

Quantifying spatial phylogenetic structures of fully stem-mapped plant communities

Guochun Shen^{1,2,3*}, Thorsten Wiegand², Xiangcheng Mi⁴ and Fangliang He^{1,5}

¹State Key Lab of Biological Control and School of Life Sciences, Guangdong Key Lab of Plant Resources, SYSU-Alberta Joint Lab for Biodiversity Conservation, Sun Yat-sen University, Guangzhou, Guangdong 510275, China; ²Department of Ecological Modelling, UFZ Helmholtz Centre for Environmental Research-UFZ, PF 500136, D-04301 Leipzig, Germany; ³Tiantong National Field Observation Station for Forest Ecosystem, East China Normal University, Shanghai 200241, China; ⁴State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, The Chinese Academy of Sciences, Beijing 100093, China; and ⁵Department of Renewable Resources, University of Alberta, 751 General Service Building, Edmonton, AB T6G 2H1, Canada

Summary

1. Analysis of the phylogenetic similarity of co-occurring species at different spatial scales is increasingly used for decoding community assembly rules. Here, we integrated the analysis of phylobetadiversity and marked point pattern analysis to yield a new metric, the phylogenetic mark correlation function, $k_d(r)$, to quantify spatial phylogenetic structure of fully stem-mapped communities.
2. $k_d(r)$ is defined as the expected phylogenetic distance of two heterospecifics separated by spatial distance r , and normalized with the expected phylogenetic distance of two heterospecifics taken randomly from a study area. It measures spatial phylogenetic turnover relative to spatial species turnover and is closely related with the spatially explicit Simpson index.
3. We used simulated fully stem-mapped plant communities with known spatial phylogenetic structures to assess type I and II errors of the phylogenetic mark correlation function $k_d(r)$ under a null model of random phylogenetic spatial structure, and to test the ability of the $k_d(r)$ to detect scale-dependent signals of phylogenetic spatial structure. We also compared the performance of the $k_d(r)$ with two existing measures of phylobetadiversity that have been previously used to analyse fully stem-mapped plots. Finally, we explored the spatial phylogenetic structure of a 24-ha fully stem-mapped subtropical forest in China.
4. Simulation tests showed that the new metric yielded correct type I and type II errors and accurately detected the spatial scales at which various processes (e.g. habitat filtering and competition) were invoked to generate spatial phylogenetic structures. The power of the $k_d(r)$ was not affected by a phylogenetic signal in species abundance and different topologies of the phylogenetic tree.
5. Replacing phylogenetic distance by functional distance allows for application of the $k_d(r)$ to estimate spatial correlations in functional community structure. Thus, the $k_d(r)$ allows trait and phylogenetic structure to be analysed in the same framework. The phylogenetic mark correlation function is a powerful and accurate tool for revealing scale-dependent phylogenetic/functional footprints in community assemblages and allows ecologists to keep up with the increasingly available data of fully stem-mapped plots, functional traits and community phylogenies.

Key-words: community assembly rules, phylogenetic mark correlation function, point pattern analysis, phylobetadiversity, scale dependence, spatial phylogenetic community pattern, type I and II errors

Introduction

One of the persistent challenges in community ecology is to explain how species coexist in communities, particularly in species-rich communities such as tropical forests (Chesson 2000). Numerous ecological and evolutionary processes have been identified to play roles in species coexistence and assembly of communities (Wright 2002), but their relative importance is

not well understood. Analysis of the phylogenetic similarity of co-occurring species has increasingly been used to aid in this task (Webb *et al.* 2002; Kembel & Hubbell 2006; Swenson *et al.* 2006; Kraft, Valencia & Ackerly 2008; Paine *et al.* 2012; Swenson 2013). For example, a relationship between local species co-occurrence and their phylogenetic relatedness may point to the operation of habitat filtering and/or competitive exclusion (Webb *et al.* 2002), and the phylogenetic relatedness of neighbours may be an important predictor of density-dependent mortality (Metz, Sousa & Valencia 2010; Paine *et al.* 2012).

*Correspondence author. E-mail: shengguochun255@hotmail.com

An important result emerging from studies of phylogenetic alpha diversity is that different processes such as competition and environmental filtering are expected to imprint signals at different spatial scales (Goldberg 1987; Webb *et al.* 2002; Johnson & Stinchcombe 2007; Swenson *et al.* 2012). For example, phylogenetic evenness (i.e. co-occurring species are phylogenetically more distantly related than expected by chance) is often observed at small scales and clustering (i.e. co-occurring species are phylogenetically more closely related than expected) at larger scales (Webb *et al.* 2002; Cavender-Bares, Keen & Miles 2006). This suggests that quantification of scale-dependent phylogenetic structure should be helpful in determining the relative importance of biotic and abiotic filters in governing community assembly if the niche conservatism assumption holds (Losos 2008).

Especially suitable for spatially explicit analyses of local-scale phylogenetic spatial structure are spatially referenced data, such as fully stem-mapped forest plots from the Center for Tropical Forest Science in which every tree with a diameter at breast height ≥ 1 cm has been identified, mapped and measured in an area of 16–50 ha (Condit 1998). Such data permit the detailed measurement of the spatial arrangement of individuals with respect to their ecological similarity and hence the analysis of correlation between spatial and phylogenetic distances of individuals, independently on the overall phylogenetic community structure. This provides means for hierarchical analyses of phylogenetic structure by decoupling analysis of the overall phylogenetic community structure of a plot (e.g. Kraft *et al.* 2007) and the smaller-scale spatial phylogenetic structures. For example, individuals of ecologically similar or dissimilar species may tend to be located close to each other (small-scale clustering or evenness, respectively).

Previous phylogenetic and functional analyses of community assemblages have generally considered alpha diversity and spatial scale (Swenson *et al.* 2012), but Graham & Fine (2008) and Swenson *et al.* (2012) proposed phylogenetic beta diversity (phylobetadiversity) as a complementary approach for assessing scale effects. For example, low phylobetadiversity and high species beta diversity at small spatial scales are expected if competitive exclusion limits local coexistence of phylogenetically closely related species in a similar habitat (Graham & Fine 2008). This suggests that metrics that relate phylobetadiversity across space to species beta diversity should be especially suitable for revealing scale effects in phylogenetic spatial structure. However, phylobetadiversity has rarely been used in the analysis of fully stem-mapped plots (but see Swenson *et al.* 2012).

To capitalize on the full power of phylogenetic analyses in community ecology when fully stem-mapped data are available, we propose to integrate the analysis of phylobetadiversity (Hardy & Senterre 2007; Graham & Fine 2008; Swenson *et al.* 2012) with marked point pattern analysis (Illian *et al.* 2008). Our method utilizes point pattern data where the spatial location and the species identity of every individual in the plot are known and where a phylogenetic distance measure between all pairs of species is available.

