track. The whole area is furrowed by rivers, and constitutes a basin filled by alluvial and colluvial deposits. The lowest parts are composed of large and small marshes. The low altitude (from 1 to 4 m a.s.l. for the largest part of the area)
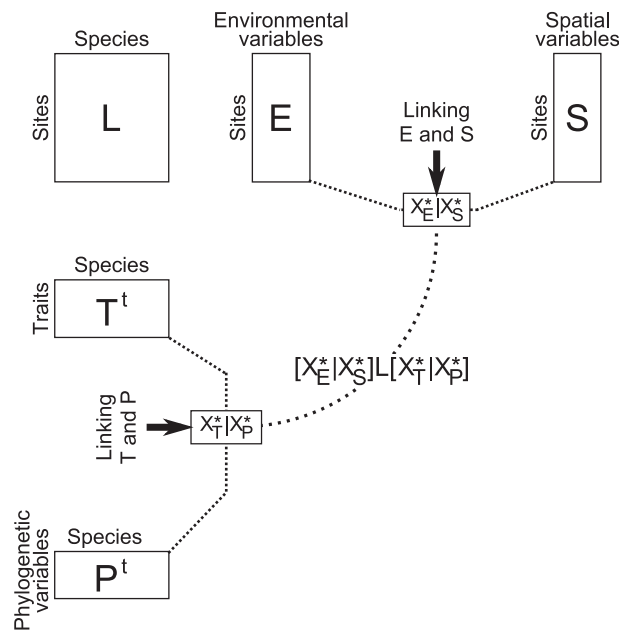
Fig. 2. Schematic summary of our combined analysis of the geographic space (**S**), environmental variables (**E**), species compositions in sampling units (**L**), biological traits (**T**) and phylogeny (**P**). $\mathbf{T}'$ and $\mathbf{P}'$ are the transposed matrices of **T** and **P**, respectively. The notations '$[X_E^*|X_S^*]$' and '$[X_T^*|X_P^*]$' mean that matrices **E** and **S** and matrices **T** and **P**, respectively, are transformed in a way that allows their linking (these matrices are explained in Materials and Methods, in Fig. 1 and the approach is extended in Appendix S1).

and dunes restrict water loss and the presence of an estuary (the Mafragh river) leads to sea water flooding during storms.

This plain is about 15 000 ha, within which is the 10 000 ha study area. Within this area, 102 sites were defined on a regular grid. Five of these sites were excluded from the analysis given their very high heterogeneity, leading to a total of 97 sites. Abundance and environmental data were collected in 1979 and we have now extended this data set by including bibliographical species traits and a phylogeny. Complementary analyses have been regularly performed over the last three decades (unpublished data). These analyses confirmed that no major perturbation has affected the study area since the data we used were collected, such that our conclusions apply to the current local situation.

Environmental variables

The pedological data were collected by de Bélair in 1979 and first analysed in de Bélair (1981) from each site. Ten soil variables were considered (Appendix S2 in Supporting Information): Clay (%), silt (%), sand (%), K_2O (‰), Mg^{2+} (mEq/100 g), Na^+ (mEq/100 g), K^+ (mEq/100 g), conductivity (mMho cm^{-1}), retention capacity (%; pF 2.5), altitude (m). We excluded conductivity and retention from the analysis because of their high correlation with concentration of Na^+ and clay, respectively.

Abundance data

The plant abundance data were collected in 1979 and originally analysed by de Bélair (1981). On each site, three relevés were investigated randomly in a circle with a radius of 100 m around the pedological pit of the site. The relevés were delimited by squares whose edges varied from 2 to 3 m depending on the degree of spatial homogeneity.

Fig. 3. Photographs of the area. (a) La Mafragh in early spring; in the foreground, *Chamaemelum fuscatum*, in the middle distance, young *Bolboschoenus maritimus* and *Schoenoplectus littoralis*; in the background, dried *B. maritimus* and *S. litoralis* from the previous year; then horizon and dunes. (b) La Mafragh in late summer; few green spots of *Cressa cretica* are visible on the xeric, central area.

