

Appendix S1. Definition, equations and references for the indices

Because of the plethora of indices developed so far to measure phylogenetic alpha and beta diversity, we had to select a subset of indices for our analysis. The selection was made to represent the wide variety of aspects of phylogenetic diversity that mathematical indices can capture and the most commonly used indices.

Contents

Notations

Alpha-diversity indices

I. Richness indices

1. Using branch lengths
2. Using phylogenetic pairwise distances
3. Using species phylogenetic isolation index

II. Divergence indices

1. Using branch lengths
2. Using phylogenetic pairwise distances
3. Using nearest distances
4. Using phylogenetic isolation index

III. Regularity indices

1. Using phylogenetic pairwise distances
2. Using nearest distance
3. Using tree topology
4. Using phylogenetic isolation

IV. Parametric indices

1. Unified frameworks
2. Hill numbers adapted to measure phylogenetic diversity

Beta-diversity indices

I. Richness indices (presence-absence data)

II. Divergence indices (using pairwise distances among species)

1. Presence/absence data
 - A. Decomposition into α , β , γ diversities
 - B. Direct dissimilarities
 1. Using all distances
 2. Using nearest distances

- 2. Abundance data
 - A. Decomposition into α, β, γ diversities
 - B. Direct dissimilarities

III. Parametric indices

- 1. Equivalent numbers
- 2. Entropy

Details on the links between several indices

Notations

B is the number of branches. Bt is the set of branches in the phylogenetic tree. Bti is the set of branches in the phylogenetic tree joining species (tip) i to the root of the tree. L_b is the length of branch b . In our paper, we considered T as the root to tip distance (all trees were ultrametric so that for any tree this distance was a constant over all tips). S is the assemblage species number or species richness. R_b is the 'branch' range size, i.e. union of range size of the species descending from branch b . A_b is the branch abundance, i.e. the sum of the abundances of the species descending from branch b . P_b is the branch relative abundance, i.e. the sum of the relative abundances of the species descending from branch b . S_b is the branch richness, i.e. the number of species descending from branch b .

d_{ij} is the phylogenetic distance between two species i and j . Unless otherwise specified we defined the distance between two species as the distance from their first common ancestor. In addition we simulated trees with height (T) standardized to equal 1 so that the maximal possible distance between two species is 1.

$d_{i\min}$ is the distance of a given species i to its closest relative in the assemblage. p_i is the probability to draw an individual of species i from the assemblage or the proportion of species i in the assemblage (measured here as a relative abundance). n_i is the abundance of species i in the assemblage.

ALPHA-diversity indices

Links among indices	Index	Equation	Definition	Reference
<u>I. Richness indices</u>				
1. Using branch lengths				
	PD Phylogenetic diversity	$\sum_{b \in Bt} L_b$	Sum of total branch lengths connecting species together	Faith (1992)

	PD _{Ab} Abundance-weighted PD <i>sensu</i> Vellend	$B * \frac{\sum_{b \in Bt} A_b L_b}{\sum_{b \in Bt} A_b}$	Sum of branch lengths, where branches are scaled by proportional abundances of subtending species. Was named PD _{Aw} in Vellend <i>et al.</i> (2010).	Vellend <i>et al.</i> (2010)
	Δ nPD Abundance-weighted PD <i>sensu</i> Barker	$\sum_{b \in Bt} A_b L_b$	Sum of branch lengths, where branches are scaled by abundances of subtending species.	Barker (2002)
	PE Phylogenetic endemism	$\sum_{b \in Bt} \frac{L_b}{R_b}$	Sum of total branch lengths connecting species together weighted by their range size	Rosauer <i>et al.</i> (2009)
2. Using phylogenetic pairwise distances				
	F Extensive quadratic entropy	$\sum_i \sum_j d_{ij}$	Sum of pairwise distances	Izsák & Papp (2000); Izsák & Szeidl (2002)
	PSR Phylogenetic species richness	$\sum_i \left(\frac{1}{S-1} \sum_j d_{ij} \right)$ with either $d_{ij} = 0.5(c_{ii} + c_{jj} - c_{ij})$ or $d_{ij} = 1 - c_{ij} / \left(\sqrt{c_{ii} c_{jj}} \right)$	Variability in an unmeasured neutral trait multiplied by species richness. c_{ii} is the sum of branch lengths from species (tip) i to the root of the phylogenetic tree; c_{ij} is the sum of branch lengths from first common ancestor for i and j to the root. (See section 'Details on the links between several indices')	Helmus <i>et al.</i> (2007)
3. Using species phylogenetic isolation index				
A phylogenetic isolation index represents the relative isolation of a given species within a phylogenetic tree. Several indices have been proposed so far but we focus here on the evolutionary distinctiveness index called 'Fair Proportion' as proposed by Redding (2003) and Isaac (2007).				
	ED Summed evolutionary distinctiveness	$\sum_i ED_i$ with $ED_i = \sum_{b \in Bti} \frac{L_b}{S_b}$	Sum of species' evolutionary distinctiveness. Note that, in our case study, this index is not equal to Faith's PD because the the ED_i are computed from the regional pool of species and summed across a given assemblage (i.e. a subset of the regional species pool)	This paper. Also Safi <i>et al.</i> (2013). The ED_i were defined by Redding (2003) and Isaac (2007)