The core of our method is a metric based on mark correlation functions (Stoyan 1984; Schlather 2001; Illian *et al.* 2008). The new metric, the phylogenetic mark correlation function $k_d(r)$, builds on previous theory of species beta diversity (Shimatani 2001; Chave & Leigh 2002) and phylobetadiversity (Hardy & Senterre 2007; Graham & Fine 2008; Swenson *et al.* 2012). It is defined as the expected phylogenetic distance of two heterospecifics separated by spatial distance r , normalized with the expected phylogenetic distance c_d of two heterospecifics taken randomly from the plot. The $k_d(r)$ measures spatial phylogenetic turnover relative to spatial species turnover and is especially powerful for revealing phylogenetic patterns across spatial scale. An especially attractive feature of the phylogenetic mark correlation function is that it is able to measure the correlation between spatial and phylogenetic distances of individuals independent of the overall phylogenetic community structure. It can therefore complement alpha diversity measures of phylogenetic community structure (e.g. NRI and NTI; Webb *et al.* 2002; Kraft *et al.* 2007) by quantifying smaller-scale spatial phylogenetic structures.

The objectives of this study are to (i) present and exemplify the phylogenetic mark correlation function for analysing phylogenetic spatial structure in fully stem-mapped plots, (ii) assess the performance and accuracy of the method over a range of variations in spatial scale, phylogenetic topology and phylogenetic signal in abundance, and (iii) compare the performance of the $k_d(r)$ metric with that of others metrics used to analyse phylogenetic spatial structure in fully stem-mapped plots. We also show explicitly that the power of our metric is independent of a phylogenetic signal in species abundance and the topology of the underlying phylogeny.

We simulated fully stem-mapped plant communities with and without signals of scale-dependent spatial phylogenetic structure that resembles typical situations expected in real census plot data and determine the type I and II error rates of the phylogenetic mark correlation function. As a ‘proof of concept’ example, we applied the phylogenetic mark correlation function to data from the 24-ha fully stem-mapped Gutianshan subtropical forest census plot in China. Finally, we compared the performance of the new metric with two existing abundance weighted and quadrat-based metrics, the mean nearest phylogenetic dissimilarity and the mean pairwise phylogenetic dissimilarity (Swenson 2011).

Materials and methods

SIMULATION OF FULLY STEM-MAPPED COMMUNITIES

All simulated fully stem-mapped communities comprised approximately 12 000 individuals of 10 species that were distributed within a 300×300 m plot (Fig. 1). A phylogeny for each simulated community was generated by randomly clustering the tips according to Paradis (2012). Branch lengths of the phylogenetic tree in each simulated community were sampled from a gamma distribution. To quantify the sensitivity of the different metrics of phylobetadiversity to the topology of the phylogenetic tree, we varied the topology of the phylogeny among simulated communities by changing the shape and rate parameters of the gamma distribution (see detailed settings of the parameters and

examples of simulated phylogenies in Appendix S1). Phylogenetic relatedness between species was represented by a matrix of pairwise distances between the pairs of tips from the simulated phylogenetic tree using its branch lengths. Species abundances in each simulated community were generated by Brownian motion along a given phylogeny. The phylogenetic signal in the simulated species abundance was quantified by the K-statistic (Blomberg, Garland & Ives 2003) (see Appendix S1). We generated six types of fully stem-mapped communities based on four assembly rules (Table 1). Simulation of communities with higher species richness or more complex assembly rules was limited by extensive computational time.

(1) *Random placement*. The spatial pattern of the community is completely random. We simulated these communities by independently superimposing the patterns of individual species generated by homogeneous Poisson processes (Wiegand & Moloney 2004)

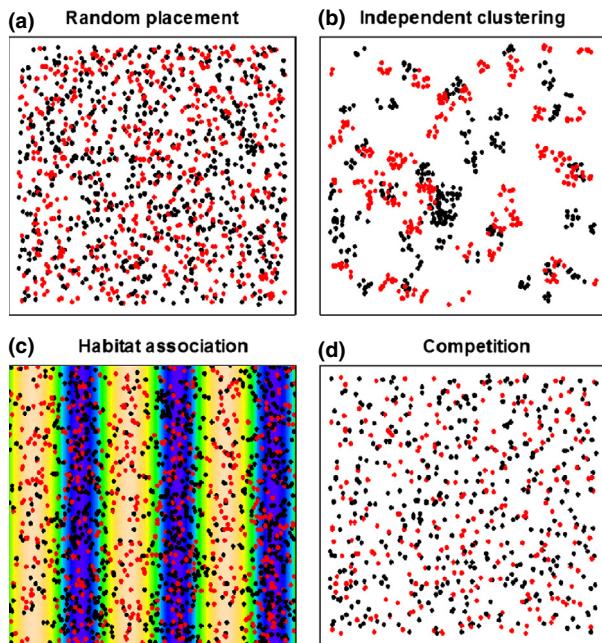


Fig. 1. Illustration of spatial patterns of the simulated communities for different assembly rules within 300×300 m plots. For clarity, we show only the spatial distributions of two of the ten species in the simulated community. Points with the same colour represent conspecifics. (a) Community generated by random placement, (b) community generated by independent cluster point processes that mimic dispersal limitation, (c) community generated by habitat association where the collared strips represent trough and peak variation in environmental variable $v(x)$ that determines the niches of the species and d) community with intra- and interspecific competition among individuals separated less than 5 m.

Table 1. Known spatial phylogenetic structure of the simulated fully stem-mapped communities that were based on four different assembly rules (see Materials and methods section for detail). For each assembly rule, we simulated 999 communities without spatial phylogenetic signal (scenarios c1–c4). Additionally, spatial phylogenetic signals were included in communities based on the habitat association and the competition assembly rules (scenarios c5 and c6). The expected phylogenetic patterns are random (r) in scenarios c1 to c4, phylogenetic clustering (+) in scenario c5, and in scenario c6, it is evenness (−) at distances $r < 5$ m and random (r) at distances $r > 10$ m. Note that the topology of the phylogenies and the phylogenetic signal in species abundance varied among all simulated communities (see Materials and methods section for detail)

	Random placement	Independent clustering	Habitat association	Competition
No phylogenetic signal	c1; (r)	c2; (r)	c3; (r)	c4; (r)
With phylogenetic signal			c5; (+)	c6; <5 m (−); >10 m (r)

(Fig. 1a). The random placement communities act as reference for communities that contain no spatial structure (i.e. no clustering, no co-occurrence or habitat association) and no phylogenetic signal (scenario c1 in Table 1).

(2) *Independent clustering*. The spatial structure of the simulated communities was characterized by intraspecific clustering, but different species were independently placed. We generated these communities by independently superimposing the patterns of individual species generated by a homogeneous Thomas process (Wiegand, Huth & Martínez 2009) with parameters $\mu = 8$ (i.e. an average of eight points per cluster) and $\sigma = 5$ m (i.e. 95% of all points were located closer than distance $2\sigma = 10$ m from the cluster centre) (Fig. 1b). This assembly rule corresponds to situations where the spatial pattern of the community is only determined by dispersal limitation, but not by habitat association or interspecific interactions (scenario c2 in Table 1). As a consequence, these communities contain no spatial phylogenetic structure.