Fig. 4. Phylogenetic tree. The families are indicated (see Appendix S2 for details).
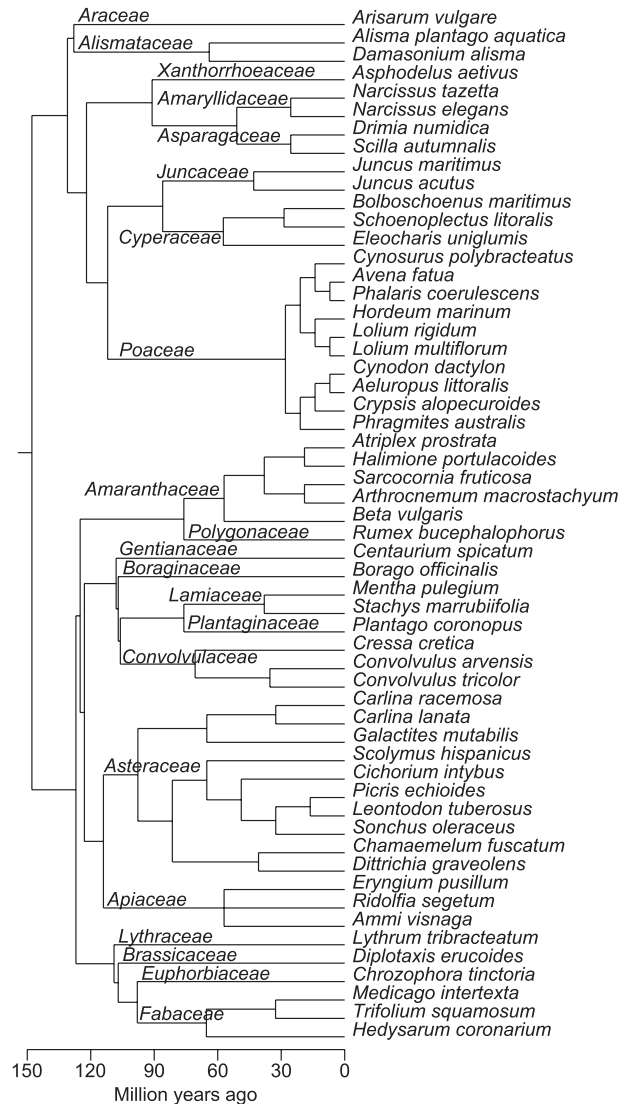
Indices of abundance were attributed to the observed species by phytosociological estimations (Braun-Blanquet, Fuller & Conrad 1932). The Braun Blanquet scale was used in its simplest form (+ , 1, up to 5) and transformed to a scale from 1, 2, up to 6. As the three relevés within sites were very similar, we considered the average index of abundance of the species over the three relevés per site.

Phylogeny

The phylogeny is given in Fig. 4. Its topology was obtained from Phylomatic software (Webb, Ackerly & Kembel 2008) that now includes the new Angiosperm Phylogeny Group classification (APG III Group 2009). Branch lengths were estimated from a set of dated nodes (mostly from Hedges & Kumar 2009) and the *bladj* algorithm (Webb, Ackerly & Kembel 2008). Full details and sources are given in Appendix S2.

Traits

Trait values would ideally be collected directly on individual plants in the field (Cornelissen *et al.* 2003). However, in our work presented

here we chose to analyse previous data that had already been published. As such we do not have contemporaneous measures of plant quantitative traits associated with our unique data set. Although the Mediterranean region is clearly an appropriate and exciting field area to test community assembly rules, the only trait data available are those collected through the literature and that based on expert knowledge. As such trait values are dispersed throughout the literature (floras, scientific papers, unpublished reports) and we could not find a single reference that gathered traits for all species. Principally, we have used the French Mediterranean floristic database BASECO (Gachet, Véla & Tatoni 2005), which compiles trait information from specified Floras. We have supplemented this with information from many different articles or other Floras, compiled specifically for the present work (listed in Table 1). Notwithstanding the analyses undertaken here, this work underlines the urgent need to build a comprehensive Mediterranean plant trait database.

Physiological traits would have been very relevant for applying our methodology to our case study. However, as far as we are aware, this information is not available. Accordingly, we have used, among the available traits, ten traits that were most likely to influence

community assembly in La Mafragh. These included life cycle, flower sexuality, barycentre and length of the flowering period, pollination, minimum and maximum plant height, presence of spiky structures, succulent leaves and hairy leaves (Appendix S2). Using the fourth-corner algorithm (Dray & Legendre 2008), we selected the four traits that were significantly correlated with environmental variables (see Appendix S3 in Supporting Information for details). These traits are listed in Table 2 and Appendix S2.

Statistical analyses

To evaluate the effects of space on the environmental variables (spatial autocorrelation) we performed a Moran's test (Cliff & Ord 1973; Thioulouse, Chessel & Champely 1995). To evaluate if the biological traits had a phylogenetic signal, we performed the root-skewness test developed in Pavoine *et al.* (2010). To assess whether there was any phylogenetic and trait clustering at La Mafragh (lower phylogenetic and trait diversity within local sites than expected from the pool of species) we used the PQE and TQE tests as described in Pavoine *et al.* (2010). These tests are designed to evaluate the degree of environmental filtering (versus limiting similarity) in a metacommunity.

In all factorial analyses, species were weighted by their overall relative abundance over La Mafragh, and sites were weighted by the relative number of plants observed. This weighting scheme is, by definition, derived from the canonical analysis of matrix **L** (that gives species abundances in sites, with sites as rows and species as columns). Accordingly, the species weights are obtained as the sum of values in **L** per column divided by the total sum of values in **L**. Similarly, the site weights are obtained as the sum of values in **L** per row divided by the total sum of values in **L**. As we work with quantitative and proportional environmental variables, the environmental matrix **E** was analysed by centred principal component analysis (PCA) after having scaled the quantitative variables by their range (see Appendix S3). The spatial matrix **S** was defined as the eigenvectors of a *neighbour* matrix (Thioulouse, Chessel & Champely 1995). Overall, a *neighbour* matrix is a site-by-site matrix with binary measures, where a value of 1 is given for two sites that are connected (*neighbours*) and 0 elsewhere. Numerous methods have been developed for defining where two sites are *neighbours* and where they are not (Legendre & Legendre 1998; Dray, Legendre & Peres-Neto 2006). Here we used a Gabriel *neighbour* matrix

adjusted to correct for the connections of vertices in the border of the study area, as defined in the package ade4 (Dray & Dufour 2007; see also Appendix S1 and S3 in Supporting Information). Matrix **S** was analysed by PCA.