	AED Abundance-weighted ED	$\sum_i AED_i$ with $AED_i = \sum_{b \in Bt_i} \frac{L_b}{A_b}$	Abundance-weighted version of ED	This paper. AED_i was defined by Cadotte <i>et al.</i> (2010)
II. Divergence indices				
1. Using branch lengths				
	avPD Average phylogenetic diversity	$\frac{PD}{S}$	Sum of total branch lengths, where branches are scaled by proportional abundances of subtending species, divided by the number of species	Clarke & Warwick (2001)
	avPD _{Ab} Abundance-weighted avPD	$B * \frac{\sum_{b \in Bt} A_b L_b}{\sum_{b \in Bt} A_b}$ $\frac{\quad}{S}$	Sum of total branch lengths connecting species together divided by the number of species	This paper
2. Using phylogenetic pairwise distances				
MPD = AvTD = S/(S-1)*J	MPD (also named AvTD and Δ ⁺) Mean pairwise distances	$\frac{\sum_{ij} d_{ij}}{S(S-1)}$	Mean distances between species	Clarke & Warwick (1998); Webb <i>et al.</i> (2002, 2008); Kembel <i>et al.</i> (2010)
	PSV Phylogenetic species variability	MPD with either $d_{ij} = 0.5(c_{ii} + c_{jj} - c_{ij})$ or $d_{ij} = 1 - c_{ij} / (\sqrt{c_{ii}c_{jj}})$	Variability in an unmeasured neutral trait or the relative amount of unshared branch length. c_{ii} is the sum of branch lengths from species (tip) i to the root of the phylogenetic tree; c_{ij} is the sum of branch lengths from first common ancestor for i and j to the root. (See Section 'Details on the links between several indices')	Helmus <i>et al.</i> (2007)

	J Intensive quadratic entropy	$\frac{\sum_{ij} d_{ij}}{S^2}$	Average distance between two randomly chosen species	Izsák & Papp (2000)
With certain definitions of the phylogenetic distances among species (see the definition of PSE) PSE = S/(S-1) Rao's QE	MPD _{Ab} = Rao's QE Rao's Quadratic Entropy or abundance- weighted MPD	$\sum_i \sum_j d_{ij} p_i p_j$	Quadratic entropy: Simpson's type diversity index where the product of species relative abundances is weighted by phylogenetic distances. This is also the phylogenetic distance between two randomly chosen individuals [drawn WITH replacement]	Rao (1982); Clarke & Warwick (1998); Pavoine <i>et al.</i> (2005); Hardy & Senterre (2007); Webb <i>et al.</i> (2002, 2008); Kembel <i>et al.</i> (2010)
	PSE Phylogenetic species evenness	$\frac{S}{S-1} \sum_{ij} d_{ij} p_i p_j$ with either $d_{ij} = 0.5(c_{ii} + c_{jj} - c_{ij})$ or $d_{ij} = 1 - c_{ij} / (\sqrt{c_{ii} c_{jj}})$	Abundance-weighted PSV. c_{ii} is the sum of branch lengths from species (tip) i to the root of the phylogenetic tree; c_{ij} is the sum of branch lengths from first common ancestor for i and j to the root. (See Section 'Details on the links between several indices')	Helmus <i>et al.</i> (2007)
	InterMPD _{Ab} (also named Δ^*) Interspecific MPD _{Ab}	$\sum_i \sum_{j \neq i} d_{ij} p_i p_j / \sum_i \sum_{j \neq i} p_i p_j$	'Interspecific' MPD _{Ab} : expected phylogenetic distance between two individuals randomly drawn conditionally on the fact that the individuals are from different species	Clarke & Warwick (1998); Miller <i>et al.</i> (2013)
3. Using nearest distances				
	MNTD Mean nearest taxon distance	$\frac{1}{S} \sum_i d_{i \min}$	Mean shortest distance from a species to all others in the assemblage	Webb <i>et al.</i> (2002, 2008); Kembel <i>et al.</i> (2010)
	MNTD _{Ab} Abundance- weighted MNTD	$\sum_{i=1}^S [d_{i \min} * p_i]$	Abundance-weighted MNTD: mean shortest distances, adjusted by species' proportions (i.e. species' relative abundances)	Webb <i>et al.</i> (2002, 2008); Kembel <i>et al.</i> (2010)
4. Using phylogenetic isolation index				
A phylogenetic isolation index represents the relative isolation of a given species within a phylogenetic tree. Several indices have been proposed so far but we focus here only on the evolutionary distinctiveness as measured by the index 'Fair proportion' (Redding, 2003; Isaac, 2007).				

	mean(ED) Mean evolutionary distinctiveness	$\frac{\sum_i ED_i}{S}$ with $ED_i = \sum_{b \in Bti} \frac{L_b}{S_b}$	Mean of species' evolutionary distinctiveness	This paper. The ED_i were defined by Redding (2003) and Isaac (2007)
III. Regularity indices				
1. Using phylogenetic pairwise distances				
	VPD (also named VarTD and Λ^+) Variance in pairwise distances	$\frac{1}{S(S-1)} \left(\sum_i \sum_{j \neq i} (d_{ij} - MPD)^2 \right)$	Variance in pairwise distances	Clarke & Warwick (2001)
	VPD _{Ab} Abundance weighted VPD	$(\sum_i \sum_j n_i n_j) \frac{\sum_i \sum_j n_i n_j (d_{ij} - MPD_{Ab})^2}{(\sum_i \sum_j n_i n_j)^2 - \sum_i \sum_j (n_i n_j)^2}$	Variance in pairwise distance weighted by species abundances.	This paper
	InterVPD _{Ab} Interspecific VPD _{Ab}	$(\sum_i \sum_{j \neq i} n_i n_j) \frac{\sum_i \sum_{j \neq i} n_i n_j (d_{ij} - InterMPD_{Ab})^2}{(\sum_i \sum_{j \neq i} n_i n_j)^2 - \sum_i \sum_{j \neq i} (n_i n_j)^2}$	Variance in pairwise distance weighted by species abundances. n_i is the abundance of species i in the assemblage.	This paper
2. Using nearest distances				
	VNTD Variance in nearest taxon distances	$\frac{1}{S} \sum_{i=1}^S [(d_{i \min} - MNTD)^2]$	Variance in nearest pairwise distance	This paper
	VNTD _{Ab} Abundance-weighted VNTD	$\frac{(\sum_i n_i) \sum_i n_i (d_{i \min} - MNTD_{Ab})^2}{(\sum_i n_i)^2 - \sum_i n_i^2}$	Variance in nearest pairwise distance weighted by species abundances	This paper
	PE _{ve} Phylogenetic evenness	Weighted evenness: $EW_l = \frac{dist(i, j)}{n_i + n_j} \bigg/ \sum_{k=1}^S n_k$	Phylogenetic version of the functional FEve index. First a minimum spanning tree (MST) is computed using the cophenetic distances obtained from the phylogenetic tree. The MST contains $S-1$ branches	Villéger <i>et al.</i> (2008); Dehling <i>et al.</i> (2014)