(3) *Habitat association*. In this case, the spatial structure of the simulated communities is only driven by habitat association. For simplicity, we assumed that the heterogeneous environment was characterized by one variable $v(x)$ generated by a sine function along the x axis with period of $32 \pi m \approx 100$ m (Fig. 1c). Next, we assigned each species s an optimal niche value n_s that was drawn from an uniform distribution between −1 and 1. The intensity function $\lambda_s(x, y)$ of species s was given by $\lambda_s(x, y) = \lambda (1 + \sin(-n_s\pi + \pi^*x/100))$, where $\lambda = N_s/(300*300)$ is the density of species s in the study plot and N_s the abundance of species s . The distribution of each species s was generated by a heterogeneous Poisson processes based on $\lambda_s(x, y)$ (Wiegand & Moloney 2004). Thus, two species with $\Delta n_s = 0$ (Δn_s is the difference of two optimal niche values of the two species) have identical niches, and with $\Delta n_s = 1$ or -1 , they have most dissimilar niches. Because the optimal niche value of each species was not related with the phylogeny of the simulated community, the expected phylogenetic spatial structure was random at all spatial scales (scenario c3 in Table 1).

To generate stem-mapped communities with phylogenetic spatial structure caused by habitat association, we defined that the niche difference Δn_s between two species was highly correlated ($R^2_{adj} > 0.95$) with their phylogenetic relatedness. Because the period of the habitat is $32 \pi m$, we expect in this case significant phylogenetic clustering within 25 m (scenario c5 in Table 1).

(4) *Competition*. Communities driven by competition were based on intraspecific and interspecific interactions and were simulated by a multitype Strauss point process described in Geyer & Møller (1994). Direct competition was limited to plants that were located closer than 5 m. We assigned each pair a, b of species an index of ecological similarity $sim(a, b)$ that was randomly drawn from a uniform distribution between 0 and 1. Thus, phylogenetic spatial structure of these communities was expected to be random at all spatial scales (scenario c4 in Table 1).

To generate communities with phylogenetic spatial structure driven by competition, we positively correlated ($R^2_{\text{adj}} > 0.95$) the strength of competition between two species in the multitype Strauss point process with their phylogenetic relatedness. Thus, more similar species tended to locally exclude each other, and as a consequence, phylogenetic evenness was expected to occur for plants located at distances below 5 m (i.e. the range of direct competition) (scenario c6 in Table 1). However, we may find frequently cases where three plants are arranged linearly as ABC where the distance between A and C is just outside the range of competition (say 6–10 m). Because AB and BC will show large phylogenetic distances (due to competition), the pair AC may show a small phylogenetic distance. Therefore, we may find a tendency towards phylogenetic clustering just outside the range of competition which will smoothly disappear with increasing spatial distances (say twice the range of competition) (scenario c6 in Table 1).

AN EXAMPLE OF A REAL COMMUNITY

We applied the phylogenetic mark correlation function to data of the 24-ha Gutianshan (GTS) subtropical forest plot, China (see Legendre *et al.* 2009 for the description of this plot). We used the data of 17 707 living, large individuals (DBH > 10 cm), belonging to 107 species, of the first census. Phylogenies of the tree community in the GTS plot were constructed using three DNA sequence regions (*rbcL*, *matK* and *trnH-psbA*) in Gang *et al.* (2012). The phylogeny used in our analysis is shown in Fig. S2 in the Appendix S1.

DEFINITION OF THE PHYLOGENETIC MARK CORRELATION FUNCTION

Mark correlation functions allow testing if marks (e.g. size of a tree) of a point pattern (e.g. trees) are spatially correlated, conditional on the spatial locations of the points (Illian *et al.* 2008). A mark correlation function $k_t(r)$ yields the expectation of a test function $t(m_i, m_j)$ involving the marks m_i and m_j of two points i and j , taken over all pairs of points that are distance r apart and is normalized with the expectation of the test function $t(m_i, m_j)$ taken over all pairs of points regardless of their distances.

The phylogenetic mark correlation function $k_d(r)$ evaluates only heterospecific pairs of individuals and uses the phylogenetic distance $d(a, b)$ (or a functional distance) between two species a and b as test function. It therefore yields the expected phylogenetic distance of two heterospecific individuals separated by spatial distance r and is normalized with the expected phylogenetic distance c_d of two heterospecific individuals taken randomly from the plot. More formally, the $k_d(r)$ can be estimated as:

$$\hat{k}_d(r) = \frac{\sum_{i=1}^N \sum_{j=1}^N \mathbf{I}(sp_i \neq sp_j) d(sp_i, sp_j) \kappa(\|x_i - x_j\| - r)}{\sum_{i=1}^N \sum_{j=1}^N \mathbf{I}(sp_i \neq sp_j) \kappa(\|x_i - x_j\| - r)} \quad \text{eqn 1}$$

where the indicator function $\mathbf{I}(sp_i \neq sp_j)$ yields one if the individuals i and j are different species (i.e. $sp_i \neq sp_j$) and zero otherwise, the $d(sp_i, sp_j)$ is the phylogenetic (or functional) distance measure between individuals i and j of species sp_i and sp_j , respectively, the kernel function $\kappa(\|x_i - x_j\| - r)$ yields one if the individuals i and j are located at spatial distance $r \pm h/2$ (h is the bandwidth in the kernel function κ) and zero otherwise, and N is the total number of individuals in the surveyed area. The normalization constant c_d represents the overall phylogenetic community structure and is estimated as:

$$\sum_{i=1}^N \sum_{j=1}^N \mathbf{I}(sp_i \neq sp_j) d(sp_i, sp_j) / \sum_{s=1}^S N_s (N - N_s),$$

where S is the total number of species in the plot, and N_s is the abundance of species s .

LINK TO INDICES OF BETA DIVERSITY

Equation 1 can be rewritten to represent the $k_d(r)$ as the (normalized) ratio of two quantities representing phylobetadiversity $\beta_{\text{phy}}(r)$ and species beta diversity $\beta_S(r)$ (see Appendix S2 of Supporting Information):

$$k_d(r) = \frac{(\beta_{\text{phy}}(r)) / (\beta_S(r))}{\beta_{\text{phy}}^*} = \frac{1}{c_d} \frac{\beta_{\text{phy}}(r)}{\beta_S(r)} \quad \text{eqn 2}$$

where $\beta_S(r)$ is the spatially explicit Simpson index that measures species betadiversity (Shimatani 2001; Chave & Leigh 2002), β_S^* is the classical Simpson index (Simpson 1949), the $\beta_{\text{phy}}(r)$ is a phylogenetic extension of $\beta_S(r)$ which measures phylobetadiversity, the β_{phy}^* is the phylogenetic analogue of β_S^* , and $c_d = \beta_{\text{phy}}^* / \beta_S^*$.

First, the spatially explicit Simpson index $\beta_S(r)$ is a mark correlation function with test function $\mathbf{I}(sp_i \neq sp_j)$ that can be estimated as:

$$\hat{\beta}_S(r) = \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \mathbf{I}(sp_i \neq sp_j) \kappa(\|x_i - x_j\| - r)}{\sum_{i=1}^N \sum_{j=1}^N \kappa(\|x_i - x_j\| - r)}, \quad \text{eqn 3}$$

where the summation goes over all pairs of individuals i and j (with $i \neq j$). Thus, $\beta_S(r)$ yields the probability that two arbitrarily chosen individuals a distance r apart are heterospecifics and quantify species turnover.