As the biological traits were of different statistical types (which cannot be handled by conventional factorial analyses – multichoice and ordinal), we applied the mixed-variables coefficient of distance (Pavoine *et al.* 2009) to compare species trait states, leading to a matrix **T** of pairwise distances between species. This trait distance matrix **T** was analysed by principal coordinate analysis (PCoA). We defined the phylogenetic matrix **P** as a species × species matrix with pairwise phylogenetic distances among species. The phylogenetic distance between two species is evaluated as the square root of the sum of branch lengths along the shortest path that connects species. The square root provides Euclidean distances (Ollier 2004) that were analysed using the PCoA approach (Legendre, Desdevises & Bazin 2002).

We tested the significance of the connection between matrices **R** and **Q** with the multivariate version of the fourth-corner approach and an appropriate null model (Model 4 in Dray & Legendre 2008). This null model assumes that the environmental and spatial variables are fixed, and that the species are randomly distributed in space and environment (whatever their traits and phylogeny). Given the high number of variables included in the analysis, we chose not to test all pairwise connections of environmental and spatial variables with traits and phylogenetic variables, as was suggested in the fourth corner approach. This would have led to a high number of tests, with a drastic chance of obtaining false significant tests. Alternatively, we applied the multivariate version of the fourth-corner approach to matrices **E** and **T** (hypothesis tested: species traits are associated with the environment), **E** and **P** (hypothesis tested: species phylogenies are associated with the environment), **S** and **T** (hypothesis tested: species traits are structured spatially) and **S** and **P** (hypothesis tested: species phylogenies are structured spatially).

By mixing graphical exploratory analysis with formal statistical tests, this approach both confirms our general hypotheses (environmental filtering, phylogenetic signal in traits and spatial signals in the environment) and has the potential to identify new, more precise, hypotheses (Tukey 1977). All analyses were completed in R (R Development Core Team 2010) and the data set (together with the instructions for completing the analyses performed in this paper) is available in Appendices S3–S5.

Table 2. Traits used for the description of plant species

| Traits	Type	Attributes	References
Life cycle	M	4 attributes: Perennial; Annual; Biennial; Seasonal	Following de Bolòs & Vigo (1984–2001) (and additional floras*)
Pollination	M	3 attributes: respective frequency of Autogamous, Entomogamous = pollination by insects and Anemogamous = pollination by wind	Compiled from BASECO (Gachet, Véla & Tatoni 2005), Julve (1998–2008), and additional floras*
Spikiness	O	0 = Absence of spiky structures; 1 = occasional spiky structures; 2 = presence of spiky structures	Various sources compiled* and completed by Errol Vela (present work)
Hairy leaves	O	0 = No; 1 = Sometimes; 2 = Yes	Various sources compiled* and completed by Errol Vela (present work)

Trait type codes: M. Multichoice; O. Ordinal.
*Cuenod, Pottier-Alapetite & Labbe 1954; Pottier-Alapetite 1979–1981; Pignatti 1982; Castroviejo 1984–2010; de Bolòs *et al.* 1993; Jauzein 1995; Valdés *et al.* 2002; Jeanmonod & Gamisans 2007; Jauzein & Tison in press.

Results

All tests were performed with a nominal α error term equal to 5%. The PQE and TQE tests confirmed phylogenetic and trait clustering at La Mafragh (*P*-value = 0.008 with phylogeny and *P*-value < 0.001 with traits). According to Moran's test, all environmental variables had significant spatial autocorrelation (Appendix S3). All traits retained for the **RLQ** analysis had a significant phylogenetic signal although the signal was lowest for life cycle (Table S1 in Appendix S3).

The associations of the environmental and spatial variables with the biological traits and phylogenetic variables were significant: the global fourth-corner test on both space and environment, and on both traits and phylogeny, was strongly significant with the observed value far from the theoretical values (*P* < 0.001). Our more focused tests were all significant (*P* < 0.001 between environment and traits; *P* = 0.004 between environment and phylogeny; *P* = 0.002 between space and traits; and *P* = 0.009 between space and phylogeny). The first axis of the **RLQ**, applied to both space and environment, and both traits and phylogeny, explains 65% of the total variation. The positive side of this axis corresponds to areas that were rich in clay and in salts (Na$^+$/100 g, K$^+$, K$_2$O and Mg^{2+}) (Fig. 5a). These areas were mostly located on
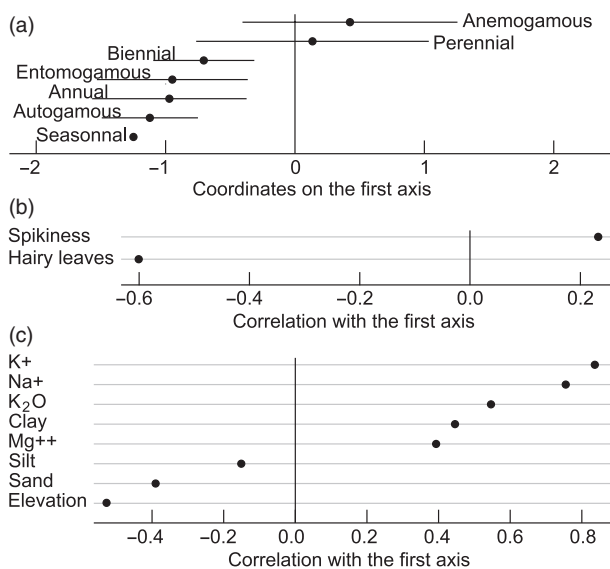
Fig. 5. Detailed effects of the environmental variables and species traits on the first axis of the **RLQ** analysis. (a) The attributes of the multichoice biological traits (life cycle, pollination) are located at the average coordinates of the species that possess them. For a given attribute, the standard deviation of the scores of the species that possess this attribute is given by the length of a segment. Codes are given in Table 1. (b) Spearman correlations (based on ranks) between the ordinal traits and the coordinates of species on the first axis. (c) Pearson correlations (based on raw data) between the numeric environmental variables and the coordinates of the sites on the first axis. From this figure we can deduce that the species located on sites with clay and high concentration of salts are rather anemogamous, perennial and have spiky structures, whereas the species located on sites with sand, low concentration of salts and highest elevation are rather annual or biennial, autogamous or entomogamous and have hairy structures.