		<p>Partial weighted evenness:</p> $PEW_l = \frac{EW_l}{\sum_{l=1}^{S-1} EW_l}$ $PE_{ve} = \frac{\sum_{l=1}^{S-1} \min(PEW_l, 1/S - 1) - 1/S - 1}{1 - 1/S - 1}$	<p>connecting the S species. We denote l a branch on the MST, $dist(i,j)$ is the length of the branch l that connects species i and j. n_i is, as defined above, the abundance of species i in the assemblage.</p>	
3. Tree topology				
	<p>IAC Imbalance of abundances at the clade level</p>	$\frac{\sum_{i=1}^S n_i - \hat{n}_i }{v}$ <p>with</p> $\hat{n}_i = \frac{N}{\prod_{k \in s(i, root)} \eta_k}$	<p>IAC quantifies the relative deviation in the abundance distribution from a null case where individuals are evenly partitioned between clade splits. v is the number of nodes in the phylogenetic tree. n_i is, as defined above, the abundance of species i in the assemblage. \hat{n}_i is the expected abundance species i would have if the abundance was randomly split among lineages in the phylogenetic tree at each speciation event. η_k is the number of lineages originating at node k in the set $s(k, root)$, which contains the nodes located on the path between node k and the root of the phylogenetic tree. N is the total assemblage abundance</p>	<p>Cadotte <i>et al.</i> (2010)</p>
	<p>Ic Colless index</p>	$\sum_{k=1}^v S_{1k} - S_{2k} $	<p>Sum of absolute differences in species richness between sister-clades at each internal node. For fully resolved trees, each internal node defines two sister-clades. S_{1k} is the number of species descending from the first clade defined by node k and S_{2k} that of the second clade. v is, as defined above, the number of nodes in the phylogenetic tree</p>	<p>Colless (1982)</p>

	γ Gamma index	$\frac{\left(\frac{1}{S-2} \sum_{i=2}^{S-1} (\sum_{k=2}^i kt_k)\right) - \frac{1}{2} \sum_{j=2}^S jt_j}{\left(\sum_{j=2}^S jt_j\right) \sqrt{\frac{1}{12(S-2)}}}$	<p>The index characterizes the distribution of branching events within the tree. Trees with $\gamma < 0$ have relatively longer branches towards the tips of the phylogeny (tippy trees), whereas trees with $\gamma > 0$ have relatively longer inter-nodal distances towards the root of the phylogeny (stemmy trees). t_k represents an 'evolutionary period' (limits are given by two speciation events) or equivalently an internode distance.</p>	Pybus & Harvey (2000)
--	----------------------	---	--	-----------------------

4. Using phylogenetic isolation

	H_{ED} Entropy measure of evolutionary distinctiveness	$-\sum_{i=1}^S \frac{ED_i}{\sum_{i=1}^S ED_i} \ln \left(\frac{ED_i}{\sum_{i=1}^S ED_i} \right)$	Shannon index applied to evolutionary distinctiveness values	Cadotte <i>et al.</i> (2010)
	E_{ED} Equitability of evolutionary distinctiveness	$H_{ED}/\ln(S)$	H_{ED} controlled for species richness	Cadotte <i>et al.</i> (2010)
	$\text{var}(ED)$ Variance in evolutionary distinctiveness	$\text{var}(ED) = \frac{1}{S-1} \sum_{i=1}^S \left(ED_i - \frac{\sum_{i=1}^S ED_i}{S} \right)^2$	Variance of species' evolutionary distinctiveness	This paper
	H_{AED} Abundance-weighted version of H_{ED}	$-\sum_{i=1}^S \frac{n_i AED_i}{\sum_{i=1}^S n_i AED_i} \ln \left(\frac{n_i AED_i}{\sum_{i=1}^S n_i AED_i} \right)$	Abundance-weighted version of H_{ED}	Cadotte <i>et al.</i> (2010)

IV. Parametric indices

1. Unified frameworks

Entropies: unified by Pavoine et al. (2009): "Tsallis Number" (for ultrametric phylogenetic trees)				
General formula of the framework	I_q Rarity-adjustable index of PD	$\sum_k [t_k - t_{k-1}] H_{q,k}$ <p>with</p> $H_{q,k} = \frac{1 - \sum_i P_i^q}{q - 1}$	t_k represents an 'evolutionary period' (limits are given by two speciation events). q is the factor that tunes the weight given to rare versus abundant species. P_i represents the relative abundance of lineage i descending from this period (sum of the relative abundances of the species descending from this lineage)	Pavoine et al. (2009)
$q = 2$	I_2 =Rao's QE if patristic distances among species are used in Rao's QE (see Section 'Divergence')			
$q \rightarrow 1$	$I_1=H_p$ Phylogenetic entropy	$- \sum_{b \in B_t} L_b P_b \ln(P_b)$	Phylogenetic entropy	Allen et al. (2009)
$q = 0$	$I_0 = PD - T$ (See Section 'Richness')			
2. Hill numbers adapted to measure phylogenetic diversity				
2A. Chao et al. framework				
First version of the framework	${}^q D(T)$	$\left\{ \sum_{b \in B_t} \frac{L_b}{T} P_b^q \right\}^{1/(1-q)}$	Represents the 'mean effective number of species' over any time interval of interest. L_b represents length of branch b while P_b is the sum of the relative abundances of species descending from branch b	Chao et al. (2010)
$q=2$		${}^2 D(T) = 1 / \left\{ \sum_{b \in B_t} \frac{L_b}{T} P_b^2 \right\}$	${}^2 D(T) = 1 / (1 - \text{Rao's QE} / T)$. See Chao et al. (2010) for details	Chao et al. (2010)
$q \rightarrow 1$		e^{H_p}	See above for the definition of H_p	Chao et al. (2010)
$q=0$		$\frac{PD}{T}$	(see 'Richness section')	Chao et al. (2010)
Second version of the framework	${}^q PD(T)$	${}^q PD(T) = T * {}^q D(T)$	'Effective number of lineages'	Chao et al. (2010)