Secondly, we generalized the spatially explicit Simpson index $\beta_S(r)$ to yield an index $\beta_{\text{phy}}(r)$ of spatial phylobetadiversity (Graham & Fine 2008). This can be done by replacing the binary test function $\mathbf{I}(sp_i \neq sp_j)$ in $\beta_S(r)$ (eqn 3) with a continuous measure $d(sp_i, sp_j)$ of phylogenetic (or functional distance) between two species sp_i and sp_j (Hardy & Senterre 2007). The ‘spatially explicit phylogenetic Simpson index’ $\beta_{\text{phy}}(r)$ is therefore the mean phylogenetic distance of all pairs of individuals a distance r apart and quantifies phylobetadiversity (Graham & Fine 2008). The β_{phy}^* is the phylogenetic analogue to β_S^* and yields the mean pairwise phylogenetic distances (MPD) of all individuals in the plot (i.e. the index D^P of Hardy & Senterre 2007).

Because $\beta_S(r)$ and $\beta_{\text{phy}}(r)$ quantify spatial species turnover and spatial phylogenetic turnover, respectively, the phylogenetic mark correlation function (eqn 1) can be interpreted as a measure of spatial phylogenetic turnover relative to the spatial species turnover (eqn 2). The constant $c_d = \beta_{\text{phy}}^* / \beta_S^*$ in eqns 1 and 2 yields the MPD of all heterospecific individuals in the plot and normalizes the phylogenetic mark correlation function $k_d(r)$ to a value of one if there is no spatial phylogenetic signal in the data (i.e. phylogenetic turnover is perfectly correlated with species turnover; Fig. S3 in Appendix S2). However, if individuals separated in space by distance r are more closely related than expected by species turnover, we have $k_d(r) < 1$ (i.e. phylogenetic clustering), and if they are more distantly related than expected, we have $k_d(r) > 1$ (i.e. phylogenetic evenness).

NULL MODELS OF RANDOM PHYLOGENETIC SPATIAL STRUCTURE

The selection of the null model is dependent on the objective of the study and the summary statistic used. In our study, the primary objec-

tive is to detect small-scale spatial phylogenetic structures such as phylogenetic evenness or clustering, generated by phylogenetically related processes. Thus, we need to test the null hypothesis of a community without phylogenetic spatial structure. This can be achieved by contrasting the observed data to null communities generated by the species shuffle null model that randomly shuffles the species label over the species present in the community phylogeny (Fine & Kembel 2011; Swenson *et al.* 2012). Because this randomization of the phylogenetic distance matrix $d(a, b)$ breaks down the actual phylogenetic relationships among species (while keeping the matrix elements intact), the resulting $k_d(r)$ functions are representative of random spatial phylogenetic structure, conforming the null hypothesis to be tested (Hardy & Senterre 2007). Note that this null model constrains all potentially confounding properties of the data and is particularly powerful for studies of beta diversity, because it fixes all observed spatial patterns except spatial phylogenetic structure (Hardy & Senterre 2007; Swenson *et al.* 2012).

SIGNIFICANCE TEST

We generated for each simulated community 999 Monte Carlo realizations of the species shuffle null model. To test departure of the observed $k_d(r)$ from that of the null model, we used the 25th largest and smallest values of $k_d(r)_{\text{null}}$ as approximate 95% simulation envelopes of $k_d(r)$ at spatial distance r . If the observed $k_d(r)$ was outside of the simulation envelopes, the community showed a significant departure from the null model at spatial distance r . If a community contains no phylogenetic spatial structure, the expectation yields $k_d(r) = 1$ and we expect $k_d(r) < 1$ for phylogenetic clustering and $k_d(r) > 1$ for phylogenetic evenness.

CALCULATIONS OF TYPE I AND TYPE II ERRORS

For each of the six types of simulated communities (Table 1), we generated 999 replicates and calculated the type I and type II error rates for the phylogenetic mark correlation function at different spatial scales. Type I and II errors were calculated as the percentage of cases (out of the 999 realizations) where a metric rejected a true hypothesis (according to Table 1) or accepted a false hypothesis, respectively. For communities simulated with the competition assembly rule, type II and I errors were calculated before 5 m and after 10 m, respectively (the expected errors between 5 to 10 m cannot be calculated accurately because this is a transition zone according to our competition setting).

COMPARISON WITH OTHER METRICS

We compared the performance of the phylogenetic mark correlation function with two metrics of phylobetadiversity that have been applied to fully stem-mapped plots, the abundance weighted mean nearest phylogenetic dissimilarity, $D'_{mn}(r)$, and the abundance weighted mean pairwise phylogenetic dissimilarity, $D'_{pw}(r)$. Both are estimated between local communities in two small quadrats that are distance r away (see detailed explanation of D'_{mn} and D'_{pw} in Appendix S3 and Swenson (2011) and Swenson *et al.* (2012)). A simulated community was divided into 10000 non-overlapping 3×3 m quadrats, but 1×1 m quadrats were used to evaluate the small-scale performance of the conventional metrics for the communities in scenario c6 (Table 1). Significance of phylogenetic spatial structure was also tested by the species shuffling null model.

IMPLEMENTATION

The algorithm for calculating the phylogenetic mark correlation function and the significance test under the species shuffling null model were implemented in C and embedded into the R environment (see <https://github.com/guochunshen/sce>) and in the software *Programita* (Wiegand & Moloney 2004). The software can be requested by the first two authors.

Results

TYPE I AND TYPE II ERRORS

The phylogenetic mark correlation function $k_d(r)$, when teamed with the species shuffle null model, performed well under the different communities without phylogenetic spatial structure. For all assembly rules, it yielded the expected type I error (around 0.05) at all spatial scales (blue lines in Fig. 2). In comparison, D'_{mn} and D'_{pw} yielded on average relatively large type I errors (Fig. 2).

The phylogenetic mark correlation function $k_d(r)$ yielded below 5% type II error rates (Fig. 3a) for communities where habitat associations were correlated with phylogenetic relatedness (i.e. closely related species tended to occur in the similar habitat; scenario c5 in Table 1). The D'_{mn} and D'_{pw} performed well at small scales, but showed larger type II error rate above the 20 m scales (Fig. 3a). The $k_d(r)$ correctly revealed the signal of competition at scales of 1–5 m (Fig. 3b), but D'_{mn} and D'_{pw} had high type II error rates within 5 m (Fig. 3b). The $k_d(r)$ correctly identified the absence of competition at distances larger than the double of the 5 m competition range, but D'_{mn} and D'_{pw} produced higher type I errors at distances above 10 m (Fig. 3c).

IMPACT OF A PHYLOGENETIC SIGNAL IN SPECIES ABUNDANCE AND THE TOPOLOGY OF THE UNDERLYING PHYLOGENY ON TYPE I AND TYPE II ERRORS

The strength of phylogenetic signal in species abundance (see the Blomberg' K -statistic and its P -value in Appendix S1) and the topology of the phylogeny (see the skewness and kurtosis of the phylogenetic distances among tips in the simulated community in Appendix S1) did not influence the type I and type II error rates of the phylogenetic mark correlation function across spatial scales (Figs 2–4). However, the performances of D'_{mn} and D'_{pw} were influenced by topology (skewness and kurtosis) of the phylogeny (Fig. 4). The error rates of D'_{mn} and D'_{pw} increased substantially with increasing strength of the phylogenetic signal in species abundance (Fig. 4).

