ECOLOGY LETTERS

Ecology Letters, (2011) **14**: 141–149 doi: 10.1111/j.1461-0248.2010.01563.x

LETTER

Spatial patterns of phylogenetic diversity

Hélène Morlon[1]*, Dylan W. Schwilk[2], Jessica A. Bryant[1,3], Pablo A. Marquet[4,5,6], Anthony G. Rebelo[7], Catherine Tauss[8], Brendan J. M. Bohannan[1] and Jessica L. Green[1,6]

Abstract

Ecologists and conservation biologists have historically used species–area and distance–decay relationships as tools to predict the spatial distribution of biodiversity and the impact of habitat loss on biodiversity. These tools treat each species as evolutionarily equivalent, yet the importance of species' evolutionary history in their ecology and conservation is becoming increasingly evident. Here, we provide theoretical predictions for phylogenetic analogues of the species–area and distance–decay relationships. We use a random model of community assembly and a spatially explicit flora dataset collected in four Mediterranean-type regions to provide theoretical predictions for the increase in phylogenetic diversity – the total phylogenetic branch-length separating a set of species – with increasing area and the decay in phylogenetic similarity with geographic separation. These developments may ultimately provide insights into the evolution and assembly of biological communities, and guide the selection of protected areas.

Keywords

Community phylogenetics, conservation, distance–decay relationship, evolutionary history, Mediterranean-type ecosystems, phylogenetic beta-diversity, phylogenetic diversity, spatial scaling, species–area relationship.

Ecology Letters (2011) 14: 141–149

INTRODUCTION

Community ecologists and conservation biologists are increasingly analysing phylogenetic information and community data in tandem (Webb *et al.* 2002; Purvis *et al.* 2005; Forest *et al.* 2007; Cavender-Bares *et al.* 2009; Vamosi *et al.* 2009; Winter *et al.* 2009; Devictor *et al.* 2010). For example, the phylogenetic structure of local communities is compared with that of larger species pools to understand the processes driving community assembly (Webb *et al.* 2002; Heard & Cox 2007; Graham & Fine 2008; Cavender-Bares *et al.* 2009; Kraft & Ackerly 2010). Similarly, phylogenetic diversity (PD) is mapped across landscapes to select conservation areas that optimize the preservation of evolutionary history (Rodrigues & Gaston 2002; Ferrier *et al.* 2007; Forest *et al.* 2007; Winter *et al.* 2009; Devictor *et al.* 2010).

Despite the growing interest in PD, spatial biodiversity research has historically been centred on patterns of species diversity. Hundreds of publications have documented the species–area relationship (Preston 1962; Rosenzweig 1995), which describes the increase in species richness with geographic area. Touted as one of the few general laws in ecology, the species–area relationship has been crucial to the development of ecological theory (Preston 1962; MacArthur & Wilson 2001; Chave *et al.* 2002) and for estimating extinction risk in the face of environmental change (Pimm & Askins 1995; Guilhaumon *et al.* 2008). Similarly, analytical characterizations of the curve describing how the similarity in species composition between two communities decays with the geographic distance separating them (the distance–decay relationship) have been used to infer the relative importance of dispersal limitation and environmental filtering in explaining patterns of diversity (Preston 1962; Nekola & White 1999; Chave & Leigh 2002; Condit *et al.* 2002; Morlon *et al.* 2008), and to predict the complementarity of sites within reserve networks (Ferrier *et al.* 2007).

In contrast to the decades of research on the spatial scaling of species diversity, research on the spatial scaling of PD remains in its infancy. Empirical observations of the increase of PD with area (Rodrigues & Gaston 2002), and of the decay in phylogenetic similarity with geographic or environmental distance (Chave *et al.* 2007; Hardy & Senterre 2007; Bryant *et al.* 2008) have recently emerged. However, there have been no attempts to generalize the shape or mathematical form of these diversity patterns. This is a major gap, given that patterns explicitly incorporating information on evolutionary history will likely be more powerful than patterns that do not (such as the species–area and distance–decay relationships) for testing, and estimating parameters of, biodiversity theory (Jabot & Chave 2009). Furthermore, phylogeny-based spatial patterns are needed for setting conservation priorities aimed at protecting evolutionary history in a spatial context (Rodrigues & Gaston 2002; Purvis *et al.* 2005; Ferrier *et al.* 2007; Winter *et al.* 2009; Devictor *et al.* 2010).

