cati: an R package using functional traits to detect and quantify multi-level community assembly processes

Adrien Taudiere and Cyrille Violle

A. Taudiere (adrien.taudiere@cefe.cnrs.fr) and C. Violle, CEFE UMR 5175, CNRS – Univ. de Montpellier – Univ. Paul-Valéry Montpellier – EPHE – 1919 route de Mende, FR-34293 Montpellier, Cedex 5, France.

Community ecologists are active in describing species by their functional traits, quantifying the functional structure of plant and animal assemblages and inferring community assembly processes with null-model analyses of trait distribution and functional diversity indices. Intraspecific variation in traits and effects of spatial scale are potentially important in these analyses.

Here, we introduce the R package cati (Community Assembly by Traits: Individuals and beyond) available on CRAN, for the analysis of community assembly with functional traits. cati builds on a recent approach to community assembly that explicitly incorporates individual differences in community assembly analyses and decomposes phenotypic variations across scales and organizational levels, based on three phenotypic variance ratios, termed the T-statistics. More generally, the cati package 1) calculates a variety of single-trait and multi-trait indices from interspecific and intraspecific trait measures; 2) it partitions functional trait variation among spatial and taxonomic levels; 3) it implements a palette of flexible null models for detecting non-random patterns of functional traits. These patterns can be used to draw inferences about hypotheses of community assembly such as environmental filtering and species interactions.

The basic input for cati is a data frame in which columns are traits, rows are species or individuals, and entries are the measured trait values. The cati package can also incorporate a square distance matrix into analyses, which could include phylogenetic or genetic distances among individuals or species. Users select from a variety of functional trait metrics and analyze these relative to a null model that specifies trait distributions in a regional source pool.

Beyond species: trait distribution and individual differences in community ecology

Breaking down phenotypes into functional traits (Violle et al. 2007) has long been the basis of comparative research in ecology (Grime 1979, Weiher et al. 1999, Shipley 2007). More recently, the employment of functional traits has stimulated the study of a new facet of biodiversity: the diversity of traits within a study unit (e.g. a community), namely functional diversity (Tilman 2001, Petchey and Gaston 2002, Weiher 2010, Cadotte et al. 2011, Mason and de Bello 2013). From a community-ecology perspective, a functional characterization of species, provides a way of describing patterns and of inferring hypotheses about the processes leading to local species coexistence (Weiher and Keddy 1995, Weiher et al. 1998, 2011, McGill et al. 2006, Adler et al. 2013, Enquist et al. 2015).

Trait-based community assembly approaches generally fall into two categories (Anderson et al. 2011): 1) comparison of observed trait distributions/metrics with null distributions generated by random draws, and 2) analysis of trends in trait distributions/metrics among communities along environmental gradients. Ecologists usually propose a myriad of functional trait metrics to describe the distributions of traits within communities, namely the functional structure of communities (Mouchet et al. 2010, Schleuter et al. 2010, Aiba et al. 2013, Chalmandrier et al. 2013, Mason et al. 2013, Swenson 2014; reviewed in Supplementary material Appendix 1, Table A1). Community ecologists commonly interpret the functional structure of ecological communities as a signature of past and ongoing community assembly processes (Enquist et al. 2015). Examples include: habitat filtering (or environmental filtering, following the terminology of Kraft et al. (2014)) expected to restrict the trait distribution within a community relative to the trait distribution within the regional pool; and niche differentiation processes expected to increase phenotypic differences among individuals and/or species in a community, compared to a random situation. Null models (frequentist statistical tests that control for simple sampling effects on metrics of community patterns) help to interpret the functional structure of a community. However, a pioneering study (Jung et al. 2010 and Box 1) has shown that a lack of consideration of individual differences (e.g. using species' mean-trait values instead of accounting for intraspecific variation) in the characterization

Among the most challenging criticisms of trait-based approaches to community assembly is the illusion of assigning a specific assembly process to a specific spatial scale (Grime 2006, Cavender-Bares et al. 2009, Mayfield and Levine 2010, Violle et al. 2012). For instance, habitat filtering can be considered an 'outside-the-community' process, since an abiotic factor (e.g. a climatic factor) acts at a regional scale and thus should have a uniform effect on all the elements of a local community. However, due to micro-environmental heterogeneity and to biotic processes that can affect the local abiotic environment, this external vision of habitat filtering is caricatured, if not wrong (Adler et al. 2013, Kraft et al. 2014). Violle et al. (2012) proposed a practical way to define community assembly filters based on the identification of a generic external filter (all assembly processes taking place outside the community) and a generic internal filter (all assembly processes internal to the community). Community ecologists can further interpret these as results of biotic or abiotic processes, depending on their background knowledge of the study system. In cati, the hypotheses of randomization underlying the null models are based on this dichotomy (see also Table 3).

Recently, several authors have challenged the mean-field approach to community assembly – i.e. the use of species' trait means to describe their position along a niche axis and subsequently to test for multiple community assembly processes (Bolnick et al. 2003, 2011, Jung et al. 2010, Laughlin et al. 2012, Zaccarelli et al. 2013). Indeed, some field studies have found a considerable and unexpected amount of intraspecific variation (relative to interspecific variation) including variation in animal prey selection (Estes et al. 2003, Araújo and Gonzaga 2007), life-history traits of freshwater fishes (Blanck and Lamouroux 2007, Villéger et al. 2012), and plant functional traits (Albert et al. 2010, Messier et al. 2010, Paine et al. 2011). This implies that intraspecific variation can be central to detecting assembly processes: 1) within-species genetic variability and/or phenotypic plasticity can play a key role in explaining the actual presence of species within communities and species turnover along gradients (Jung et al. 2010, Leps et al. 2011, Schreiber et al. 2011, Siefert 2012); 2) interactions between organisms are more likely to occur among spatially-close individuals and thus should be better captured when accounting for individual phenotypic variation (Gross et al. 2009). More generally, theoretical models have considered the relative importance of intra- and interspecific phenotypic variation as a key parameter of species coexistence (MacArthur and Levins 1967). Building on these findings, Violle et al. (2012) proposed three phenotypic variance ratios, termed the T-statistics (to echo the F-statistics in population genetics, 'T' referring here to traits), to account for intraspecific variation, relative to interspecific variation, in community assembly studies. The three T-statistics are ratios of variances developed to test for internal and external filtering of a given community (Violle et al. 2012; see also Table 3). The primary goal of cati is to describe individual differences within communities (in particular through the implementation of the T-stats) and to evaluate their potential involvement in community assembly.

of the functional structure of communities and related null models, can lead to misleading interpretations of the processes driving the assembly of communities. This represents a continuing shortcoming in the functional trait-based literature, so Violle et al. (2012) proposed investigation of its consequences for community ecology through decomposition of phenotypic variation across levels of scale and organization, using three phenotypic variance ratios. These are termed the T-statistics (where T refers to traits; see also Box 1; echoing the F-statistics of population genetics).