PHYLOGENETIC MARK CORRELATION FUNCTION FOR THE SIMULATED AND REAL COMMUNITIES

To illustrate the behaviour and interpretation of the phylogenetic mark correlation function, we show in Fig. 5 the results for simulated fully stem-mapped communities and the GTS

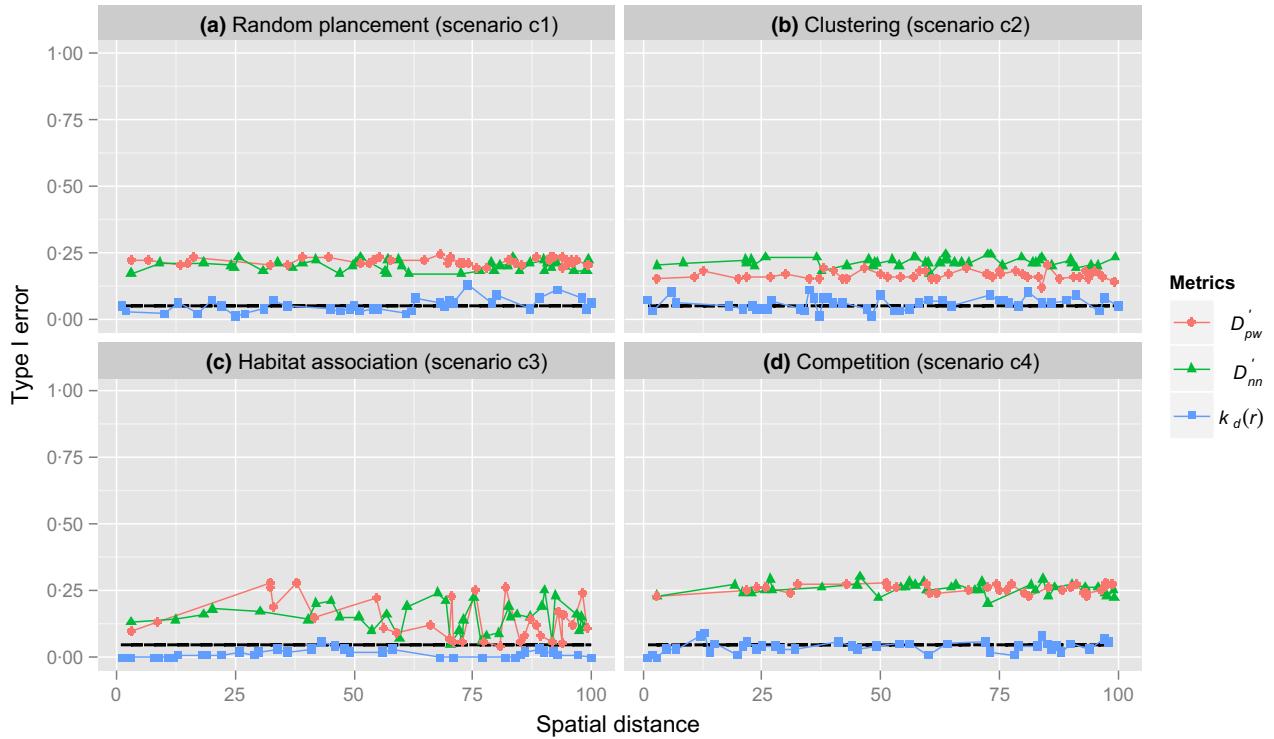


Fig. 2. Type I errors estimated for the phylogenetic mark correlation function $k_d(r)$ (blue lines), the abundance weighted mean nearest phylogenetic dissimilarity, $D'_{nn}(r)$ (green lines), and the abundance weighted pairwise phylogenetic dissimilarity, $D'_{pw}(r)$ (red lines) under the species shuffle null model across spatial scale. To assess type I error, we used simulated communities in scenarios c1–c4 in Table 1. A suitable metric should detect only in approximately 5% of all cases a significant effect (i.e. a 5% error; dashed black horizontal line).

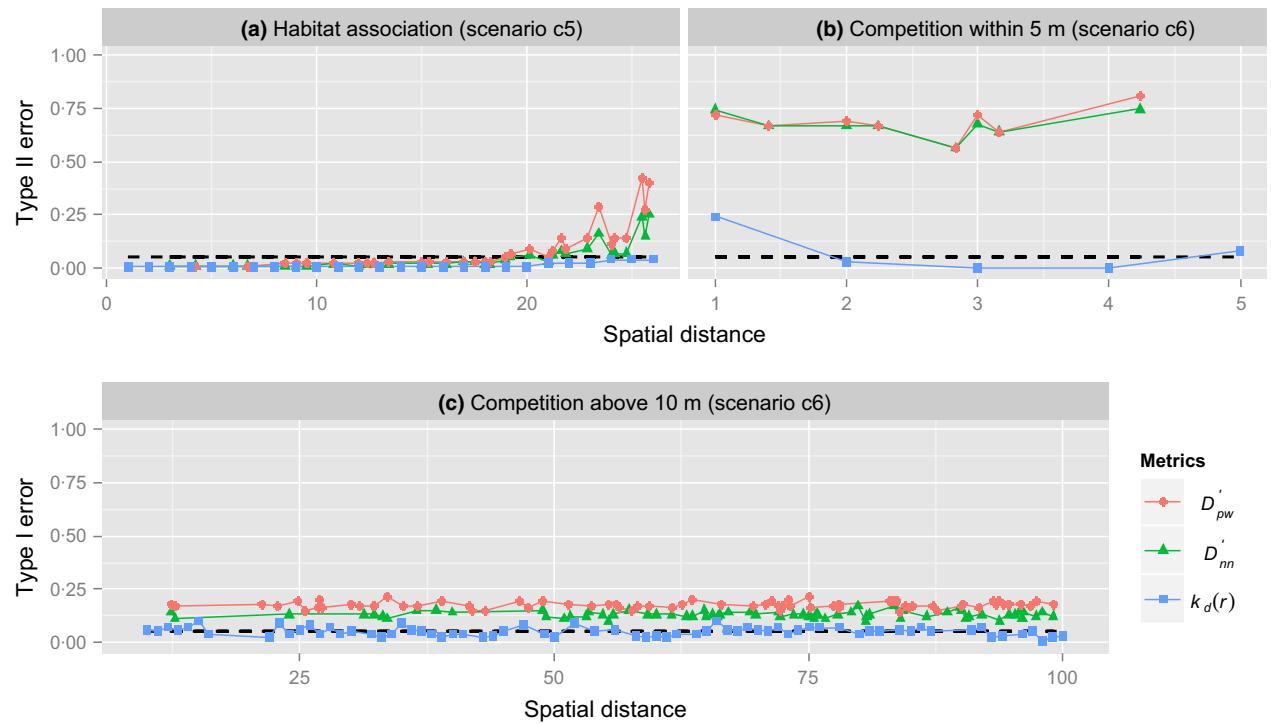


Fig. 3. Type II errors in dependence on spatial scale estimated for the phylogenetic mark correlation function $k_d(r)$ (blue lines), the $D'_{nn}(r)$ (green lines) and the $D'_{pw}(r)$ (red lines) under the species shuffle null model. To assess type II error, we used simulated communities where phylogenetic relatedness was correlated with the niche (panel a) or the pairwise competition strength (panels b and c). A suitable metric should detect departures from the null model in most cases (e.g. approximately 95% cases or 0.05 type II error). Because competition reached only up to 5 m, we calculated in scenario c6 type II error for distances $r \leq 5$ m (panel b), but type I error at distance $r \geq 10$ m (panel c).