<i>2B. Leinster & Cobbold framework</i>				
<i>General formula of the framework</i>	${}^qD^z(\mathbf{p})$	$\left(\sum_{i=1}^S p_i (Z_p)_i^{q-1} \right)^{\frac{1}{1-q}}$ <p style="text-align: center;">with</p> $(Z_p)_i = \sum_{j=1}^S Z_{ij} p_j$	<p>$(Z_p)_i$ is the expected similarity between an individual of the i^{th} species and an individual chosen at random. The phylogenetic similarity for two species i and j is given by Z_{ij}. Here p_i refers to abundance of species i. \mathbf{p} is the vector of species' relative abundances. Here we used $Z_{ij} = 1 - d_{ij}$</p>	Leinster & Cobbold (2012)
<i>2C. Scheiner framework</i>				
	${}^qD(P)$	$\left\{ \sum_{i=1}^S \left(\frac{ED_i}{\sum_{i=1}^S ED_i} \right)^q \right\}^{1/(1-q)}$	Hill numbers applied to species relative evolutionary distinctiveness. For $q=0$, this index reduces to S	Scheiner (2012)
	${}^qD(AP)$	$\left\{ \sum_{i=1}^S \left(\frac{n_i ED_i}{\sum_{i=1}^S n_i ED_i} \right)^q \right\}^{1/(1-q)}$	Abundance-weighted version of ${}^qD(P)$	Scheiner (2012)

BETA-diversity indices

Class of indices	Index	Equation	Definition	Reference
<u>I. Richness indices (presence-absence data only)</u>				
	Phylosor (Phylosor _{Ab}) Phylogenetic Sørensen index	$\frac{2 * BL_{ij}}{(BL_i + BL_j)}$	Generalization of Sørensen index. BL_{ij} is the branch length common to both communities i and j , and BL_i and BL_j are the total lengths of assemblage i and j , respectively. So phylosor represents the proportional shared branch length between two communities. Branches can be weighted by abundances	Bryant <i>et al.</i> (2008)
	Unifrac (Unifrac _{Ab})	$BL_{ij} / (BL_i + BL_j - BL_{ij})$	Generalization of Jaccard index (same notations as above). Branches can be weighted by abundances.	Lozupone & Knight (2005)
	Faith's PD β	$\frac{BL_{ij} + 0.5 * BL_0}{BL_i + BL_j + BL_0 - BL_{ij}}$	Same notations as above. BL_0 is the branch lengths not present in the two communities i and j but present in the region (pool)	Nipperess <i>et al.</i> (2010)
<i>Pavoine & Ricotta (2014) generalization of species turnover measures</i>	Although the Pavoine & Ricotta (2014) indices can incorporate abundance data, we chose to apply them here to presence-absence data, so that they represent 'richness indices'. In the equations below, x_i & z_j are presences/absences of species in the two compared communities. When formulated with absolute abundances, they are also 'richness indices'; when formulated with relative abundances they are 'divergence indices'. $\sigma_{ij} = 1 - \delta_{ij}$ with δ_{ij} being a measure of pairwise phylogenetic similarity among species. The phylogenetic distances should be between 0 and 1 and satisfy certain mathematical properties (see Pavoine & Ricotta, 2014, for calculations of phylogenetic similarities).			
	S _{Jaccard}	$\frac{\sum_{ij} x_i z_j \sigma_{ij}}{\sum_{ij} x_i x_j \sigma_{ij} + \sum_{ij} z_i z_j \sigma_{ij} + \sum_{ij} x_i z_j \sigma_{ij}}$	Generalization of Jaccard measure of species turnover	Pavoine & Ricotta (2014)
	S _{Ochial}	$\frac{\sum_{ij} x_i z_j \sigma_{ij}}{\sqrt{\sum_{ij} x_i x_j \sigma_{ij}} * \sqrt{\sum_{ij} z_i z_j \sigma_{ij}}}$	Generalization of Ochial measure of species turnover	Pavoine & Ricotta (2014)
	S _{Sorensen}	$\frac{\sum_{ij} x_i z_j \sigma_{ij}}{0.5 * \sum_{ij} x_i x_j \sigma_{ij} + 0.5 * \sum_{ij} z_i z_j \sigma_{ij}}$	Generalization of Sørensen measure of species turnover	Pavoine & Ricotta (2014)
	S _{Sokal-Sneath}	$\frac{\sum_{ij} x_i z_j \sigma_{ij}}{2 * \sum_{ij} x_i x_j \sigma_{ij} + 2 * \sum_{ij} z_i z_j \sigma_{ij} - 3 * \sum_{ij} x_i z_j \sigma_{ij}}$	Generalization of Sokal & Sneath measure of species turnover	Pavoine & Ricotta (2014)