Raison d'être and scope of cati

Here we present cati (Community Assembly by Traits: Individuals and beyond) an R package developed to meet the key challenges facing community ecology (Box 1) and based on the most recent developments in trait-based ecology.

The purpose of cati emerges from two limitations common to previous R packages and other tools and methodologies. These limitations are: 1) the absence of a specific package dedicated to trait-based analysis for community assembly analysis, and 2) the need for a tool for testing the influence of individual differences and intraspecific variation in the assembly of ecological communities.

1) A useful grouping of R packages is already available for characterizing the functional structure of communities

(e.g. FD (Laliberté and Shipley 2011), entropart (Marcon and Herault 2013), hypervolume (Blonder et al. 2014)). Surprisingly, as yet, there is no R package specifically dedicated to the analysis of community assembly using functional traits (see Supplementary material Appendix 2, Table A2 for a list of R packages widely used in community and functional ecology). R users often build their own scripts, sometimes depositing these on their websites or publishing them in peer-reviewed journals. At other times, packages have been customized that were designed for other tasks, including phylogenetic or taxonomic analysis (e.g. vegan (Oksanen et al. 2013), spacodiR (Eastman et al. 2013), picante (Kembel et al. 2010) and ade4 (Dray and Dufour 2007)). Consequently, trait-based community assembly analyses remain difficult, if there is no prior knowledge of the implementation of null models or of the choice of relevant functional diversity indices sensu lato (Mouchet et al. 2010, Pavoine and Bonsall 2010, Schleuter et al. 2010, de Bello 2012, Aiba et al. 2013, Chalmandrier et al. 2013, Mason et al. 2013). Overall, the lack of a package dedicated to trait-based community ecology may already have led (and so will continue to lead) to confusing interpretations and difficulties in establishing cross-study generality.

2) There is a growing consensus on the importance of accounting for individual differences and intraspecific variation in community ecology (Box 1). However, this concern is recent and so no software exists to implement metrics

based on individual variation (including the T-statistics) and to evaluate the influence of intraspecific variation through null models. However, there are two notable exceptions to this. First, the RInSp package (Zaccarelli et al. 2013) investigates inter-individual specialization in resource use. While primarily built to characterize within- and among-population variations, the package can also be used to quantify community-wide individual differences. However, RInSp does not characterize the functional structure of ecological communities. Next, the spacodiR package (Eastman et al. 2013) was built primarily to analyze phylogenetic information. This implements the ratio of Rao's diversities at different scales, and it also associates the null models proposed by Hardy and Senterre (2007). However, spacodiR does not incorporate the decomposition of functional variance nor calculate only three metrics (cf. Supplementary material Appendix 1, Table A1). Usefully, cati and spacodiR can be used together to calculate these metrics for further comparisons (see below).

Furthermore, cati offers a tool: 1) to characterize the functional structure of ecological communities, using both classical functional-diversity descriptors and novel metrics designed to evaluate the relative importance of intra- and interspecific variation within a study unit; and 2) to implement null models that can account for individual differences. Designed to account for intraspecific variation of single traits or multi-trait spaces (functional space, hereafter), cati is flexible enough to integrate traits at higher organizational levels (e.g. populations or functional groups; Table 1). For the first time, cati provides functions to implement the T-statistics and null models that allows comparison with random expectation. Moreover, cati also provides functions to calculate other community assembly metrics available in cati or already implemented in other packages (e.g. FD, hypervolume and spacodiR; cf. Supplementary material Appendix 1, Table A1). Finally, the delineation of regional pools is a major issue in community ecology (Lessard et al. 2012) – these are essential for accurate detection of non-random assembly processes. cati proposes several alternatives to delineate the regional pool when implementing null models, using more- or less-strict delineations. Overall, cati can be considered a useful, general package for community assembly given its facility for implementing a range of metrics, regional pools and null models.

Describing and quantifying the amount of inter- and intraspecific trait variation and the functional structure of ecological communities

Overview

In a straightforward and flexible way, cati describes, quantifies and analyzes the amount of intra- and interspecific phenotypic variation in a site. With cati, several types of analyses can be envisioned through the comparison of phenotypic variation: 1) analysis of different study units (e.g. organism, population, vegetation stratum, grid cell, river, catchment, coral reef); 2) analysis of different states with time (e.g. functional structure of a community before and after a disturbance (Mouillot et al. 2013); Table 1), and 3) analysis of different organizational levels (e.g. within-individual, -guild, -functional group, -trophic) and taxonomic levels (e.g. genus, family, operational taxonomic units (OTUs), Table 1).

The critical development in cati, is the implementation of the T-statistics (new routines) to describe and analyze phenotypic variation across both scale and organizational levels. More generally, cati represents a unified platform that creates opportunities: 1) to call other community assembly metrics (existing routines) from other packages (e.g. FD, hypervolume and spacodiR: Supplementary material Appendix 1, Table A1), 2) to use a comprehensive framework of null models, and 3) to control the composition of regional pools. The 'SE' argument of the three main cati functions Tstats, ComIndex and ComIndexMulti handle measurement errors in the trait distributions within null models. It is hoped this option will encourage community ecologists to more satisfactorily manage these errors in their analyses.

Input data

Inputs (Fig. 1) required are: an individuals × traits matrix (traits), a vector assigning a site name to each individual (ind.plot), and a vector assigning a species name to each individual (sp). For example, we present here the input data for the famous Darwin's finches dataset detailed in the last section (see also Fig. 1 and 2 for other illustrations of input data).