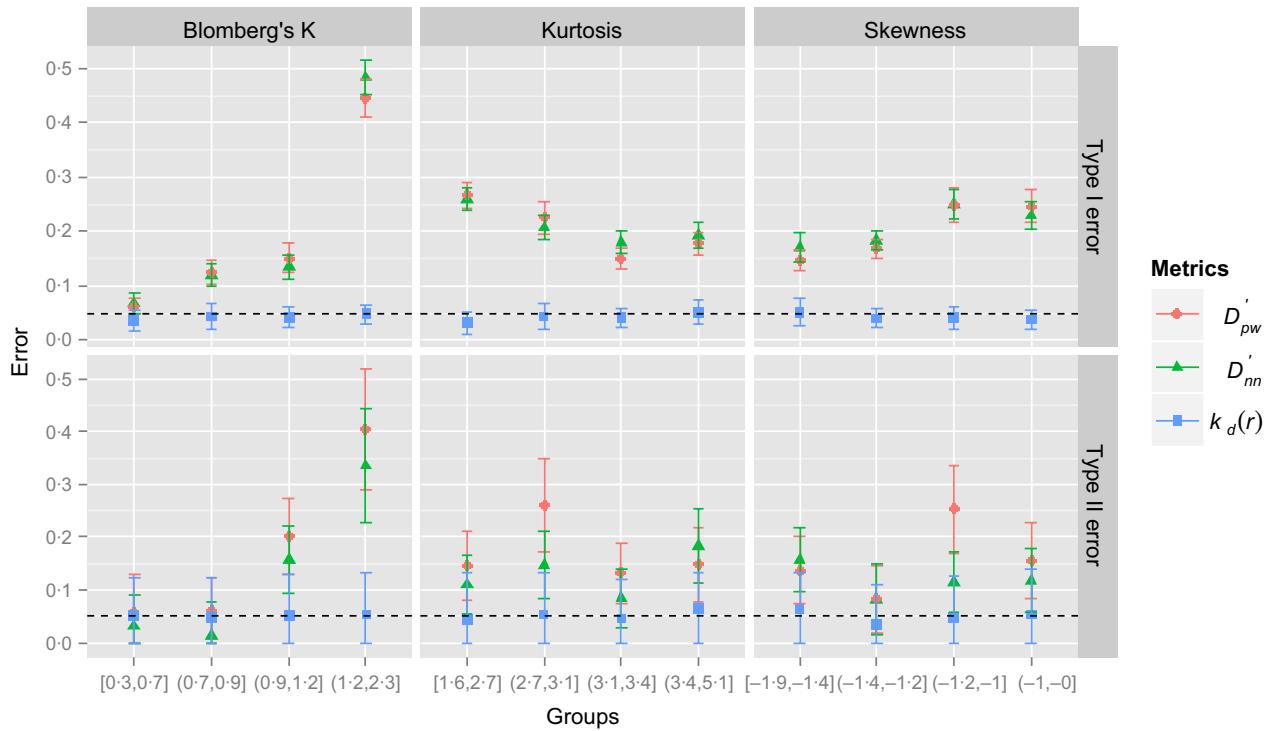


Fig. 4. Means (points) and standard deviations (vertical bars) of type I and II error rates of the three phylobetadiversity metrics under different groups of phylogenetic signal in species abundance, kurtosis and skewness of the phylogeny of the simulated communities. The strength of the phylogenetic signal in species abundance was quantified by Blomberg's K -statistic.

plot. The fully stem-mapped example community assembled by independent clustering showed strong spatial structure as expected by dispersal limitation (Fig. 1b), but no spatial phylogenetic structure was expected. As expected, the phylogenetic mark correlation function $k_d(r)$ was completely located within the simulation envelope of the species shuffle null model (Fig. 5a).

The habitat-driven community showed both spatial and phylogenetic structure because phylogenetic distance and habitat preference of species were correlated (Fig. 1c). The $k_d(r)$ recovered the spatial scales of the phylogenetic structure imprinted by the periodic habitat association (Fig. 5b; the period of sin like habitat was 10π in this example). It revealed that individuals at spatial distances $10n\pi$ ($n = 1, 2, \dots$) were phylogenetically more similar than expected (they showed maximal niche overlap, Fig. 5b) and more dissimilar at spatial distances $(1 + 2n)5\pi$ ($n = 0, 1, \dots$) (they showed minimal niche overlap, Fig. 5b).

In the community-driven by intraspecific and interspecific competition (Fig. 1d), phylogenetic distances between species were negatively correlated with the strength of competition between the two species (e.g. Metz, Sousa & Valencia 2010); thus, nearby individuals had a larger phylogenetic distance than expected. Indeed, Fig. 5c showed that the phylogenetic mark correlation function yielded values larger than the simulation envelope (i.e. phylogenetic evenness) at spatial distances within the competition range (<5 m) and within the simulation envelopes for individuals located more than 10 m apart. For individuals located in the transition zone between 5 and 10 m,

however, the phylogenetic distance was smaller than expected. This is the spatial correlation effect mentioned previously. The results of the phylogenetic mark correlation function were highly consistent among communities created by the same assembly rules (Fig. S4 in Appendix S4).

Finally, the phylogenetic mark correlation function of the real fully stem-mapped GTS tree community revealed that two heterospecific individuals (DBH > 10 cm) located at spatial distances <90 m were on average less related than expected under the species shuffle null model (Fig. 5d).

Discussion

Analysis of the ecological and evolutionary similarity of co-occurring species has increasingly been used to determine the processes underlying the diversity and assembly of communities (Swenson 2013). Here, we integrated previous work on phylobetadiversity (Hardy & Senterre 2007; Graham & Fine 2008; Swenson *et al.* 2012) with marked point pattern analysis (Schlather 2001; Illian *et al.* 2008) to yield a framework that is especially adapted to data sets of fully stem-mapped plots that include the exact position of all individuals of a community (e.g. Condit 1998).