	S_β	$\frac{4 \sum_{ij} x_i z_j \sigma_{ij}}{2 \sum_{ij} x_i x_j \sigma_{ij} + \sum_{ij} z_i z_j \sigma_{ij} + \sum_{ij} x_i z_j \sigma_{ij}}$	Standardized version of Rao's D_{Ab} .	Pavoine & Ricotta (2014)
II. Divergence indices (using pairwise distances among species)				
1. Presence/absence data				
<i>1A. Decomposition into α, β, γ diversities</i>				
	Π_{ST} MPD-based proportional β - diversity	$\frac{\Delta_T^p - \Delta_S^p}{\Delta_T^p}$	Additive decomposition of MPD. Δ_T^p is MPD measured in the regional pool. Δ_S^p is the average MPD within communities. See details in Hardy & Senterre (2007). Note that we used the coefficients described by the authors, not their estimators	Hardy & Senterre (2007)
<i>1B. Direct Dissimilarities</i>				
<i>Using all distances</i>				
	Dpw = COMDIST = Rao's D MPD-based β - diversity	$\sum_{i=1}^{S_1} \frac{1}{S_1} \sum_{j=1}^{S_2} \frac{1}{S_2} d_{ij}$	Mean phylogenetic distance between a species from assemblage 1 and a species from assemblage 2. Webb <i>et al.</i> 's COMDIST = Swenson's Dpw. S_1 and S_2 are the numbers of species in communities 1 and 2, respectively	Webb <i>et al.</i> (2008); Swenson (2011)
<i>Using nearest distances</i>				
	Dnn = COMDISTNT MNTD-based β - diversity	$\left(\frac{\sum_{i=1}^{S_1} \min_{k \text{ in } 2} (d_{ik}) + \sum_{i=1}^{S_2} \min_{k \text{ in } 1} (d_{ik})}{S_1 + S_2} \right)$	COMDISTNT uses the minimum pairwise distance (MNTD) for each taxon in a sample to all taxa in the other sample and calculates the mean. Same notations as above. 'k in 1' means species k from assemblage 1. Webb <i>et al.</i> 's COMDISTNT = Swenson's Dnn.	Webb <i>et al.</i> (2008); Swenson (2011)
	PCD Phylogenetic community dissimilarity	$PCD = \frac{n_1 PSV_{(1 2)} + n_2 PSV_{(2 1)}}{n_1 PSV_1 + n_2 PSV_2} \frac{1}{\bar{D}(n_1, n_2, C_{pool})}$ with $\bar{D}(n_1, n_2, C_{pool}) = \frac{n_1 \overline{psv}(i j)(n_2) + n_2 \overline{psv}(i j)(n_1)}{n_1 PSV_{pool} + n_2 PSV_{pool}}$	PSV is the α -diversity phylogenetic metric described earlier. tr is the trace of the matrix. Communities 1 and 2 have n_1 and n_2 species, respectively. C_{11} , C_{22} and C_{12} represent the covariance matrix of species of assemblage 1, 2 or between species of the two communities, respectively.	Ives & Helmus (2010)

		<p>and</p> $PSV_{(1 2)} = \frac{n_2 tr S_{22} - \sum S_{22}}{n_2(n_2 - 1)}$ $S_{22} = C_{22} - C'_{12} C_{11}^{-1} C_{12}$	<p>PSV_{pool} is the unconditional PSV calculated for all N species in the species pool, C_{pool} is their phylogenetic covariance matrix and $\bar{psv}_{(i j)}(n_j)$ is the mean conditional $PSV_{(i j)}$ for a community i, given the composition of n_j species randomly drawn from the species pool</p>	
2. Abundance data				
<i>2A. Decomposition into α, β, γ diversities</i>				
	<p>H_β Phylogenetic β-entropy</p>	$H_\beta = H_\gamma - H_\alpha$	<p>Additive decomposition of Allen et al. (2009) phylogenetic entropy (H_p). H_γ is calculated in the regional pool and H_α is the average diversity within communities</p>	<p>Mouchet & Moullot (2011)</p>
	<p>P_{ST} Rao's QE-based proportion abundance-weighted β-diversity</p>	$\frac{D_T^P - D_S^P}{D_T^P}$	<p>Additive decomposition of Rao's QE. D_T^P is Rao's QE calculated in the regional pool and D_S^P the average value of Rao's QE calculated per assemblage. Note here that we used the coefficients described by the authors, not their estimators</p>	<p>Hardy & Senterre (2007)</p>
	<p>B_{ST} InterMPD_{Ab}-based proportion abundance-weighted β-diversity</p>	$\frac{D_T^B - D_S^B}{D_T^B}$	<p>Additive decomposition of interMPD_{Ab}. This index is similar to P_{ST} but does not consider conspecific individuals comparisons (i.e. interMPD_{Ab} is used instead of Rao's QE)</p>	<p>Hardy & Senterre (2007); Hardy & Jost (2008)</p>

	Rao's DISC Raos' QE-based dissimilarity coefficient	$\sum_i \sum_j d_{ij} p_{i1} p_{j2} - 0.5 \left(\sum_i \sum_j d_{ij} p_{i1} p_{j1} + \sum_i \sum_j d_{ij} p_{i2} p_{j2} \right)$	Rao's dissimilarity coefficient based on Rao's QE; p_{i1} is the relative abundance of species i in assemblage 1, and p_{j2} the relative abundance of species j in assemblage 2.	Rao (1982)
<i>2B. Direct dissimilarities</i>				
<i>Using all distances</i>	Dpw _{Ab} (also named Dpw') Partially abundance- weighted version of the average species dissimilarity between two assemblages	$\sum_{i=1}^{S_1} p_i \frac{\sum_{k \text{ in } 2} d_{ik}}{S_2} + \sum_{j=1}^{S_2} p_j \frac{\sum_{k \text{ in } 1} d_{jk}}{S_1}$	Abundance-weighted version of Dpw/COMDIST. S_1 and S_2 are the numbers of species in communities 1 and 2, respectively	Swenson (2011)
	COMDIST _{Ab} = Rao's D _{Ab} Average species dissimilarity between two assemblages	$\sum_{i=1}^{S_1} p_{i1} \sum_{j=1}^{S_2} p_{j2} d_{ij}$	Alternative abundance-weighted version of Dpw/COMDIST. p_{i1} is the relative abundance of species i in assemblage 1, and p_{j2} the relative abundance of species j in assemblage 2	Rao (1982); Webb <i>et al.</i> (2008)
<i>Using nearest distances</i>	Dnn _{Ab} (also named Dnn') Average nearest- neighbour distance between two assemblages	$\sum_{i=1}^{S_1} \min_{k \text{ in } 2} (d_{ik}) p_i + \sum_{i=1}^{S_2} \min_{k \text{ in } 1} (d_{ik}) p_j$	Abundance-weighted version of Swenson's Dnn/COMDISTNT. S_1 and S_2 are the numbers of species in communities 1 and 2, respectively	Webb <i>et al.</i> , (2008); see Weiher & Keddy (1995) in a functional context
III. Parametric indices				
1. Equivalent numbers				
We also used Chiu <i>et al.</i> 's (2014) decomposition of Chao <i>et al.</i> 's (2010) adaptation of Hill numbers to phylogenetic diversity where				
${}^q D_\beta(T) = \frac{{}^q D_\gamma(T)}{{}^q D_\alpha(T)}$				
See Chiu <i>et al.</i> (2014) for definitions of parameters.				