There are three main determinants to the spatial scaling of PD: the spatial scaling of species diversity, the phylogenetic tree describing the evolutionary history of these species and their position in the phylogeny. In turn, these three components are driven by multiple evolutionary and ecological processes, including speciation and

[1]*Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR, USA*
[2]*Texas Tech University, Lubbock, TX, USA*
[3]*Massachussetts Institute of Technology, Cambridge, MA, USA*
[4]*Center for Advanced Studies in Ecology and Biodiversity and Departamento de Ecología, Pontificia Universidad Católica de Chile, Santiago, Chile*
[5]*Instituto de Ecología y Biodiversidad, Casilla 653, Santiago, Chile*
[6]*Santa Fe Institute, Santa Fe, NM, USA*
[7]*South African National Biodiversity Institute, Kirstenbosch, South Africa*
[8]*School of Plant Biology, University of Western Australia, Perth, Australia*
**Correspondence and present address: College of Natural Sciences, University of California, Berkeley, CA, USA. E-mail: hmorlon@berkeley.edu*

extinction, dispersal limitation, environmental filtering, and intra- and inter-specific interactions. Recently, much focus has been given to the third component (the position of co-occurring species in a phylogeny), often referred to as community phylogenetic structure. Phylogenetic structure measures the extent to which species assemblages deviate from random assemblages and has been used as a tool to infer the processes underlying community assembly (Webb *et al.* 2002; Heard & Cox 2007; Graham & Fine 2008; Cavender-Bares *et al.* 2009; Kraft & Ackerly 2010).

In this article, we use a model where species are randomly assembled with respect to phylogeny to derive predictions for the spatial scaling of PD in the absence of phylogenetic structure. This reduces the task to two well-studied problems, usually considered separately in the literature: modelling spatial patterns of species diversity, and modelling cladogenesis. Under the random assembly model, the link between species-based diversity patterns and the spatial scaling of PD is given by the species–PD curve, which describes how PD increases with an increasing number of species randomly sampled from a given phylogeny (Fig. 1). The species–PD curve has been studied in conservation, as it provides estimates for the potential loss of PD due to extinctions (Nee & May 1997; Heard & Mooers 2000; Diniz-Filho 2004; Purvis *et al.* 2005; Soutullo *et al.* 2005). The species–PD curve is a function only of the underlying phylogeny, not the spatial configuration of communities, and can thus be studied using models of cladogenesis developed in macroevolution (Nee & May 1997; Heard & Mooers 2000; Nee 2006; Morlon *et al.* 2010).

We first derive testable predictions of how PD increases with geographic area, and how phylogenetic similarity decays with geographic distance. We then demonstrate the validity of these predictions in nature using a spatially explicit dataset collected in the four Mediterranean-type ecosystems of Australia, California, Chile and South Africa. Finally, we discuss implications of our study for community ecology, biogeography and conservation.

MATERIALS AND METHODS

Mediterranean flora data

Data for woody angiosperms in the Mediterranean climate shrublands of Australia, California, Chile and South-Africa were collected between April and December 2006 (see Appendix S1 of Supporting Information). On each continent, we sampled 30 quadrats (120 quadrats total), separated by geographic distances ranging from 20 m (adjacent) to 170 km (Appendix S1). Within each quadrat, presence/absence data were recorded at the 2.5 × 2.5, 7.5 × 7.5 and 20 × 20 m scales, except in California where data were only recorded at the 20 × 20 m scale (Fig. S1). We sampled in a relatively homogeneous flora and environment within each Mediterranean-type ecosystem. Specifically, plots were sampled on the same parent material, and slope, aspect and fire history were kept as constant as possible. A total of 538 species encompassing 254 genera and 71 families were identified: 177 in the Australian kwongan, 27 in the Californian chaparral, 44 in the Chilean matorral and 290 in the South African fynbos (Fig. S2).