The phylogenetic mark correlation function $k_d(r)$ we presented measures phylogenetic turnover at spatial distance r relative to the corresponding species turnover. The $k_d(r)$ together with the species shuffle null model performed consistently well in a wide range of situations. Because the species shuffle null model only randomized phylogenetic spatial structure while

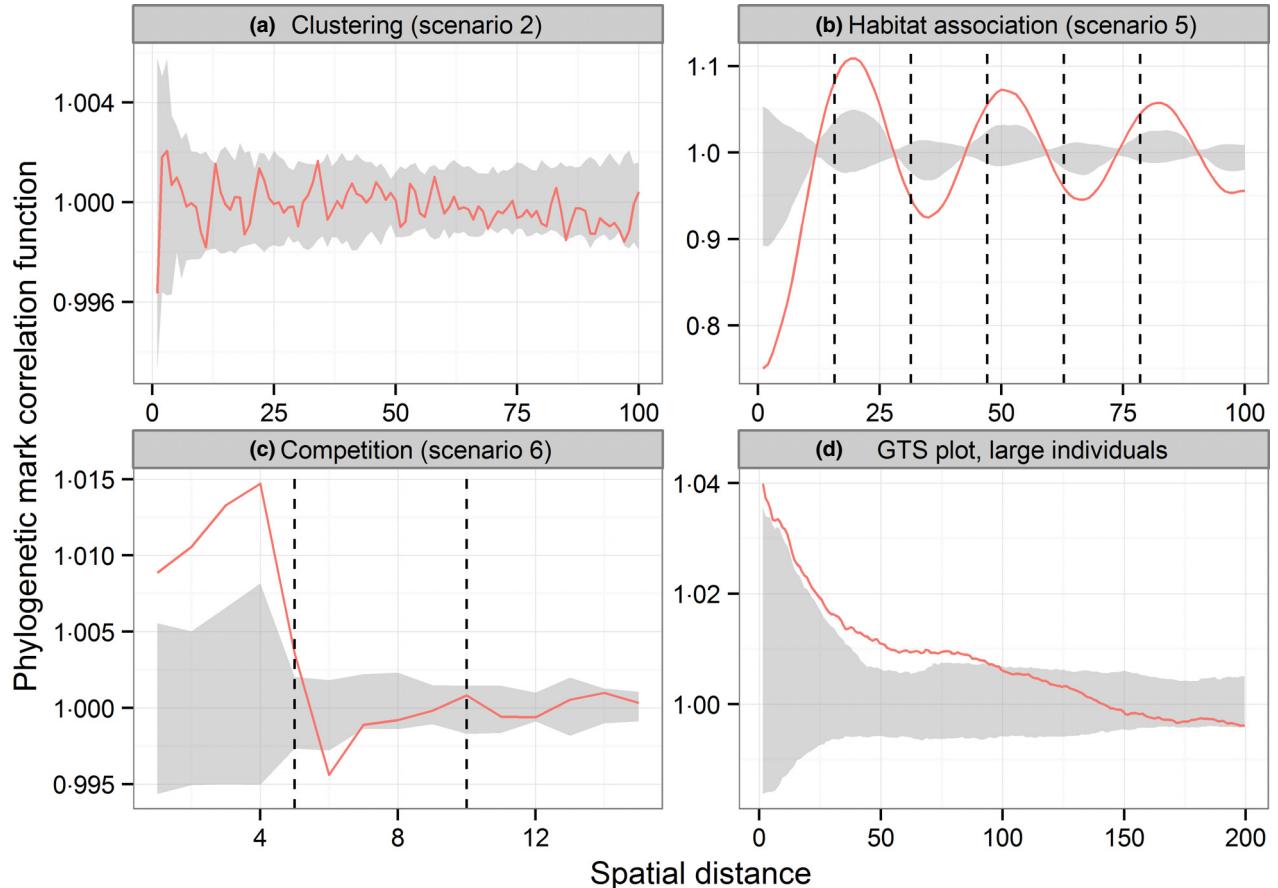


Fig. 5. Shapes of the phylogenetic mark correlation function (red lines) in simulated communities assembled by independent clustering without phylogenetic signal (panel a; c2 in Table 1), habitat association with a phylogenetic signal (panel b; c5 in Table 1, 10π period of sine waved environment along x axis), competition with a phylogenetic signal (panel c; c6 in Table 1, competition occurred within 5 m distance) and the data from the GTS forest plot (panel d). Spatial distances in bottom panels were log-transformed. The grey area represents the 95% simulation envelopes under the species shuffle null model. Phylogenetic evenness or clustering is indicated where the phylogenetic mark correlation function falls above or below the simulation envelopes, respectively. Vertical dashed lines are the reference lines at $5n\pi$ ($n = 1, 2, \dots$) distances (panel b) and 5, 10 m distance (panel c).

conditional on all spatial structures (Swenson *et al.* 2012), our analyses factored out the potentially confounding effects of spatial structures (e.g. caused by dispersal limitation) that were independent of phylogenetic relatedness (i.e. scenarios c3 and c4 in Table 1). Importantly, the $k_d(r)$ precisely detected the effect of processes that generated scale-dependent phylogenetic structure (i.e. scenarios c5 and c6 in Table 1). Additionally, our method is explicitly designed to allow trait and phylogenetic structure to be analysed in the same, and hence in a directly comparable, framework (see Baraloto *et al.* 2012 for a common framework for community-wide phylogenetic and functional structure). The distance matrix that defines in the $k_d(r)$ distances between species can be based equally on phylogenetic or functional distance. These features make the metric $k_d(r)$ a powerful tool for revealing scale-dependent phylogenetic or functional spatial structures in fully stem-mapped plant communities and should substantially enhance our ability to infer ecological processes which are expected to imprint phylogenetic or functional signals at different spatial scales (Webb *et al.* 2002; Graham & Fine 2008; Swenson *et al.* 2012).

An interesting feature of the $k_d(r)$ is that it allows for an analysis of the correlation between spatial and phylogenetic distances of individuals, independent of the overall phylogenetic community structure. This has two consequences. First, this separation of scales between non-spatial effects on the plot scale and local spatial effects allows for specific and unbiased assessment of phylogenetic signal that directly influence the locations of individuals. Secondly, because the $k_d(r)$ is independent of the overall phylogenetic community structure, its power is not affected by a phylogenetic signal in species abundance and the topology of the phylogeny which confounded most of the conventional phylogenetic metrics (Kraft *et al.* 2007; Hardy 2008). Our results in Figs 2–4 confirmed this expectation from the conditional property of $k_d(r)$. The mathematical reason for this is that the $k_d(r)$ is normalized with the term c_d (i.e. the mean pairwise phylogenetic distances of all heterospecific individuals in the plot, in eq. 1), which controls the influence of the overall phylogenetic community structure. As a consequence, the $k_d(r)$ is exclusively focused on the correlation between spatial and phylogenetic distances of individuals. In contrast, the two measures of phylobetadi-

versity D'_{mn} and D'_{pw} mix up phylogenetic patterns caused by spatial arrangement of individuals and overall phylogenetic community structure. As a consequence, a phylogenetic signal in species abundance and the topology of phylogeny influence the power of D'_m and D'_{pw} under the species shuffling null model.

Because a principal motivation in the development of the phylogenetic mark correlation function is application to species-rich communities, future work should test the performance of the method for species-rich communities. Systematic variation in local diversity (e.g. due to habitat structuring) may potentially affect the statistical properties of our methods. However, restricting the analysis to single habitat (e.g. Kembel & Hubbell 2006) can remove much of this effect (T. Wiegand unpublished analysis). The highly consistent performance of the phylogenetic mark correlation for the data sets presented here suggests that it may perform equally well for more complicated communities.