the centre of La Mafragh, and especially in two limited zones with the highest concentration of salts (Fig. 6a). The species found in these areas had similar trait states and were mostly perennial, anemogamous, with spiky structures but without hairy leaves (Fig. 5b). They were species in the Amaranthaceae and some species in the Poales (Fig. 6b). In contrast, the negative side of the first axis represents slightly elevated areas (up to 7 m a.s.l.) that are poorer in clay and poor in salts, where plants were more likely to be annual or seasonal, entomogamous or autogamous, with few or no spiky structures, but with hairy leaves. The phylogenetic variables emphasized a distinction between monocots that were, on average, more abundant on central soils with salts and clay and the dicots that were, on average, more abundant on higher elevation with sand and less salts. They also highlighted the dominance of the Juncaceae and Cyperaceae in the centre of La Mafragh (Fig. 6b).

Discussion

We have developed a novel statistical approach that analyses environmental filtering in an explicitly geographic and phylogenetic context, and applied it to the structure of plant communities at La Mafragh. This case study allows us to discuss the relevance of such factorial analysis to studies of community assembly, possible applications at larger spatial scales and extensions to other key issues in ecology.

CASE STUDY: ENVIRONMENTAL FILTERING IN PLANT COMMUNITIES OF LA MAFRAGH

From the plant assemblage study at La Mafragh, we detected significant trait and phylogenetic clustering (lower trait and phylogenetic diversity within each sampled site than expected by chance from the pool of species in La Mafragh). These results suggest that both environmental filtering and phylogenetic signal affect the distribution of traits (Webb *et al*. 2002).

According to our particular **RLQ** analysis, a salinity gradient (involving NA$^+$, Mg^{2+}, K$^+$, K$_2$O) indeed acts as a dominant environmental filter organizing the distribution of plant species in space at La Mafragh. The gradient is organized in three dimensions with two relatively small areas with low elevation and clay soils having particularly high salinity levels (areas A and B on Fig. 6a). The lower elevation and a high proportion of clay soils increase water retention. The high salinity on area A might be due to the low slope and the presence of physical obstacles (with greater elevation) that increase water trapping. Area B is the farthest from the sea (12 km) on a clay-rich soil at low elevation. Its configuration restricts sea water loss after flooding. In these salt-enriched areas, local communities are mostly composed of *Juncus*, Cyperaceae and other monocots, and of several Amaranthaceae species (dicots). Species associated with the highest salinity regimes were often perennial, anemogamous species. The fact that these species are anemogamous is likely related to unfavourable habitats for pollinators due to the high level of disturbance in these areas through regular flooding. These perennial species are maintained through the drier season as they are distributed
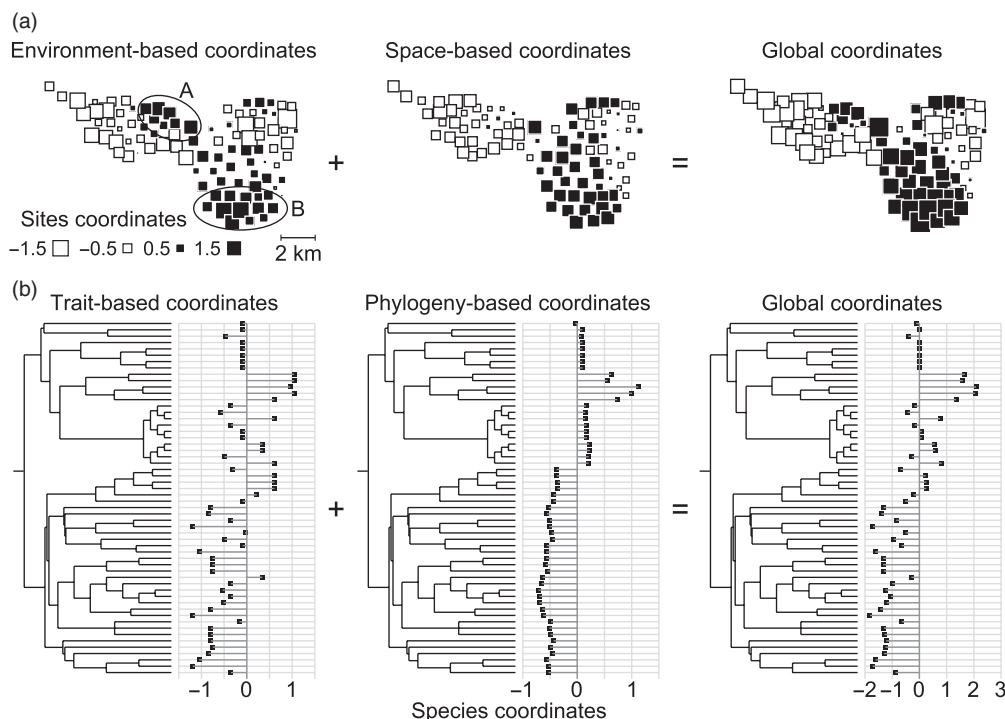
Fig. 6. Result of the **RLQ** analysis visualized on the geographic area and on the phylogeny. The coordinates of sites and species are analysed on the first axis only. (a): The global coordinates of the sites are defined as the sum of a combination of environmental variables and a combination of spatial variables. The sites are positioned at their geographical location in the 16 km × 8 km area of study. Areas A and B with highest salinity are identified. The sizes of the squares are proportional to the absolute values of the site coordinates; white indicates a negative coordinate, and black a positive coordinate. (b): The coordinates of the species are defined as the sum of a combination of trait variables and a combination of phylogenetic variables. The coordinates are given by a Cleveland (1994) dot plot next to the phylogenetic tree (see Fig. 4 for species names). From this figure, we can deduce that monocots and especially Juncaceae and Cyperaceae, in addition to several Amaranthaceae species, are more likely found in the centre of La Mafragh and especially in two areas with similar environment (areas A and B in panel a). The synthetic interpretation of Fig. 5 with Fig. 6 is given in the *Results* section.