Phylogenetic construction

We used a megatree approach to construct a hypothesized dated phylogenetic tree for the species present in our dataset (Webb & Donoghue 2005). We first built an angiosperm backbone tree by supplementing the Phylomatic2 phylogenetic data repository (http://svn.phylodiversity.net/tot/trees/), which is based on resolutions from the Angiosperm Phylogeny Group, with additional data found in the literature (Appendix S8). We then grafted the 538 species in our dataset onto the backbone tree; the resulting phylogeny is thereafter

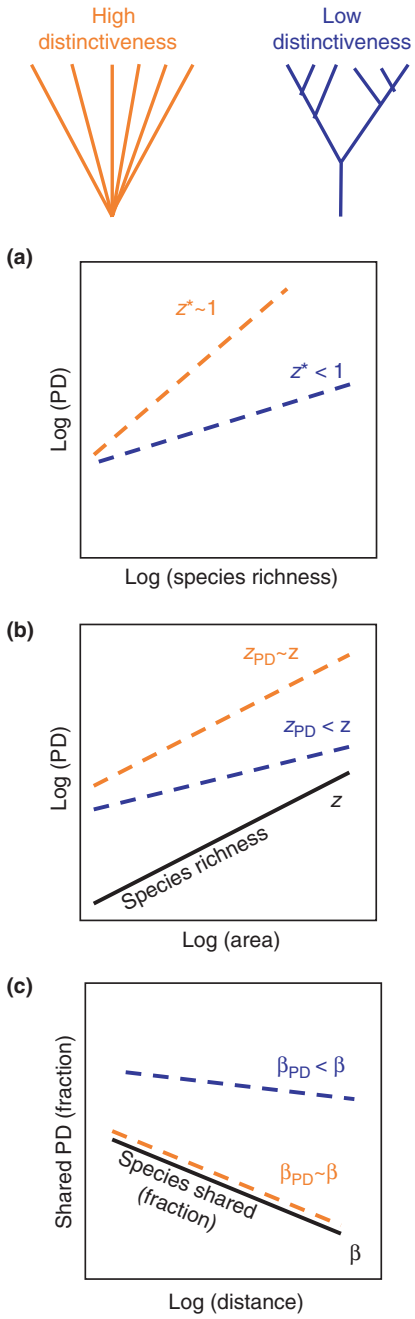


Figure 1 Conceptual figure illustrating, under the random community assembly model, the expected effect of phylogenetic tree shape on the relationship between (a) phylogenetic diversity (PD) and species richness, (b) PD and habitat area and (c) phylogenetic similarity and geographic distance. Here the PD of a set of species is measured as the phylogenetic branch-length joining all species in the set to the root. Star-like phylogenetic trees (with high distinctiveness, in orange) are characterized by steep species–PD curves (slope $z^* \sim 1$). Phylogenies with decreasing distinctiveness (in blue) have shallower species–PD curves, resulting in shallower PD–area curves, and shallower phylogenetic distance–decay curves.

referred to as the 'full phylogeny'. To assign branch-lengths, we spaced undated nodes evenly between dated ones using a slightly modified version of the Branch Length Adjuster (BLADJ) algorithm, as described next (Webb *et al.* 2008; Cam Webb, personal communication; code available at: http://www.schwilk.org/research/data.html). The full phylogeny included terminal nodes that were not species. Specifically, the full phylogeny had 874 terminal nodes, 538 of which corresponded to the species in the dataset; the 336 remaining terminal nodes were families or genera in the backbone tree with no representative in the dataset. To ensure that we included during the branch-length assignment procedure all clades for which a node age estimate was available (Wikström *et al.* 2001), we fixed the terminal nodes corresponding to family or genera to their estimated ages before running the BLADJ algorithm. The phylogeny of the entire dataset (the 'combined phylogeny') was then obtained by removing nodes with no representative in the dataset. Individual phylogenies for each of the four regional datasets (the 'regional phylogenies') were obtained by pruning to the corresponding set of species (Appendix S1).

Phylodiversity metrics

There are several ways to measure PD within and among communities (see Lozupone & Knight 2008; Vamosi *et al.* 2009; Cadotte *et al.* 2010 for reviews). Given that our goal was to build spatial phylogenetic patterns readily comparable with the species-based species–area and distance–decay relationships, we chose metrics that most closely capture the notion of total amount of evolutionary history contained within, and shared between, communities. In addition, we excluded abundance-based metrics (Chave *et al.* 2007; Cadotte *et al.* 2010) because we collected incidence data only.