2. Entropy				
	$I_{q\beta}$ Entropy-based parametric phylogenetic β - diversity	$I_{q\gamma} - I_{q\alpha}$	Additive decomposition of I_q . $I_{q\gamma}$ is I_q measured in the regional pool. $I_{q\alpha}$ is the average I_q within communities.	Pavoine <i>et al.</i> (2009)

Details on the links between several indices

Among the indices that use phylogenetic distances among species, some are very general in the definition of these distances (for example it might be sufficient that they are non-negative, however they have been calculated), others slightly restrict the choice of the distances (for example distances bounded between 0 and 1) and others impose a way of calculating the phylogenetic distances as for Helmus *et al.* (2007) PSV and PSE indices.

We start below with notations defined by Helmus *et al.* (2007).

PSV is a special case of MPD:

$$PSV = \frac{n \operatorname{tr}(\mathbf{C}) - \sum \mathbf{C}}{n(n-1)}$$

n is the number of tips (species), $\operatorname{tr}(\mathbf{C})$ is the trace of \mathbf{C} (sum of diagonal values) (see below for the definition of \mathbf{C}), $\sum \mathbf{C}$ is the sum of all values in \mathbf{C} .

First scenario:

\mathbf{C} is the matrix of covariances. Let us assume that there is no loss of generality but to make it more concrete that they are Brownian covariances. c_{ii} = sum of branch lengths from tip i to root; c_{ij} = sum of branch lengths from first common ancestor for i and j to root.

$$\operatorname{tr}(\mathbf{C}) = \sum_i c_{ii} \cdot \sum \mathbf{C} = \sum_{ij} c_{ij}$$

$$PSV = \frac{n \sum_i c_{ii} - \sum_{ij} c_{ij}}{n(n-1)} = \frac{\frac{1}{2} \left(n \sum_i c_{ii} + n \sum_j c_{jj} \right) - \sum_{ij} c_{ij}}{n(n-1)}$$

$$PSV = \frac{\frac{1}{2} \left(\sum_i \sum_j c_{ii} + \sum_i \sum_j c_{jj} \right) - \sum_i \sum_j c_{ij}}{n(n-1)}$$

$$PSV = \frac{\sum_i \sum_j \frac{1}{2}(c_{ii} + c_{jj} - 2c_{ij})}{n(n-1)}$$

Let $d_{ij} = \frac{1}{2}(c_{ii} + c_{jj} - 2c_{ij})$

$$PSV = \frac{\sum_i \sum_j d_{ij}}{n(n-1)} = MPD\{(d_{ij})\}$$

Here d_{ij} is half the sum of branch lengths in the shortest path that connects two tips (half patristic distances). For the dated tree, this would be the time to first common ancestor.

Second scenario:

PSV is applied to \mathbf{Z} , i.e. the matrix of correlations. We still assume no loss of generality but to make it more concrete we use Brownian covariances. c_{ii} = sum of branch lengths from tip i to root; c_{ij} = sum of branch lengths from first common ancestor for i and j to root.

$$z_{ii} = 1. \quad z_{ij} = c_{ij} / (\sqrt{c_{ii}c_{jj}})$$

$$\text{tr}(\mathbf{Z}) = n. \quad \sum \mathbf{Z} = \sum_{ij} c_{ij} / (\sqrt{c_{ii}c_{jj}})$$

$$PSV = \frac{n^2 - \sum_{ij} c_{ij} / (\sqrt{c_{ii}c_{jj}})}{n(n-1)} = \frac{\sum_{ij} [1 - c_{ij} / (\sqrt{c_{ii}c_{jj}})]}{n(n-1)}$$

Let $d_{ij} = 1 - c_{ij} / (\sqrt{c_{ii}c_{jj}})$, d_{ij} is bounded between 0 and 1 (see Pavoine & Izsák, 2014, for an application of $c_{ij} / (\sqrt{c_{ii}c_{jj}})$ as an index of phylogenetic similarities among species)

$$PSV = \frac{\sum_i \sum_j d_{ij}}{n(n-1)} = MPD\{(d_{ij})\}$$

Here d_{ij} is one minus the phylogenetic correlation between two tips i and j .

In both scenarios, Helmus *et al.*'s (2007) measure PSR is equal to

$$\frac{\sum_i \sum_j d_{ij}}{(n-1)}$$

PSE is a special case of Rao's QE:

$$PSE = \frac{m \text{diag}(\mathbf{C})' \mathbf{M} - \mathbf{M}' \mathbf{C} \mathbf{M}}{m^2 - \binom{m}{n}}$$

n is the number of tips; m_i is the abundance of tip i ; m is the sum of abundances across tips = $\sum_i m_i$; m/n is the average abundance per tip and \mathbf{M} is the vector of m_i for all i .