Table 1. Examples of uses of cati with different sorts of traits and ecological distances. The three main arguments of functions – Tstats, ComIndex and ComIndexMulti – can be used to analyze different features of community assembly rules. Trait data traits can be quantitative or qualitative, including phylogenetic information. ind.plot corresponds to the vector of sites for each 'individual', allowing different definitions of a community. sp corresponds to the taxonomic affiliation of each 'individual', allowing investigation of different taxonomic scales.

| Arguments	Features	Examples (possible mathematical transformations)
traits	quantitative	SLA, size, isotopic content (scaled, log)
	qualitative	dietary regime, mutualistic interactions, genetic diversity, leaf microbial diversity (pcoa of distances using Gower distance)
	evolutionary history	phylogenetic distance in the tree (pcoa of distances)
ind.plot	community described spatially	localities, strata, sea depth
	community described temporally	chronosequence, succession
	community described ecologically	environmental gradient: stress, disturbance, altitude
sp	Phylum	species, genus
	Guild	insectivore vs granivore, liana vs tree
	Individual	organ-level traits
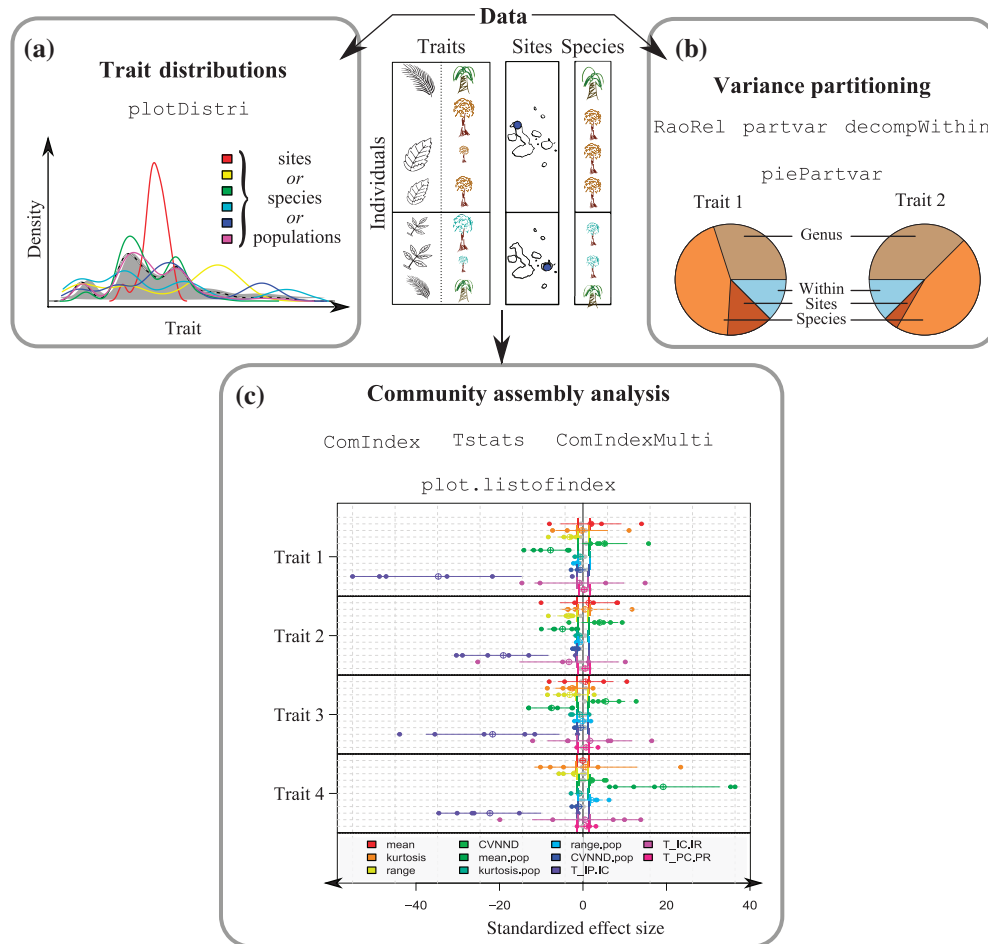
Figure 1. An example of results using cati which allows (a) plots of trait distribution using kernel density, (b) decomposition of trait variances in several ways (including decomposition across ecological scales as represented here on the pie chart) and (c) testing for the departure of observed trait distributions from randomized ones, using multiple metrics and comparing these metrics using standardized effect sizes (SES; Eq. 2). The plot illustrated here compares several metrics for each trait. See Fig. 3 for another example of this plot.

$$\begin{array}{cc} \textbf{BeakH} & \textbf{WingL} \\ 15.2 & 10.6 \\ 14.2 & 9.8 \\ \vdots & \vdots \\ NA & 10.3 \end{array}$$

traits

$$\begin{array}{c} \textbf{Sites} \\ DMaj \\ DMaj \\ \vdots \\ EspHd \end{array}$$

ind.plot

$$\begin{array}{c} \textbf{Species} \\ Camarhynchus\ pallidus \\ Camarhynchus\ parvulus \\ \vdots \\ Platyspiza\ crassirostris \end{array}$$

sp

Individual censuses are often time-consuming and difficult to carry out in many projects. In this case, the vector ind.plot can be replaced by a community matrix species (or population) × sites (argument com) with or without abundances, and traits, corresponds to a species (or populations) × traits matrix.

In the case of qualitative traits, trait values must first be transformed by, e.g. using a principal coordinate analysis (function pcoa in the package ape; Paradis et al. 2004) or a Gower distances analysis (function gowdis in the package FD). cati can also integrate ecological distances (e.g. genetic or phylogenetic). Therefore, most functions of cati can accommodate continuous, integer and factor values if first transformed into a distance or a continuous vector (Table 1).

All but one (Fred) of the functions of the package are able to deal with missing values. cati incorporates basic error trapping in all complex functions and a progress bar for lengthy

calculations. cati depends on the packages ade4, ape and nlme (Pinheiro et al. 2014).

Analysis of the importance of inter- and intraspecific phenotypic variation in ecological communities and implementation of the T-statistics

As a preliminary step, cati offers the opportunity to represent individual differences, and inter- and intraspecific variations within a study unit using the plot function plotDistri. plotDistri can plot the distribution of a given trait (i.e. its kernel density) within a community or at a larger scale (e.g. at the regional pool level). Several visualization tools are available, including distinct curves for all species within a community, an overall curve for a given species across communities and community-wide distribution curves (Fig. 1a).