We quantified the PD of a given sample (alpha diversity) as the total phylogenetic branch-length joining the basal node (here the angiosperm node) to the tips of all the species in the sample ('PD'; Faith 1992). This metric is proportional to species richness for a star phylogeny (i.e. a phylogeny where species share no branch-length), rendering comparisons with the traditional species–area relationship possible. PD has the added advantage of being the phylodiversity metric of choice in conservation research (Faith 1992; Nee & May 1997; Rodrigues & Gaston 2002; Purvis *et al.* 2005; Forest *et al.* 2007; Winter *et al.* 2009). Diversity metrics based on pairwise taxon distances between species (Chave *et al.* 2007; Hardy & Senterre 2007) are not proportional to species richness for a star phylogeny, and they are rarely used for conservation purposes (Cadotte *et al.* 2010). Faith's PD retains the root of the species pool phylogeny, and this may reduce the variance in PD among samples (Crozier 1997; Crozier *et al.* 2005). However, as illustrated next, including the root is useful for constructing metrics of phylogenetic beta-diversity.

We quantified the phylogenetic similarity between two communities (an inverse measure of phylogenetic beta-diversity) with the incidence-based PhyloSor index χ_{PD}, which measures the PD shared between communities (noted $PD_{1,2}$) divided by the average PD in each community: $\chi_{PD} = \frac{PD_{1,2}}{\frac{1}{2}(PD_1 + PD_2)}$, where PD_1 and PD_2 represent the PD of each community (Bryant *et al.* 2008). Equivalently, $\chi_{PD} = \frac{PD_1 + PD_2 - PD_{1+2}}{\frac{1}{2}(PD_1 + PD_2)}$, where PD_{1+2} is the PD of the two communities combined. This index is closely related to indices suggested by Ferrier *et al.* (2007) to measure complementarity for conservation purposes, as well as to the Unifrac metric, widely used in microbial ecology research (Lozupone & Knight 2008). For a star phylogeny, the Phylosor index reduces to the Sorenson index of similarity, which is commonly used to characterize distance–decay relationships (Preston 1962; Nekola & White 1999; Morlon *et al.* 2008). If the root is not retained in the calculation of PD, $PD_1 + PD_2 - PD_{1+2}$ can take negative values (e.g. if communities 1 and 2 are composed of distinct, distantly related clades), which is biologically unrealistic.

Random assembly hypothesis

Our approach to deriving predictions for the increase of PD with area and the decay in phylogenetic similarity with geographic distance is to assume that the curves describing the increase in species richness with area and the decay in species similarity with geographic distance are known. This approach allows leveraging decades of research on the species–area and distance–decay relationships to understand how PD is distributed spatially.

Once species richness and species spatial turnover are known across a landscape, there are several ways to map a given phylogeny onto this landscape. We chose the simplest approach, which is to randomly assign a tip to each species in the landscape. This random assembly model is increasingly being used in community phylogenetics and consists of randomizing the position of species on a phylogeny while keeping species richness and turnover constant (Bryant *et al.* 2008; Graham *et al.* 2009). This model corresponds to the hypothesis that species are randomly assembled with respect to phylogeny within and across communities. Here, our primary interest in using this model is to provide a tractable theoretical approach for investigating spatial PD patterns.

To evaluate the validity of the random assembly hypothesis in our data, we tested for deviations from the random assembly model at each spatial scale within each 20×20 m plot. To do this, we compared the total PD of the observed communities with that of communities composed of the same number of species assembled by random sampling from each regional phylogeny. We also compared the observed phylogenetic similarity between pairs of communities, sampled at the 20×20 m scale, with that of communities composed of, and sharing, the same number of species assembled by random sampling from each regional phylogeny. In other words, we randomized species across the tips of regional phylogenies while holding alpha- and beta-diversity constant (Bryant *et al.* 2008; Graham *et al.* 2009; Appendix S2).