First scenario:

\mathbf{C} is the matrix of covariances (see notations above).

$$PSE = \frac{m \sum_i c_{ii} m_i - \sum_{ij} c_{ij} m_i m_j}{m^2 \binom{n-1}{n}} = \frac{\sum_i c_{ii} \frac{m_i}{m} - \sum_{ij} c_{ij} \frac{m_i m_j}{mm}}{\binom{n-1}{n}}$$

Let p_i be the relative abundance for tip i ,

$$PSE = \frac{\sum_i c_{ii} p_i - \sum_{ij} c_{ij} p_i p_j}{\binom{n-1}{n}} = \frac{\frac{1}{2} \left(\sum_i c_{ii} p_i + \sum_j c_{jj} p_j \right) - \sum_{ij} c_{ij} p_i p_j}{\binom{n-1}{n}}$$

$$PSE = \frac{\frac{1}{2} \left(\sum_{ij} c_{ii} p_i p_j + \sum_{ij} c_{jj} p_i p_j \right) - \sum_{ij} c_{ij} p_i p_j}{\binom{n-1}{n}}$$

$$PSE = \frac{\sum_{ij} \frac{1}{2}(c_{ii} + c_{jj} - 2c_{ij}) p_i p_j}{\binom{n-1}{n}}$$

Let $d_{ij} = \frac{1}{2}(c_{ii} + c_{jj} - 2c_{ij})$

Here (as above for PSV) d_{ij} is half the sum of branch length in the shortest path that connects two tips (half patristic distances). For a dated tree, this would be the time to first common ancestor.

$$PSE = \frac{n}{n-1} \sum_{ij} d_{ij} p_i p_j = \frac{n}{n-1} QE \{ (d_{ij}), \mathbf{p} \}$$

where \mathbf{p} is the vector of all p_i .

Second scenario:

PSE is applied to \mathbf{Z} , the matrix of correlations (see definition above for PSV). We still assume no loss of generality but to make it more concrete we use Brownian covariances. c_{ii} = sum of branch lengths from tip i to root; c_{ij} = sum of branch lengths from first common ancestor for i and j to root.

$$z_{ii} = 1. \quad z_{ij} = c_{ij} / (\sqrt{c_{ii} c_{jj}})$$

$$PSE = \frac{m \times \text{diag}(\mathbf{Z})' \mathbf{M} - \mathbf{M}' \mathbf{Z} \mathbf{M}}{m^2 - \binom{m}{n}}$$

$$PSE = \frac{m \sum_i m_i - \sum_{ij} (c_{ij} / (\sqrt{c_{ii} c_{jj}})) m_i m_j}{m^2 \binom{n-1}{n}} = \frac{1 - \sum_{ij} (c_{ij} / (\sqrt{c_{ii} c_{jj}})) \frac{m_i m_j}{mm}}{\binom{n-1}{n}}$$

$$PSE = \frac{\sum_{ij} (1 - c_{ij} / (\sqrt{c_{ii} c_{jj}})) \frac{m_i m_j}{mm}}{\binom{n-1}{n}}$$

Let $d_{ij} = 1 - c_{ij} / (\sqrt{c_{ii}c_{jj}})$, d_{ij} is bounded between 0 and 1 (see Pavoine & Izsák, 2014 for an application of $c_{ij} / (\sqrt{c_{ii}c_{jj}})$ as an index of phylogenetic similarities among species)

$$PSE = \frac{n}{n-1} \sum_{ij} d_{ij} p_i p_j = \frac{n}{n-1} \text{Rao's QE} \{(\mathbf{d}_{ij}), \mathbf{p}\}$$

Here d_{ij} is one minus the phylogenetic correlation among two tips. Rao's QE $\{(\mathbf{d}_{ij}), \mathbf{p}\}$ is Rao's QE applied to the d_{ij} 's and the vector of all p_i .

Appendix S2. Description of simulations run using *scape* - a phylogenetically informed community assembly simulation platform in the R package *pez*

This appendix describes the R-based simulation platform (Pearse *et al.*, 2015) (known hereafter as '*scape*') used to assemble communities within a spatially explicit landscape, *via* phylogenetic and biogeographical assembly mechanisms. In the *scape* simulation, 'taxa' with known phylogenetic relationships are assembled into local 'communities' which, when combined, define a 'landscape'.

Generating phylogenetic trees

Phylogenetic trees defining the evolutionary relationship between taxa were generated using the *sim.bdtree* function in the R package *geiger* (Harmon *et al.*, 2008). This function simulates trees under a uniform birth–death process, stopping when the desired number of tips is obtained. An existing phylogenetic tree could also be used. Trees must be made ultrametric before proceeding. We generated 100 random trees, each with 64 tips, observing indices of tree structure (δ statistic and I_c values) to ensure an even distribution of tree shapes was achieved.

Simulating landscapes

Landscape size is specified as the square of the number of communities forming a single side. This square landscape is also described spatially with paired X and Y coordinates, such that each community in the landscape has a unique pair of X and Y coordinates. Environmental conditions are defined along both the X and Y

dimension, currently, the same gradient of values occurs along both dimensions such that the environmental conditions in the environment are symmetrical along the diagonal. Taxa then have optima or 'niches' which relate to these environmental conditions. We generated landscapes in a 16 by 16 grid, thus comprising 256 total communities.

Once the number of taxa, their evolutionary relationships (i.e. the phylogeny), and the landscape size and conditions are decided, the desired assembly parameters must be chosen. In general, choices reflect whether there is a phylogenetic signal for species' environmental optima, how strong that signal is, and whether it has the effect of repulsion of related species (similar species are less likely to co-occur) or attraction of related species (similar species are more likely to co-occur) or neither. Where there is a phylogenetic signal for environmental optima, this is achieved by using the *corBlomberg* function in the R package *ape* (Paradis, Claude & Strimmer, 2004) to modify the phylogenetic tree's variance-covariance matrix accordingly. In the case of repulsion, the resulting pattern of co-occurrences is similar to that from competition, while for attraction it is more similar to the expectation from environmental filtering or facilitation. Parameters also control whether there is a phylogenetic signal for range size, and its strength. Related species may, for example, have similar range sizes.

Landscape characteristics were also varied so as to alter species' distributions. The average range width was set as a proportion of the total landscape size (here, 0.2).