on areas likely to retain water. These areas are unfavourable for most other species because, after sea water flooding, evaporation leads to moderate to high salinity concentration. The association of water with areas of high salt concentration thus filters species according to their resistance and determines their distribution. In areas with medium to high salinity, Juncaceae and Cyperaceae monocots with subtropical and/or subcosmopolitan biogeographical origin survive the xeric hot season of the Mediterranean climate by living in the most permanently humid soils within the lowest topographic regions of the study area. These species are tolerant to salts (Pignatti 1982; de Bolòs & Vigo 1984–2001). The Amaranthaceae species that co-occurred with *Juncus* and Cyperaceae at La Mafragh were halophytes and thus resistant to the effects of salt marshes with silt-rich soils, allowing their establishment in the wettest, highly salt-enriched basins (Pignatti 1982; de Bolòs & Vigo 1984–2001).

In contrast, in the areas with lower salinity, annual Mediterranean dicots from numerous families (Apiaceae, Asteraceae, Brassicaceae, Boraginaceae, Convolvulaceae, Euphorbiaceae, Fabaceae, Gentianaceae, Lamiaceae, Lythraceae species) complete their cycle before the onset of the hot summer season (annual and biennial species). These species also had hairy leaves, which facilitate water retention and reflect solar radiation, and were entomogamous and/or autogamous.

Other species, such as a range of Mediterranean monocots (species with coordinates close to zero on the fist axis in Fig. 6b, including Alismataceae, Amaryllidaceae, Araceae, Asparagaceae and Xanthorrhoeaeceae species) are not related with the salinity gradient. These species are dominant in the rainy seasons (autumn to spring) when the clay-rich soils at La Mafragh favour the water retention thus allowing these geophytes to complete both their vegetative and reproductive cycle. In Mediterranean climates, geophytic species are much more prevalent than would be expected and these species are mainly terrestrial and heliophilous (Blondel & Aronson 1999).

Overall our exploratory approach raises three patterns emerging: (i) a combination of humidity and salt filters the species that can be maintained on the central part of La Mafragh; (ii) these species are mostly monocots and especially *Juncus*-like, but are associated with salt-resistant Amaranthaceae; (iii) hairy structures and a short life cycle allow dicots to be maintained in the most xeric areas of La Mafragh.

In addition to environmental filtering through the link between environment and traits, we also found that species were distributed according to geographical space and their phylogenetic relatedness. This key spatial distinction separates the monocots in the middle of the plain from the dicots on the east and west sides (Fig. 6a). A common hypothesis used in community assembly studies over the last decade is that of a

phylogenetic signal (or more 'strongly' phylogenetic conservatism) in traits (Losos 2008). This hypothesis was considered to justify the use of phylogenetic distances among species as predictors of the trait distances among species (e.g. Webb 2000; Gerhold *et al.* 2008). Here we obtained significant signals in all traits. Nevertheless, phylogeny cannot be used as a proxy for traits because, even if the phylogenetic signals are significant, they were not homogeneous in all part of the phylogenetic tree. In particular, we obtained biological similarities between the *Juncus*, Cyperaceae and Amaranthaceae species; these species being perennial, entomogamous and found on soils with moderate to high salinity. Both trait conservatism and trait convergence seem to have shaped trait values of these species that co-occur at La Mafragh.