In cati we implement the calculation of three T-statistics to quantify the relative amounts of intra- to interspecific variation, and of within-community to regional variation (see also Box 1). T-statistics partition phenotypic variances across organizational levels (individual I; population P; community C and region R; Violle et al. 2012; function Tstats; Table 2 and 3). 1) $T_{IP/IC}$ is the ratio of within-population variance
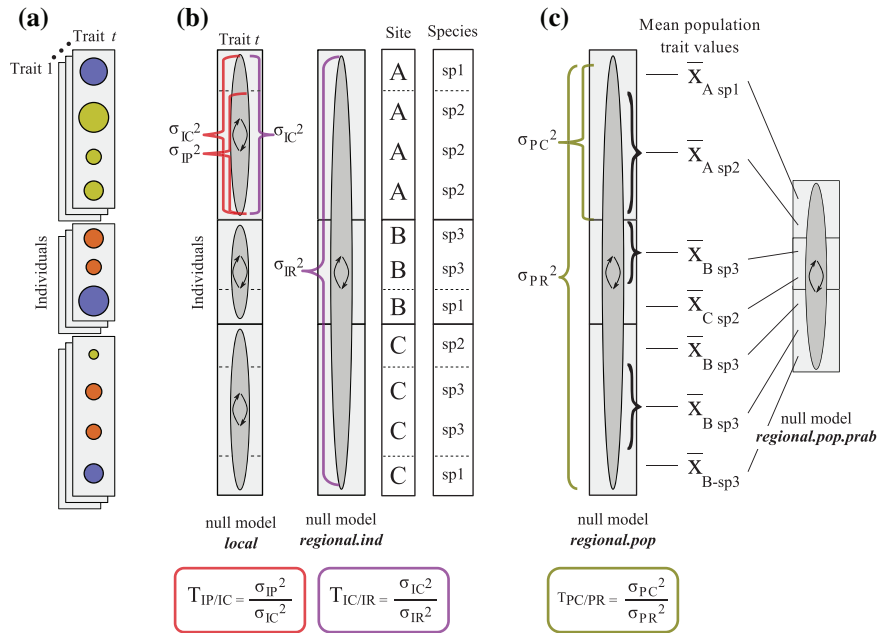
Figure 2. Schematic view of the four null models defined in Table 3. In this example, 11 individuals belonging to three species (sp1, sp2 and sp3) occur on three sites (A, B, C). (a) Distribution of individual values (circles) for the trait t. The three rectangles define the three communities, circle colors define the species, and circle size is proportional to the individual trait value. (b) Null models 'local' and 'regional.ind' use individual trait values. (c) Null models 'regional.pop' and 'regional.pop.prab' are obtained using mean values for each population (e.g. the population of species 1 in site A is labeled A sp1 in the scheme). The three ratios of variance (T-statistics) are also shown in regard to their associated null model. See Table 3 for more details on null models. See Supplementary material Appendix 2–3, Table A2 and Fig. A3 for illustration of other community assembly metrics.

(Individual within Population) to total within-community variance (Individual within Community). It measures the strength of internal filtering, i.e. the strength of niche packing among the species of the community. The higher the overlap of intraspecific trait variation (thus the higher the niche overlap among coexisting species), the higher the value of $T_{IP/IC}$ (see Hulshof et al. 2013 and Le Bagousse-Pinguet et al. 2014 for recent applications in woody and herbaceous communities). 2) $T_{IC/IR}$ is the ratio of community-wide variance (Individual within Community) to total variance in the regional pool (Individual within Region), assessed at the individual level. It measures the strength of external filtering when accounting for individual differences. The higher the overlap of community trait distributions, the higher the value of $T_{IC/IR}$. 3) $T_{PC/PR}$ is the same ratio as $T_{IC/IR}$, but with population-level means only (no intraspecific variation).

Table 2. Main functions in cati and short descriptions.

	Functions	Description	Ref.
Quantify intra-specific variation	RaoRel	The Rao function computes α, β and γ components for taxonomic, functional and/or phylogenetic diversity with: $\gamma = mean (\alpha) + \beta$ where γ is the diversity of the regional pool, α is the diversity of the local community and β is the turnover between local communities. Diversity is estimated using the Rao quadratic entropy indices	1, 2
	partvar	Variance partitioning across nested scales using the decomposition of variance on restricted maximum likelihood (REML) method (lme function)	3
	decompCTRE	This function decomposes the variation in community trait composition into three sources: 1) intraspecific trait variability, 2) variability due to species turnover and 3) their covariation	4
Test for community assembly	Tstats	Computes observed T-statistics (T for Traits) as three ratios of variance, namely $T_{IP/IC}$, $T_{IC/IR}$ and $T_{PC/PR}$. This function can also return the distribution of these three statistics under null models (cf. Table 3)	5
	ComIndex	Computes the moments of the trait distribution (e.g. mean and kurtosis) and other uni-traits metrics (e.g. range and CVNND) to test and quantify the non-random assembly of communities. This function allows researchers to use their own metrics and to choose a null model corresponding to each metric	
	ComIndexMulti	Computes multi-trait metrics (e.g. functional hypervolume) to test and quantify the non-random assembly of communities. This function allows researchers to use their own metrics and to choose a null model corresponding to each metric	

References: 1) de Bello et al. 2011, 2) Rao 1982, 3) Messier et al. 2010, 4) Leps et al. 2011, 5) Violle et al. 2012.

Table 3. The four types of null models implemented in cati, their related null and alternative hypotheses, randomization design and associated T-statistics. All null models can be used with the functions ComIndex and ComIndexMulti.

	Null hypothesis	Randomization procedure	Unilateral alternative hypothesis	T-statistics
local	There is no internal filtering: the distribution of trait values of all individuals within a given community does not depend on species identity	Randomization of individual trait values within the community	Internal filtering significantly affects the distribution of trait values within a given community: two individuals belonging to a population have more-similar trait values than two individuals drawn randomly from the community	$T_{IP/IC}$
regional.ind	There is no external filtering: the distribution of trait values of individuals within a given community, is a random drawing from the regional pool	Drawn without replacement of individual trait values belonging to the regional pool (keeping the actual number of individuals in each community)	Two individuals belonging to a community have more-similar trait values than two individuals drawn randomly from the regional pool	$T_{IC/IR}$
regional.pop and regional.pop.prab	There is no species-based external filtering: the distribution of mean trait values of species within a given community is a random drawing from the regional pool	1) Assigned a population-level value to each individual and 2) drawn without replacement of population-level trait values belonging to the regional pool (keeping the actual number of individuals in each community (regional.pop) or not (regional.pop.prab))	Two individuals belonging to a community have more-similar population-based trait values than two individuals drawn randomly from the regional pool with (regional.pop) or without (regional.pop.prab) taking abundance into account	$T_{PC/PR}$ (regional.pop)

It measures the strength of external filtering at the species level.