This represents the number of cells in the landscape a species is, on average, likely to be found in. Where range size lacks a phylogenetic signal, species' ranges are drawn randomly from a normal distribution with a given mean size. Ranges may be spatially coherent, where species' probabilities of presence have a hump-shaped distribution. Range sizes need not be spatially coherent, in which case sites in which species are present are more dispersed through space.

Scape works by updating continually the probability of presence of each species on the landscape based on the assembly parameters chosen. The probability of presence may be affected by species' environmental optima, the presence of other species and their degree of relatedness, the average range size, the range size of similar species, and whether or not range sizes are identical. In addition, the amount of stochasticity in the application of these assembly rules can be adjusted by adjusting the amount of random variation in range size and species' environmental niche values.

The resulting matrix shows the probability of presence of each species for each site (i.e. a site * species matrix), which is used to produce a site-by-species presence/absence matrix (Fig. S1). A species abundance matrix, where probabilities are scaled by a provided maximum species abundance (K), was also produced. Here maximum abundance was set at 100 individuals per community. This matrix was used to calculate abundance-based metrics.

For each tree, we simulated eight types of communities meant to reflect a wide range of possible parameter combinations, from strong niche (and so, phylogenetically correlated) structure to random assembly (Table S1).

Landscape type	Parameter name							
	g.center	g.range	g.repulse	repulse	signal.center	signal.range	same.range	
1	0.2	1	0.2	TRUE	TRUE	FALSE	FALSE	
2	20	1	1	FALSE	TRUE	FALSE	FALSE	
3	0.2	0.2	0.2	TRUE	TRUE	TRUE	FALSE	
4	20	20	1	FALSE	TRUE	TRUE	FALSE	
5	1	1	1	FALSE	FALSE	FALSE	FALSE	
6	1	1	1	FALSE	FALSE	FALSE	TRUE	
7	1	1	1	FALSE	FALSE	TRUE	FALSE	
8	1	1	1	FALSE	FALSE	TRUE	TRUE	

Table S1. Parameter values used for the eight types of landscapes simulated using *scape*.

scape workflow

Inputs

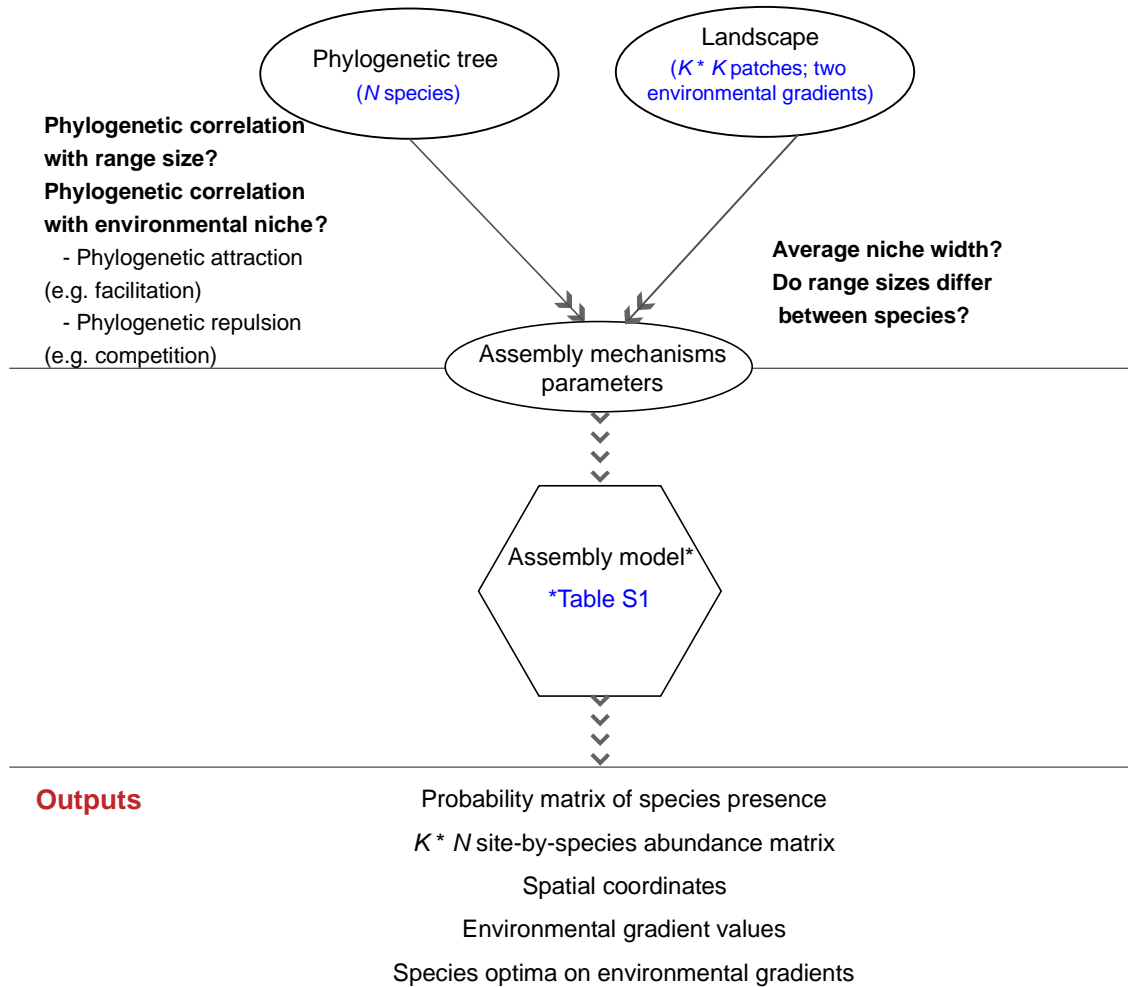


Fig. S1. *scape* workflow showing the necessary input information and the path by which these are used to produce output from the assembly model. R code for the *scape* simulation is available in the R package *pez* (Pearse *et al.*, 2015). K is the carrying capacity (maximum number of individuals) for each community.