We have thus demonstrated here that phylogenetic distances between species were not sufficient to describe the distribution of species across the salinity gradient because of evidence of convergence events. Contrary to previous studies (e.g. Cavender-Bares, Keen & Miles 2006), here, in the small area of La Mafragh, this convergence did not lead to a complete absence of phylogenetic signal so that both phylogenetic clustering and trait clustering was found locally. Trait variation thus resulted both from the effect of niche conservatism and unique and independent adaptive responses of each species to environmental conditions (Diniz-Filho *et al.* 2007). This highlights the importance of identifying which trait states and which lineages are filtered by the environment and to describe more precisely how traits evolved among the lineages of the phylogeny.

EXTENSIONS

Obviously, the ecological extensions of this methodology are wide. Our approach is one of the first methods to map and analyse phylogenetic variation in geographical space (Diniz-Filho *et al.* 2007). Using the matrix treatments we proposed in Table 1, the **RLQ** can be applied on the phylogeny (matrix **P**) and space (matrix **S**) alone. Such an approach could be used to identify the spatial scale at which local phylogenetic overdispersion shifts to clustering, which can provide insights on the local and regional mechanisms that affect community assembly (Swenson *et al.* 2006).

Adding traits and environment, we chose to give equal weights to traits and phylogeny and equal weights to environment and space in our **RLQ** framework. This was done by associating the **RLQ** with a multiple factorial analysis. However, there are clearly numerous alternative ways to combine two matrices (i.e. the trait **T** and the phylogenetic matrix **P**, or the environmental **E** and the spatial matrix **S**) in an integrated analysis. These might include redundancy analysis (Rao 1964; Sabatier, Lebreton & Chessel 1989; Legendre & Legendre 1998), co-inertia (Dray, Chessel & Thioulouse 2003) and canonical correlation analysis (Méot, Legendre & Borcard 1998). Accommodating these alternative approaches with the **RLQ** provides a wide range of possibilities and, instead of including phylogenetic and spatial effects, these alternatives allow their removal before associating traits and environment as suggested in Kühn, Nobis & Durka (2009) (Appendix S1).

Furthermore, we applied our analysis of environment, space, traits and phylogeny to understanding community organization at a local scale (10 000 ha). An exciting perspective would be to develop this methodology at a larger scale (for instance, at a continental scale). Indeed, a discrepancy between phylogenetic patterns and trait patterns at large scale might indicate the relative importance of biogeographic, historical factors (e.g. colonization and endemic speciation events) versus local ecological factors that drive assemblage formation (Ingram & Shurin 2009). Our methodology can be adapted to distinguish whether (and when) phylogeny is more related with geographical space and/or whether trait variation is more associated with the local environment. The use of both traits and phylogeny to determine community assembly rules at large scales is still in its infancy (Beche & Statzner 2009; Ingram & Shurin 2009) and there is clearly scope for further theoretical and empirical work here. In addition our methodology goes beyond recent analyses by determining which trait states are associated with which environmental variables, and which lineages of the phylogeny are structured and affected by space.

Obviously, the applications of this extended version of the **RLQ** are not restricted to the analysis of environmental filtering processes. For instance, potential areas for further work include host–parasite, plant–pollinator and predator–prey interactions where phylogenetic and trait relatedness within and between species in resource-consumer interactions can be thoroughly analysed. In a host–parasite interaction for instance, the objective would be to associate the traits and phylogeny of the host with the traits and the phylogeny of parasites. Such an application extends Legendre *et al.* (2002) ideas of associating the phylogenies of hosts and parasites in tests for host–parasite co-evolution.

In conclusion, we provide a novel mathematical framework that develops the modern ordination method (the **RLQ** method) with numerous possibilities of extensions and applications in ecology. In our application to analyse environmental filtering, the inclusion of phylogeny allows genetic and evolutionary components associated with the set of traits to be included in the analysis of community structures. The inclusion of geographic space also allows the role of other environmental factors, such as anthropogenic disturbances and habitat fragmentation, to be appropriately considered as drivers of community organization. The overall novelty of this application is that environmental filtering is described by identifying the trait states and the lineages that are selected by the filters, the value of the abiotic variables that act as filters and the geographic areas or gradients where these filters act. The approach can reveal where in the phylogenetic tree, the trait states are conserved and where they are convergent such that the diversity in traits and the environment can be thoroughly understood.

Acknowledgements

References

APG III Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society*, **161**, 105–121.

Beche, L.A. & Statzner, B. (2009) Richness gradients of stream invertebrates across the USA: taxonomy- and trait-based approaches. *Biodiversity and Conservation*, **18**, 3909–3930.

de Bélair, G. (1981) *Biogéographie et aménagement: la plaine de La Mafragh (Annaba, Algérie)*. PhD Thesis, Université Paul Valéry, Montpellier.

Blondel, J. & Aronson, J. (1999) *Biology and Wildlife of the Mediterranean Regions*. Oxford University Press, Oxford.

de Bolòs, O. & Vigo, J. (1984–2001) *Flora dels països catalans*. Barcino, Barcelona.

de Bolòs, O., Vigo, J., Masalles, R. & Ninot, J.M. (1993) *Flora manual dels països catalans*, 2nd edn. Editorial Portic, Barcelona.