We provide three additional and commonly-used methods to decompose diversity and phenotypic variance (Fig. 1b, Table 2): 1) Rao's decomposition into alpha, beta and gamma components of any biodiversity facet (taxonomic, functional or phylogenetic; de Bello et al. 2011; function RaoRel), 2) a gradient analysis based on the decomposition of among-communities trait variation into species turnover, intraspecific trait variability and their covariation (Leps et al. 2011; function decompCTRE), and 3) variance partitioning across nested scales (Messier et al. 2010; function part-var). In summary, we set up simple generic functions and bring together several well-recognized methods in trait-based community ecology to partition phenotypic variation.

A toolbox for the calculation of indices describing the functional structure of communities

Beyond the focus on within-species variation, users of cati can choose to implement a palette of further metrics to describe and quantify trait distributions, accounting for species' abundances or not. The ComIndex function allows calculation of the moments of the distribution of any trait (e.g. mean, variance, kurtosis, skewness; Grant et al. 1985, Kraft et al. 2008, Enquist et al. 2015) and key functional diversity indices such as functional richness FRic, functional evenness FEve and functional divergence FDiv (Villéger et al. 2008; function Fred). A list of indices is required as a vector in the index option of ComIndex. For example, to calculate the variance and the functional divergence FDiv, for all traits, in all the plots studied, and to compare these with local randomizations (see below), the code is:

```
ComIndex(traits,      index = c("var(x)",
"Fred(x, ind.plot)$FDiv"), nullmodels =
"local", ind.plot, sp)
```

In the case of a species-based community matrix (i.e. no individual variation within species), the argument com can be used (species × sites matrix with or without abundance) instead of ind.plot (vector of sites for individuals; Table 1).

Most importantly, it is easily possible to implement any other index in ComIndex by calling other packages (e.g. FD: Laliberté and Shipley 2011; spacodiR: Eastman et al. 2013; Supplementary material Appendix 1–2, Table A1 and A2). For instance, calculating the Tst metric (the proportion of the overall trait diversity expressed among sites) of Hardy and Senterre (2007) involves the following options in ComIndex:

```
index = "spacodi.calc(table(ind.plot, 1:
length(ind.plot)),  sp.traits = as.data.
frame(as.matrix(x)))$Tst"
```

with table(ind.plot, 1:length(ind.plot)), corresponding to the community matrix individuals × sites.

cati is able to calculate indices based on Euclidian space, a minimum spanning tree, a distance matrix, or a hierarchical classification tree (Supplementary material Appendix 1, Table A1). ComIndexMulti is the sister function of ComIndex, but suitable for multi-trait indices. For example, functional dispersion (FDis) and functional evenness (FEve) (existing routines of the FD package) can easily be calculated by this function. As a follow-up of multi-trait and phylogeny-based approaches to biodiversity, a new perspective is to account for the whole phenotype instead of isolated components (Laughlin 2014). Indeed, covariation among traits may occur during assembly processes, so single-trait approaches

can produce contrasting results depending on the trait under examination (Bernard-Verdier et al. 2012). It is thus relevant to investigate the effects of environmental filters on the integrated phenotype that can be assessed by its functional trait space. This is because natural selection and ecological filters most likely act on ecological strategies rather than on single phenotypic traits. The functional space of a population, a community or any other grouping, can be mathematically assessed by an n-dimensional hypervolume (Blonder et al. 2014; see also Lamanna et al. 2014 for an application at local and continental scales). cati takes into account recent developments in this area by using the existing hypervolume routine (R package hypervolume, Blonder et al. 2014) in the function ComIndexMulti. Other hypervolume-like methods can be implemented in ComIndexMulti such as the calculation of the convex hull volume (Cornwell et al. 2006).

ComIndex and ComIndexMulti are generic functions used for community analysis, in this way metrics are also calculated for randomized communities. nperm = NULL returns only observed metrics. This feature speeds the calculation where the requirement is just for descriptors of the functional structure of observed communities.

Community assembly analysis: null models from local to regional scales

Four null models in cati: outlook

To compare observed patterns with random ones, cati implements two sets of null models based on different assumptions – a null model local (internal-to-the community) or one of the three nulls models regional (external-to-the community). These models can also incorporate individual differences (see Table 3 for the hypotheses and randomization procedures associated with each null model).

The null model local randomizes trait values for all individuals within a community, irrespective of taxon identity. This randomization breaks the link between taxonomic identity and trait values within the community (Fig. 2, Table 3). If individual data are not available, but only information about the abundance of each species (e.g. relative abundances in herbaceous or microbial communities), the argument com 'reinterprets' abundance data as individual-like data (internal function AbToInd). In this case, variance decomposition via the T-statistics is impossible because there is no within-species trait variation. The test can be performed with metrics such as the coefficient of variation of nearest neighboring distances (CVNND) (Jung et al. 2010; function CVNND). In the null model local, the departure from the null distribution can be interpreted as an influence of the internal filter. The internal filter tends to force two individuals belonging to a given population (in a community-wide perspective, all individuals belonging to a species in the given community) to display similar trait values compared with two individuals, randomly drawn from the same community (niche packing; Violle et al. 2012).

The null model regional.ind randomizes trait values for all individuals in all communities (or more generally in the regional pool: see below how the regional pool can be delineated in cati) while keeping the actual number of individuals of the communities constant. This randomization breaks the link between taxonomic identity and trait values at the regional scale (Fig. 2, Table 3). The results from regional.ind can be interpreted as the influence of external filtering (any ecological process outside the community that tends to narrow the trait distribution within the community; Violle et al. 2012). The external filter tends to force two individuals belonging to the same community to display more similar traits values than two individuals randomly drawn in the region, irrespective of the species (Violle et al. 2012).