Finally, although the phylogenetic mark correlation function has an outstanding ability to detect spatial phylogenetic structures, caution is needed in interpreting these patterns because many processes can generate similar spatial phylogenetic structure in ecological communities (Losos 2008). Once spatial structures are correctly detected, more complex rules for simulating null communities in concert with field experiments are required to make further inference on the underlying processes.

Conclusions

The proposed new method has broad ramifications. It provides a general framework for developing other metrics for analysis of the correlation between spatial distance of individuals and their phylogenetic or functional distance. For example, we can define a cumulative mark correlation function using all pairs of individuals with spatial distances $< r$ instead of spatial distances $\approx r$. This cumulative metric quantifies the expected phylogenetic distance of two heterospecifics separated by spatial distances smaller than r , normalized with the expected phylogenetic distance c_d of two heterospecifics taken randomly from the plot. Additionally, a wide array of analyses is possible by selecting subsets of the full community for the pairs of individuals analysed. First, we can conduct species-centred analyses where the first individual of the pair is taken from a given focal species and the second individual is selected from all other species in the community. This analysis supplements the community-wide analysis and reveals which species drive the observed community-wide spatial phylogenetic structures and if all species show the same or opposed patterns. Secondly, we may analyse phylogenetic spatial structures between different life stages such as juveniles and adult trees in the community by restricting the focal tree to adult trees and the second tree to juvenile trees. Finally, we can find out if dead trees (taken as first individual of the pair) and surviving trees (taken as the second individual of the pair) show spatial phylogenetic structure; for example, dead trees may be surrounded by more phylogenetically similar species than expected (Metz, Sousa & Valencia

2010). In this case, the appropriate null model is to conduct ‘random labelling’ (Wiegand & Moloney 2004) where the mark ‘dead’ is randomly re-allocated over all surviving and dead individuals in the data set analysed. Adopting the framework of phylogenetically marked point patterns will thus allow ecologists to keep up with the increasingly available data of fully stem-mapped plots, species traits and community phylogenies to empower the inference on processes that shape community assemblages.

Acknowledgments

This work was supported by Sun Yat-sen University and NSERC (Canada) to FH, the ERC advanced Grant 233066 to TW, the NSFC 31170401 to XC and the NSFC 31100309 to GS. We thank Jinlong Zhang for preparing the molecular phylogeny of the GTS plot. We also thank Luke Harmon and five referees for their constructive suggestions and comments.

References

- Baraloto, C., Hardy, O.J., Paine, C.E., Dexter, K.G., Cruaud, C., Dunning, L.T. *et al.* (2012) Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *Journal of Ecology*, **100**, 690–701.
- Blomberg, S.P., Garland, T. Jr & Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.
- Cavender-Bares, J., Keen, A. & Miles, B. (2006) Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology*, **87** (Supplement), S109–S122.
- Chave, J. & Leigh, E.G. (2002) A spatially explicit neutral model of beta-diversity in tropical forests. *Theoretical Population Biology*, **62**, 152–168.
- Chesson, P. (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, **31**, 343–366.
- Condit, R. (1998) *Tropical Forest Census Plots*. Springer, Berlin.
- Fine, P.V.A. & Kembel, S.W. (2011) Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography*, **34**, 552–565.
- Gang, F., Zhang, J.L., Pei, N.C., Rao, M.D., Mi, X.C., Ren, H.B. & Ma, K.P. (2012) Comparison of phylobetadiversity indices based on community data from Gutianshan forest plot. *Chinese Science Bulletin*, **57**, 623–630.
- Geyer, C.J. & Möller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, **21**, 359–373.
- Goldberg, D.E. (1987) Neighborhood competition in an old-field plant community. *Ecology*, **68**, 1211–1223.
- Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.
- Hardy, O.J. (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.
- Hardy, O.J. & Senterre, B. (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology*, **95**, 493–506.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester, UK.
- Johnson, M.T.J. & Stinchcombe, J.R. (2007) An emerging synthesis between community ecology and evolutionary biology. *Trends in Ecology & Evolution*, **22**, 250–257.
- Kembel, S.W. & Hubbell, S.P. (2006) The phylogenetic structure of a neotropical forest tree community. *Ecology*, **87**, 86–99.
- Kraft, N.J.B., Valencia, R. & Ackerly, D.D. (2008) Functional traits and niche-based tree community assembly in an Amazonian forest. *Science*, **322**, 580–582.
- Kraft, N.J.B., Cornelius, W.K., Webb, C.O. & Ackerly, D.D. (2007) Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist*, **170**, 271–283.
- Legendre, P., Mi, X., Ren, H., Ma, K., Yu, M., Sun, I.F. & He, F. (2009) Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, **90**, 663–674.
- Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1003.

- Metz, M., Sousa, W. & Valencia, R. (2010) Widespread density-dependent seedling mortality promotes species coexistence in a highly diverse Amazonian rainforest. *Ecology*, **91**, 3675–3685.
- Paine, C.E.T., Norden, N., Chave, J., Forget, P.-M., Fortunel, C., Dexter, K.G. & Baraloto, C. (2012) Phylogenetic density dependence and environmental filtering predict seedling mortality in a tropical forest. *Ecology Letters*, **15**, 34–41.
- Paradis, E. (2012) *Analysis of Phylogenetics and Evolution With R* (Second Edition). Springer, New York.
- Schlather, M. (2001) On the second-order characteristics of marked point processes. *Bernoulli*, **7**, 99–117.
- Shimatani, K. (2001) On the measurement of species diversity incorporating species differences. *Oikos*, **93**, 135–147.
- Simpson, E.H. (1949) Measurement of diversity. *Nature*, **163**, 688.
- Stoyan, D. (1984) On correlations of marked point processes. *Mathematische Nachrichten*, **116**, 197–207.
- Swenson, N.G. (2011) Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *PLoS ONE*, **6**, e21264.
- Swenson, N.G. (2013) The assembly of tropical tree communities - the advances and shortcomings of phylogenetic and functional trait analyses. *Ecography*, **36**, 264–276.
- 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.
- Swenson, N.G., Erickson, D.L., Mi, X., Bourg, N.A., Forero-Montana, J., Ge, X. et al. (2012) Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology*, **93**, S112–S125.
- 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.
- Wiegand, T., Huth, A. & Martínez, I. (2009) Recruitment in tropical tree species: revealing complex spatial patterns. *The American Naturalist*, **174**, E106–E140.
- Wiegand, T. & Moloney, K.A. (2004) Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**, 209–229.
- Wright, J.S. (2002) Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, **130**, 1–14.

Received 22 May 2013; accepted 17 September 2013

Handling Editor: Robert B. O'Hara

Supporting Information

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

Appendix S1. Distributions of Blomberg's K, skewness and kurtosis of the phylogeny in the simulated communities.

Appendix S2. Link between the phylogenetic mark correlation function and previous indices of beta diversity.

Appendix S3. Definition of the abundance weighted mean nearest phylogenetic dissimilarity and the abundance weighted mean pairwise phylogenetic dissimilarity.

Appendix S4. Results of phylogenetic pattern analysis for selected test communities.