Braun-Blanquet, J., Fuller, G.D. & Conrad, H.S. (1932) *Plant Sociology—The Study of Plant Communities (Authorized English Translation of Planzensoziologie)*. McGraw-Hill Book Company, Inc., New York and London.

Brooks, D.R. (1985) Historical ecology: a new approach to studying the evolution of ecological associations. *Annals of the Missouri Botanical Garden*, **72**, 660–680.

Brooks, D.R. & McLennan, D.A. (1993) Historical ecology: examining phylogenetic components of community evolution. *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds R.E. Ricklefs & D. Schluter), pp. 267–280. University of Chicago Press, Chicago.

Castroviejo, S. (1984–2010) *Flora Iberica: plantes vasculares de la Peninsula iberica e Islas Baleares*. Real Jardín Botánico, CSIC, Madrid, Spain.

Cavender-Bares, J., Keen, A. & Miles, B. (2006) Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology*, **87**, S109–S122.

Cavender-Bares, J., Ackerly, D., Baum, D. & Bazzaz, F. (2004) Phylogenetic overdispersion in Floridian oak communities. *American Naturalist*, **163**, 823–843.

Chase, J.M., Amarasekare, P., Cottenie, K., Gonzalez, A., Holt, R.D., Holyoak, M., Hoopes, M.F., Leibold, M.A., Loreau, M., Mouquet, N., Shurin, J.B. & Tilman, D. (2005) Competing theories for competitive metacommunities. *Metacommunities: Spatial Dynamics and Ecological Communities* (eds M. Holyoak, M.A. Leibold & R.D. Holt), pp. 335–354. The University of Chicago Press, Chicago.

Cleveland, W.S. (1994) *The Elements of Graphing Data*. AT&T Bell Laboratories, Murray Hill, New Jersey.

Cliff, A.D. & Ord, J.K. (1973) *Spatial Autocorrelation*. Pion, London.

Cornelissen, J.H.C., Lavorel, S., Garnier, E., Diaz, S., Buchmann, N., Gurvich, D.E., Reich, P.B., 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.

Cuenod, A., Pottier-Alapetite, G. & Labbe, A. (1954) *Flore analytique et synoptique de la la Tunisie—cryptogames vasculaires, gymnospermes et monocotyledones*. S.E.F.A.N., Tunis, Tunisia.

Desdevises, Y., Legendre, P., Azouzi, L. & Morand, S. (2003) Quantifying phylogenetically structured environmental variation. *Evolution*, **57**, 2647–2652.

Diniz-Filho, J.A.F. & Bini, L.M. (2008) Macroecology, global change and the shadow of forgotten ancestors. *Global Ecology and Biogeography*, **17**, 11–17.

Diniz-Filho, J.A.F., Bini, L.M., Rodriguez, M.A., Rangel, T.F.L.V.B. & Hawkins, B.A. (2007) Seeing the forest for the trees: partitioning ecological and phylogenetic components of Bergmann's rule in European Carnivora. *Ecography*, **30**, 598–608.

Dolédec, S., Chessel, D., Ter Braak, C.J.F. & Champely, S. (1996) Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics*, **3**, 143–166.

Dray, S., Chessel, D. & Thioulouse, J. (2003) Co-inertia analysis and the linking of ecological data tables. *Ecology*, **84**, 3078–3089.

Dray, S. & Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Dray, S. & Legendre, P. (2008) Testing the species traits-environment relationships: the fourth-corner problem revisited. *Ecology*, **89**, 3400–3412.

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483–493.

Elton, J. (1946) Competition and the structure of ecological communities. *Animal Ecology*, **15**, 54–68.

Emberger, L. (1955) Une classification biogéographique des climats. *Recueil des travaux des laboratoires de botanique, de géologie et zoologie de la faculté des sciences de Montpellier*, **7**, 3–43.

Emberger, L. (1966) Place de la région méditerranéenne française dans l'ensemble méditerranéen. *Bulletins de l'APGB*, **2**, 153–165.

Escofier, B. & Pagès, J. (1994) Multiple factorial analysis (AFMULT package). *Computational Statistics & Data Analysis*, **19**, 121–140.

Fortin, M.-J. & Dale, M. (2005) *Spatial Analysis: a Guide for Ecologists*. Cambridge University Press, Cambridge.

Freckleton, R.P. & Jetz, W. (2009) Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **276**, 21–30.

Gachet, S., Véla, E. & Tatoni, T. (2005) BASECO: a floristic and ecological database of Mediterranean French flora. *Biodiversity and Conservation*, **14**, 1023–1034.

Gerhold, P., Pärtel, M., Liira, J., Zobel, K. & Prinzing, A. (2008) Phylogenetic structure of local communities predicts the size of the regional species pool. *Journal of Ecology*, **96**, 709–712.

Giannini, N.P. (2003) Canonical phylogenetic ordination. *Systematic Biology*, **52**, 684–695.

Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 857–874.

Greenacre, M.J. (1984) *Theory and Applications of Correspondence Analysis*. Academic Press, London.

Hedges, S.B. & Kumar, S., eds. (2009) *The TimeTree of Life*, pp. 551. Oxford University Press, New York.