The null model regional.pop mirrors regional.ind but here an average population-level trait value is assigned to each individual in a given population (Fig. 2, Table 3). regional.pop tests the implications of disregarding or averaging-out information about within-population variation, when investigating the impact of external filtering on local community structure. As discussed in Violle et al. (2012), comparing the null models regional.pop and regional.ind allows testing of the importance of accounting for intraspecific variability in community ecology. The null model regional.pop takes species abundances into account, whereas the null model regional.pop.prab does not.

Lastly, several studies use local information about species' composition but generic information about species' traits. For instance it is common to perform community assembly analyses by extracting a single mean trait value from worldwide databases. In plants, the TRY database (Kattge et al. 2011) is proving particularly useful. In this case, all individuals of a given species are assigned the same trait value. Thus, it is not possible to use T-statistics to partition inter- and intraspecific variation but other metrics can be used – for instance the range of trait values displayed by species co-occurring in a community.

Three cati functions designed to run null models

The four null models described above, can be implemented in three cati functions: Tstats, ComIndex and ComIndexMulti.

The Tstats function automatically assigns one specific null model to its related T-statistic (Table 3). Indeed, each T-statistic has been built in relation to a specific scale (see above and Fig. 2). For example, in Tstats, the value of $T_{IP/IC}$ for an observed community will be compared to n randomized communities (option nperm = n) having the same species composition as the observed community, based on null model local.

Users of cati can also apply the four null models with other community assembly metrics using the functions ComIndex and ComIndexMulti – the latter being for multi-trait indices (including distance-based and hypervolume-like metrics). In this case, any metric can be used with any null model (Supplementary material Appendix 3, Fig. A3). The index option selects the list of indices. The nullmodels option selects the null models to consider. Species data, including abundances data, can be included in the analyses (com and type options).

cati allows the delineation of different regional pools

Several authors have stressed the dangers of delineating the regional pool too loosely in community assembly studies

(Lessard et al. 2012). Indeed, most studies combine species (or individuals) from all the communities present in their analyses as a proxy for the regional pool. In cati too, this is the default delineation of the regional pool, though this may lead to underestimation of the regional pool values and to spurious interpretations regarding the community assembly processes at play. Therefore, cati offers several alternatives for delineating the regional pool when implementing regional null models, by allowing attribution of a specific regional pool to each community and also by extending the set of traits values to enlarge the regional pool (argument reg.pool).

Significance tests

In Tstats, ComIndex and ComIndexMulti, the significance of a metric is tested using one-tailed permutation tests. The probability that an observed value departs from the simulated values (i.e. the quantile of the null distribution in which the observed value is found) is:

$$p.value = (x + 1)/(y + 1) \qquad (1)$$

where x is the number of randomized values greater or less than the observed value and y is the total number of null values obtained with the randomization (function Pval). For example, p.value = 0.03, means that 3% of the randomized values are less (or greater) than the observed value. The magnitude of these differences is calculated based on standardized effect size (functions ses and ses.listofindex; Gotelli and McCabe 2002) as:

$$SES = (I_{obs} - I_{sim})/\sigma_{sim} \qquad (2)$$

where I_{obs} is the observed value, I_{sim} the mean of values calculated with the null model and σ_{sim} the standard deviation of these simulated values.

Graphical representations

To facilitate the use of cati's functions and the production of customized graphs, we have developed S3 methods (plot, print and summary functions) linked to classes Tstats, ComIndex, ComIndexMulti, and listofindex. The majority of plot functions represents standardized effect sizes (SES) instead of observed metrics values (e.g. Fig. 1c and Fig. 3). SES values (Eq. 2; functions ses and ses.listofindex) allow a comparison of the magnitude of departure from the null model for different metrics, communities or traits.

Test of bias, power and robustness of the T-statistics

Using simulations, we test bias, power and robustness of the T-statistics in the Supplementary material Appendix 4. We bring out a low type-I error (alpha-error is below 0.05 for all T-statistics; see Supplementary material Appendix 4 for more details). We also studied the type-II error in relation to the type of hypotheses specified in the models. The power to detect the external filter at the individual level ($T_{IC/IR}$) predominates over the power to detect it at the population level ($T_{PC/PR}$) in all simulation cases (Supplementary material Appendix 4). Therefore, users of cati must be careful when
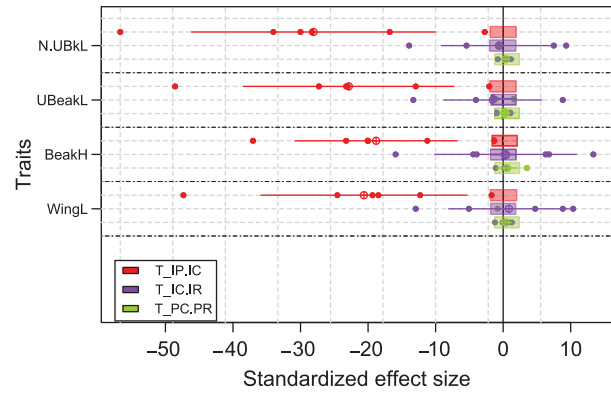

Figure 3. Standardized effect size (SES) of T-statistics for the four traits of the Darwin's finches dataset. This is the unchanged output of the command plot(res) (see main text). Four traits are represented. N.UBkL: nostril upper beak length, UBeakL: upper beak length, BeakH: beak height, and WingL: wing length. For a given trait and a given metric, each colored dot represents the SES value for one community (six communities – here islands – total) when it is different from the null model. T_IC.IR: community-wide variance relative to the total variance in the regional pool, T_IP.IC: ratio of within-population variance to total within-community variance, and T_PC.PR: inter-community variance relative to the total variance in the regional pool. The crossed circles and the segments represent, respectively, the mean and the standard deviation of the SES values for a given T-statistics and a given trait (i.e. mean and standard deviation of community values). For a given T-statistics, the mean of the SES (crossed circle) is significantly different from the null distribution if not embedded within the colored.

comparing $T_{IC/IR}$ and $T_{PC/PR}$. Finally, we detected no bias in the functions partvar and decomCTRE (Supplementary material Appendix 4). This appendix can be rerun with different parameters using the text document available at < https://github.com/adrientaudiere/cati/blob/Package-cati/Documentation/Appendix4/Appendix4.Rnw >.