Spatial PD theory predictions

Our spatial phylogenetic theory predictions build on the random assembly hypothesis and the observation that, if there exists a consistent relationship between PD and an increasing number of species randomly sampled in a phylogeny (the species–PD curve), then spatial patterns of PD may be deduced from this curve (Fig. 1). We obtained species–PD curves for each of the four regional phylogenies and for the combined phylogeny by randomly sampling an increasing number of species in each phylogeny, 100 times at each richness value. For comparison with previous studies, we fitted a logarithmic function to the observed species–PD curves, which is the only published analytical prediction for species–PD curves we are aware of (equation 1 in Nee & May 1997). Sensitivity analyses were conducted to evaluate the influence of polytomies and the BLADJ branch-length assignment procedure on the observed species–PD curve (Appendix S3).

Using the best-fit functional form for the species–PD curve in our data, we derived theoretical predictions for the increase of PD with area and the decay of phylogenetic similarity with geographic distance under the random assembly hypothesis. To test the accuracy of these predictions, we compared the predicted PD–area relationship and decay in phylogenetic similarity with geographic distance in each region with the 95% confidence envelopes of the curves obtained by simulations of the random assembly process (Appendix S4).

We also tested the ability of the random assembly process to reproduce the observed spatial PD patterns in each region. To do this, we computed the observed PD–area relationship by quantifying PD at the 2.5×2.5, 7.5×7.5 and 20×20 m scales in each of the 30 quadrats (except in California where data were only collected at the 20×20 m scale), and the decay in phylogenetic similarity with geographic distance by quantifying χ_{PD} between each pair of communities (435 pairs in each regional dataset) at the 20×20 m scale. We compared the observed relationships with the 95% confidence envelopes of the curves obtained by simulations of the random assembly process (Appendix S4).

All analyses were carried out using the Picante software package implemented in R (Kembel *et al.* 2010).

RESULTS

Random assembly hypothesis

Within each of the four Mediterranean flora datasets, most communities did not significantly deviate from the random assembly model (Fig. S3). Similarly, the fraction of PD shared between most pairs of communities within each dataset was not significantly different than that expected by chance given their species richness and fraction of species shared (Fig. S4). The dataset was thus ideal for testing predictions about the increase in PD with area, and the decay in phylogenetic similarity with geographic distance, under the random assembly model.

Species–PD curves and the shape of regional phylogenies

When an increasing number of species (S) were randomly drawn in each regional phylogeny, the corresponding increase in PD (species–PD curve) was well approximated by a power-law relationship (Fig. 2). This pattern also held for the combined phylogeny (Fig. S5). The power-law shape was robust to the presence of polytomies and the branch-length assignment procedure (Appendix S3), suggesting that it was not an artefact of the method of phylogenetic construction.