Hill, M.O. & Smith, A.J.E. (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon*, **25**, 249–255.

Hubbell, S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, **19**, 166–172.

Hubbell, S.P. (2006) Neutral theory and the evolution of ecological equivalence. *Ecology*, **87**, 1387–1398.

Ingram, T. & Shurin, J.B. (2009) Trait-based assembly and phylogenetic structure in northeast Pacific rockfish assemblages. *Ecology*, **90**, 2444–2453.

Jauzein, P. (1995) *Flore des champs cultivés*. I.N.R.A. Eds, Sopra, Paris.

Jauzein, P. & Tison, J.M. (In Press) *Flore de France méditerranéenne continentale*. Conservatoire Botanique National Méditerranéen de Porquerolles, Hyères.

Jeanmonod, D. & Gamisans, J. (2007) *Flora Corsica*. Edisud, Aix-en-Provence.

Jetz, W., Sekercioglu, C.H. & Böhning-Gaese, K. (2008) The worldwide variation in avian clutch size across species and space. *PloS Biology*, **6**, e303.

Julve, P. (1998–2008) Baseflor: index botanique, écologique et chorologique de la flore de France (Version 8 décembre 2008).

Kühn, I., Nobis, M.P. & Durka, W. (2009) Combining spatial and phylogenetic eigenvector filtering in trait analysis. *Global Ecology & Biogeography*, **481**, 745–758.

Lacourse, T. (2009) Environmental change controls postglacial forest dynamics through interspecific differences in life-history traits. *Ecology*, **90**, 2149–2160.

Legendre, L., Desdevises, Y. & Bazin, E. (2002) A statistical test for host-parasite coevolution. *Systematic Biology*, **51**, 217–234.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*, 2nd English edn. Elsevier Science BV, Amsterdam.

LeViol, I., Julliard, R., Kerbiriou, C., Redon, L.d., Carnino, N., Machon, N. & Porcher, E. (2008) Plant and spider communities benefit differently from the presence of planted hedgerows in highway verges. *Biological conservation*, **141**, 1581–1590.

Losos, J.B. (1996) Phylogenetic perspectives on community ecology. *Ecology*, **77**, 1344–1354.

Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1007.

Mayfield, M.M., Boni, M.F. & Ackerly, D.D. (2009) Traits, habitats, and clades: identifying traits of potential importance to environmental filtering. *American Naturalist*, **174**, E1–E22.

Méot, A., Legendre, P. & Borcard, D. (1998) Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. *Environmental and Ecological Statistics*, **5**, 1–27.

Moretti, M. & Legg, C. (2009) Combining plant and animal traits to assess community functional responses to disturbance. *Ecography*, **32**, 299–309.

Ollier, S. (2004) *Tools for the integration of spatial, temporal and evolutionary constraints in ecological data analysis - Des outils pour l'intégration des contraintes spatiales, temporelles et évolutives en analyse des données écologiques*. PhD thesis, University of Lyon, Lyon, France.

Ollier, S., Couteron, P. & Chessel, D. (2006) Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. *Biometrics*, **62**, 417–477.

Pavoine, S., Baguette, M. & Bonsall, M.B. (2010) Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecological Monographs*, **80**, 485–507.

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

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.

Pignatti, S. (1982) *Flora d'Italia*. Edagricole, Bologna.

Pottier-Alapetite, G. (1979–1981) *Flore de la Tunisie – dicotyledones*. Imprimerie Officielle de la République Tunisienne, Tunis, Tunisia.

R Development Core Team (2010) *R: a Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhya, A*, **26**, 329–359.

Sabatier, R., Lebreton, J.D. & Chessel, D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. *Multiway Data Analysis* (eds R. Coppi & S. Bolasco), pp. 341–352. Elsevier Science Publishers B.V., North-Holland.

Silvertown, J., McConway, K., Gowing, D., Dodd, M., Fay, M.F., Joseph, J.A. & Dolphin, K. (2006) Absence of phylogenetic signal in the niche structure of meadow plant communities. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **273**, 39–44.

Simberloff, D. (1970) Taxonomic diversity of island biotas. *Evolution*, **24**, 23–47.

Swenson, N.G., Enquist, B.J., Pither, J., Thompson, J. & Zimmerman, J.K. (2006) The problem and promise of scale dependency in community phylogenetics. *Ecology*, **87**, 2418–2424.

Thioulouse, J., Chessel, D. & Champely, S. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics*, **2**, 1–14.

Tukey, J. (1977) *Exploratory Data Analysis*. Addison-Wesley Publishing company, Reading, MA.

Valdés, B., Rejdali, M., Achhal, A., Jury, S.L. & Montserrat, J.M. (2002) *Catalogue des plantes vasculaires du nord du Maroc, incluant des clés d'identification*. CSIC, Madrid.

Webb, C.O. (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist*, **156**, 145–155.

Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, **18**, 2098–2100.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Wold, S., Esbensen, K. & Geladi, P. (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37–52.

Supporting information

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

Appendix S1. Mathematical complements.

Appendix S2. Supplementary information on La Mafragh Case study.

Appendix S3. Instructions for completing the analyses performed in this paper in R.

Appendix S4. Data set. The 'rda' format can be read with R (see Appendix S3).

Appendix S5. R scripts. (see Appendix S3).

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