Application of cati functions using morphological traits of Darwin's finches

To illustrate the capabilities of cati, we provide an example using four morphological traits of Darwin's finches available online (< http://bioquest.org/birdd/morph.php >). In our case study, each island is considered a 'community'. We have deleted the island Cocos from the dataset because of the presence on it of only one species. The remaining dataset contains 13 species, 6 islands and 2513 individual measurements for four traits (N.UBkL: nostril upper beak length, UBeakL: upper beak length, BeakH: beak height and WingL: wing length). Detailed examples of analyses of traits of Darwin's finches by cati are provided in the package's reference manual and tutorial.

First, install and load cati from the Comprehensive R Archive Network (CRAN), and load the Darwin's finches dataset as an example.

```
install.packages("cati")
library(cati)
data(finch.ind)
```

Next, calculate the T-statistics on Darwin's finches data and plot the result.

```
res <- Tstats(traits.finch, ind.plot.
finch, sp.finch)
res
plot(res)
```

The comparison of the departure of the T-statistics from randomized situations (through SES values) is plotted in Fig. 3. The results are consistent among traits. $T_{IP/IC}$ is very low compared with randomized situations for the four traits. This metric reflects the degree of overlap of trait values displayed by co-occurring populations within a community. Then, at the community (island) level, each finch species is suspected to be packed along the niche axis (strong internal filtering) more often than randomly, confirming earlier observations by Darwin (1839) and successors (Grant et al. 1985). In other words, two individuals belonging to a particular population, display more similar trait values than two individuals drawn randomly from the island. This provides further evidence for the importance of biotic interactions in driving the coexistence of finches in each island. Conversely, the four traits displayed no significant pattern for the other T-statistics, either at the individual ($T_{IC/IR}$) or the population ($T_{PC/PR}$) levels. In other words, two individuals living in a particular island do not display significantly more similar trait values than two individuals randomly-drawn from the regional pool. This result suggests there is no external filtering for these traits, possibly due to few environmental differences between the islands.

Concluding remarks and future directions

cati is a package dedicated to the analysis of community assembly using functional traits. It is flexible enough: 1) to implement any uni- or multi-trait metric to describe the dispersion of traits within communities and at larger scales, 2) to partition phenotypic variation at multiple organizational levels (e.g. to account for infra-individual variations such as between-leaf phenotypic variation within a tree), and 3) to run null models with specific assumptions (e.g. by accounting for species abundances, by delineating the regional pool more or less precisely). cati can also implement various distances, including genetic and phylogenetic distances. This allows conjoint treatment of genetic, phylogenetic and phenotypic information at different organizational levels and/or spatial or temporal scales.

Online resources and data accessibility

The cati R package is available from CRAN and development versions, forum; Darwin's finches dataset and a tutorial are available at <https://github.com/adrientaudiere/cati>. The reference manual is available at <http://cran.r-project.org/web/packages/cati/cati.pdf> and the tutorial at <https://github.com/adrientaudiere/cati/blob/Package-cati/Documentation/vignette_Darwin_finches/vignette.pdf>.

References

Adler, P. B. et al. 2013. Trait-based tests of coexistence mechanisms. – Ecol. Lett. 16: 1294–1306.

Aiba, M. et al. 2013. Robustness of trait distribution metrics for community assembly studies under the uncertainties of assembly processes. – Ecology 94: 2873–2885.

Albert, C. H. et al. 2010. A multi–trait approach reveals the structure and the relative importance of intra- vs. interspecific variability in plant traits. – Funct. Ecol. 24: 1192–1201.

Anderson, M. J. et al. 2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. – Ecol. Lett. 14: 19–28.

Araújo, M. S. and Gonzaga, M. O. 2007. Individual specialization in the hunting wasp *Trypoxylon* (Trypargilum) *albonigrum* (Hymenoptera, Crabronidae). – Behav. Ecol. Sociobiol. 61: 1855–1863.

Bernard-Verdier, M. et al. 2012. Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. – J. Ecol. 100: 1422–1433.

Blanck, A. and Lamouroux, N. 2007. Large-scale intraspecific variation in life-history traits of European freshwater fish. – J. Biogeogr. 34: 862–875.

Blonder, B. et al. 2014. The n-dimensional hypervolume. – Global Ecol. Biogeogr. 23: 595–609.

Bolnick, D. I. et al. 2003. The ecology of individuals: incidence and implications of individual specialization. – Am. Nat. 161: 1–28.

Bolnick, D. I. et al. 2011. Why intraspecific trait variation matters in community ecology. – Trends Ecol. Evol. 26: 183–192.

Cadotte, M. W. et al. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. – J. Appl. Ecol. 48: 1079–1087.

Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. – Ecol. Lett. 12: 693–715.

Chalmandrier, L. et al. 2013. A family of null models to distinguish between environmental filtering and biotic interactions in functional diversity patterns. – J. Veg. Sci. 24: 853–864.

Cornwell, W. K. et al. 2006. A trait-based test for habitat filtering: convex hull volume. – Ecology 87: 1465–1471.

Darwin, C. 1839. Journal of researches into the geology and natural history of the various countries visited by H. M. S. Beagle, under the command of Captain FitzRoy, R. N., from 1832 to 1836. – H. Colburn, London, pp. 404–405.

de Bello, F. 2012. The quest for trait convergence and divergence in community assembly: are null-models the magic wand? – Global Ecol. Biogeogr. 21: 312–317.

de Bello, F. et al. 2011. Quantifying the relevance of intraspecific trait variability for functional diversity. – Methods Ecol. Evol. 2: 163–174.

Dray, S. and Dufour, A. B. 2007. The ade4 package: implementing the duality diagram for ecologists. – J. Stat. Softw. 22: 1–20.

Eastman, J. et al. 2013. spacodiR: spatial and phylogenetic analysis of community diversity. – R package ver. 0.13.

Enquist, B. J. et al. 2015. Scaling from traits to ecosystems: developing a general trait driver theory via integrating trait-based and metabolic scaling theories. – Adv. Ecol. Res. 52: 169–199.

Estes, J. A. et al. 2003. Individual variation in prey selection by sea otters: patterns, causes and implications. – J. Anim. Ecol. 72: 144–155.

Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. – Ecology 83: 2091–2096.

Grant, P. R. et al. 1985. Variation in the size and shape of Darwin's finches. – Biol. J. Linn. Soc. 25: 1–39.

Grime, J. P. 1979. Plant strategies and vegetation processes. – Wiley.

Grime, J. P. 2006. Trait convergence and trait divergence in herbaceous plant communities: mechanisms and consequences. – J. Veg. Sci. 17: 255–260.

Gross, N. et al. 2009. Effects of land-use change on productivity depend on small-scale plant species diversity. – Basic Appl. Ecol. 10: 687–696.

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

Hulshof, C. M. et al. 2013. Intra-specific and inter-specific variation in specific leaf area reveal the importance of abiotic and biotic drivers of species diversity across elevation and latitude. – J. Veg. Sci. 24: 921–931.

Jung, V. et al. 2010. Intraspecific variability and trait-based community assembly. – J. Ecol. 98: 1134–1140.

Kattge, J. et al. 2011. TRY – a global database of plant traits. – Global Change Biol. 17: 2905–2935.

Kembel, S. W. et al. 2010. Picante: R tools for integrating phylogenies and ecology. – Bioinformatics 26: 1463–1464.

Kraft, N. et al. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. – Science 322: 580–582.

Kraft, N. J. B. et al. 2014. Community assembly, coexistence, and the environmental filtering metaphor. – Funct. Ecol. doi: 10.1111/1365-2435.12345

Laliberté, E. and Shipley, B. 2011. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. – R package ver. 1.0-11.

Lamanna, C. et al. 2014. Functional trait space and the latitudinal diversity gradient. – Proc. Natl Acad. Sci. USA 111: 13745–13750.

Laughlin, D. C. 2014. The intrinsic dimensionality of plant traits and its relevance to community assembly. – J. Ecol. 102: 186–193.

Laughlin, D. C. et al. 2012. A predictive model of community assembly that incorporates intraspecific trait variation. – Ecol. Lett. 15: 1291–1299.

Le Bagousse-Pinguet, Y. et al. 2014. Species richness of limestone grasslands increases with trait overlap: evidence from within- and between-species functional diversity partitioning. – J. Ecol. 102: 466–474.

Leps, J. et al. 2011. Community trait response to environment: disentangling species turnover vs intraspecific trait variability effects. – Ecography 34: 856–863.

Lessard, J. P. et al. 2012. Inferring local ecological processes amid species pool influences. – Trends Ecol. Evol. 27: 600–607.

MacArthur, R. H. and Levins, R. 1967. The limiting similarity, convergence and divergence of coexisting species. – Am. Nat. 101: 377–385.

Marcon, E. and Herault, B. 2013. Entropart, an R package to partition diversity. – <http://CRAN.R-project.org/package=entropart>.

Mason, N. W. H. and de Bello, F. 2013. Functional diversity: a tool for answering challenging ecological questions. – J. Veg. Sci. 24: 777–780.

Mason, N. W. H. et al. 2013. A guide for using functional diversity indices to reveal changes in assembly processes along ecological gradients. – J. Veg. Sci. 24: 794–806.

Mayfield, M. M. and Levine, J. M. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. – Ecol. Lett. 13: 1085–1093.

McGill, B. J. et al. 2006. Rebuilding community ecology from functional traits. – Trends Ecol. Evol. 21: 178–185.

Messier, J. et al. 2010. How do traits vary across ecological scales? A case for trait-based ecology. – Ecol. Lett. 13: 838–848.

Mouchet, M. et al. 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. – Funct. Ecol. 24: 867–876.

Mouillot, D. et al. 2013. A functional approach reveals community responses to disturbances. – Trends Ecol. Evol. 28: 167–177.

Oksanen, J. et al. 2013. Vegan: community ecology package. – R package ver. 2.0-10.

Paine, C. et al. 2011. Functional traits of individual trees reveal ecological constraints on community assembly in tropical rain forests. – Oikos 120: 720–727.

Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. – Bioinformatics 20: 289–290.

Pavoine, S. and Bonsall, M. B. 2010. Measuring biodiversity to explain community assembly: a unified approach. – Biol. Rev. 86: 792–812.

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

Pinheiro, J. et al. 2014. nlme: linear and nonlinear mixed effects models. – R package ver. 3.1-117.

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

Schleuter, D. et al. 2010. A user's guide to functional diversity indices. – Ecol. Monogr. 80: 469–484.

Schreiber, S. J. et al. 2011. The community effects of phenotypic and genetic variation within a predator population. – Ecology 92: 1582–1593.

Shipley, B. 2007. Comparative plant ecology as a tool for integrating across scales. – Ann. Bot. 99: 965–966.

Siefert, A. 2012. Incorporating intraspecific variation in tests of trait-based community assembly. – Oecologia 170: 767–775.

Swenson, N. G. 2014. Functional and phylogenetic ecology in R. Springer UseR! – Springer.

Tilman, D. 2001. Functional diversity. – In: Levin, S. (ed.), Encyclopaedia of biodiversity. Academic Press, pp. 109–120.

Villéger, S. et al. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. – Ecology 89: 2290–2301.

Villéger, S. et al. 2012. Low functional β-diversity despite high taxonomic β–diversity among tropical estuarine fish communities. – PLoS One 7: e40679.

Violle, C. et al. 2007. Let the concept of trait be functional! – Oikos 116: 882–892.

Violle, C. et al. 2012. The return of the variance: intraspecific variability in community ecology. – Trends Ecol. Evol. 27: 244–252.

Weiher, E. 2010. A primer of trait and functional diversity. – In: Magurran, A. E. and McGill, P. J. (eds), Biological diversity: frontiers in measurement and assessment. Oxford Univ. Press, pp. 175–193.

Weiher, E. and Keddy, P. A. 1995. Assembly rules, null models, and trait dispersion: new questions from old patterns. – Oikos 74: 159–164.

Weiher, E. et al. 1998. Community assembly rules, morphological dispersion, and the coexistence of plant species. – Oikos 81: 309–322.

Weiher, E. et al. 1999. Challenging Theophrastus: a common core list of plant traits for functional ecology. – J. Veg. Sci. 10: 609–620.

Weiher, E. et al. 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. – Phil. Trans. R. Soc. B 366: 2403–2413.

Zaccarelli, N. et al. 2013. RInSp: an R package for the analysis of individual specialization in resource use. – Methods Ecol. Evol. 4: 1018–1023.