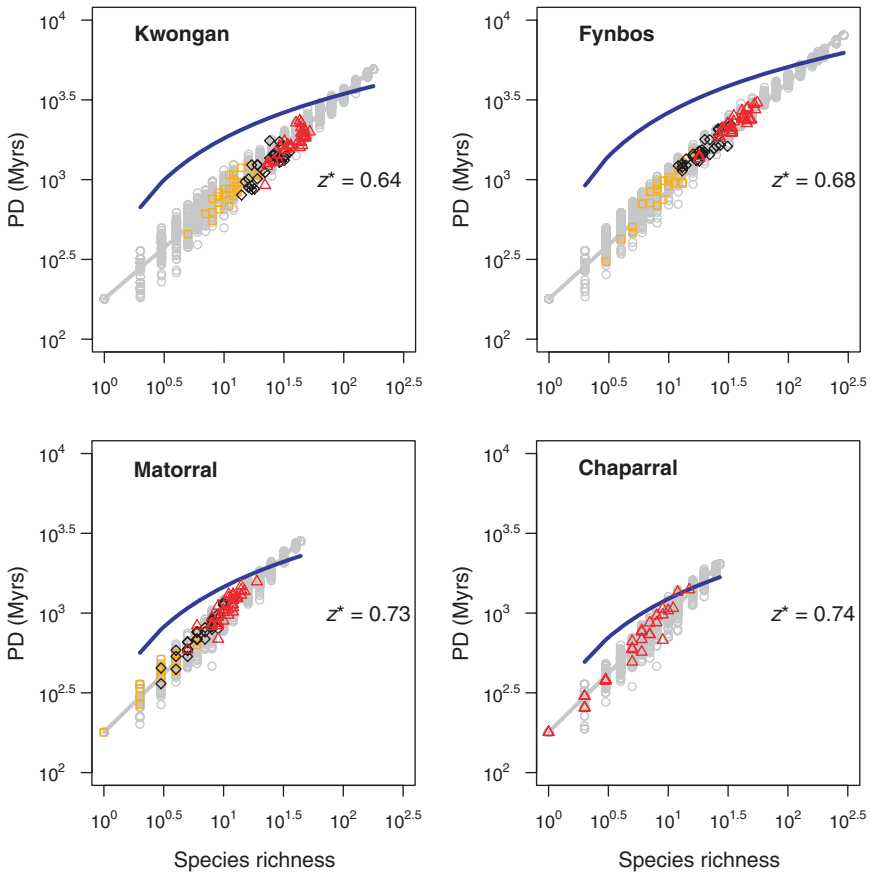


Figure 2 Species–phylogenetic diversity (PD) curves in Mediterranean-type ecosystems. The grey circles report, for each value of species richness (S), the PD of 100 communities obtained by randomly sampling S species across the tips of each phylogeny (species–PD relationship). This relationship is well fit by a power law in the four phylogenies (eqn 1, plain grey line). In particular, the power-law fit is much better than the best-fit logarithm (in blue). The intercept of both fits is constrained by the age of the most recent common ancestor, T_0. The species–PD curve corresponding to the combined dataset is also power law, with $z^* = 0.71$ (Figure S3). Coloured data points correspond to actual communities. Orange squares: communities sampled at the 2.5×2.5 m scale; black diamonds: communities sampled at the 7.5×7.5 m scale; red triangles: communities sampled at the 20×20 m scale. Most communities are not significantly different from randomly assembled communities (see Appendix S2 for details).

In particular, the power law provided a much better fit to the species–PD curve than the logarithmic function (Fig. 1 and Appendix S3).

A power-law species–PD relationship takes the form:

$$PD(S) \sim T_0 S^{z^*} \quad (1)$$

with the normalization constant given by the age T_0 of the most recent common ancestor in the phylogeny. This expression provides an expectation for the PD of a community containing S species, under the random assembly hypothesis. This expression also characterizes the species–PD curve by a single exponent z* ($z^* \leq 1$) which captures information about the phylogenetic distinctiveness of species (i.e. how evolutionarily unique species are relative to one another within a phylogeny; Vane-Wright *et al.* 1991; Fig. 1a). High z* values correspond to trees with high distinctiveness (typically, trees with long terminal branches and high imbalance), while low z* values correspond to trees with low distinctiveness (i.e. trees with short terminal branches and low imbalance). We found z* values ranging from above 0.7 in the matorral and chaparral, to 0.68 in the fynbos and 0.64 in the kwongan. z* values were slightly lower in the kwongan and fynbos due to the presence of closely related species in floras that radiated recently (Richardson *et al.* 2001).

We used the power-law species–PD curve to characterize the relationship between phylogenetic distinctiveness, the spatial distribution of species and spatial patterns of PD (Fig. 1). We used the power law because it is a convenient mathematical approximation, and also because it may be general to many phylogenetic trees. We observed a power-law relationship in all four datasets we studied. This consistency across datasets suggests generality, given that less than 25% of PD was shared between any two datasets. In cases where the power-law approximation is not accurate, our approach may be readily modified to account for alternative characterizations of species–PD curves (see next).

Increase of PD with area

Under the hypothesis that species assemblages are random with respect to phylogeny at each spatial scale, and assuming the power-law scaling between PD and species richness (eqn 1), the expected PD contained in a sample of area A is given by:

$$PD(A) \sim T_0[S(A)]^{z^*}, \quad (2)$$

where S(A) is the expected number of species contained in a sample of area A (the species–area relationship). A classic form of the species–area relationship is the power law:

$$S(A) = cA^z, \quad (3)$$

where c is a normalization constant, and z typically varies around the value of 0.25 (Rosenzweig 1995). While variations around the power-law species–area curve are common (Guilhaumon *et al.* 2008), the power law yielded a good description of the increase of species richness with area in our data (Fig. 3). The shape of the PD–area relationship may then be characterized by a power law with exponent z_{PD}, the product of the power-law exponent z of the species–area relationship and of the power-law exponent z* of the species–PD curve:

$$PD(A) \sim T_0 c^{z^*} A^{z_{PD}} = T_0 c^{z^*} A^{zz^*}. \quad (4)$$

This equation provides an expectation for the PD of a community spanning an area A, under the random assembly hypothesis. The power-law PD–area curve is shallower than the species–area curve by a factor z*, showing that PD increases with area at a slower pace than species richness (Fig. 1b). The power-law species–area and PD–area curves imply that if a fraction x of a given area is preserved, a fraction x^z of species is preserved (eqn 3), corresponding to a fraction x^{zz^*} of preserved PD (eqn 4). Equation 4 may be used to provide estimates for the loss of PD with habitat loss (see Appendix S5 for estimates in Mediterranean-type ecosystems).

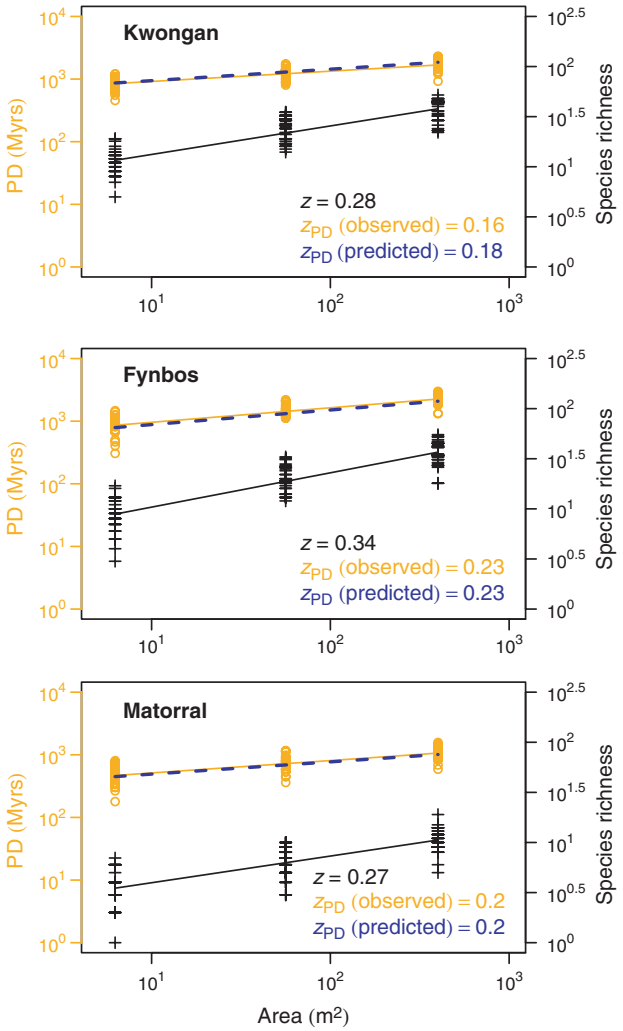


Figure 3 The increase of phylogenetic diversity (PD) with area in Mediterranean-type ecosystems. The observed PD–area relationship (in orange: circles, data; line, power-law fit) is well approximated by an expectation (eqn 4, in blue) obtained by simple power transformation of the classical species–area relationship (in black: crosses, data; line, power-law fit). The power-law exponent z_{PD} of the PD–area relationship is well approximated by the product of the power-law exponent of the species–area relationship z and the power-law exponent of the species–PD relationship z*. PD increases with habitat area at a slower pace than species, and the difference is the largest in floras where species are the least phylogenetically distinct (i.e. in the kwongan and fynbos).

The PD–area relationships observed in the three Mediterranean-type ecosystems were well described by eqn 4, which is based on power-law scaling relationships (Figs 3 and S10). Other forms of the species–PD curve and species–area relationship may better describe other systems. This would yield different shapes for the PD–area relationship that could be derived using a similar approach (Appendix S6).

Decay of phylogenetic similarity with geographic distance

To derive expectations for the decay in phylogenetic similarity with geographic distance, we maintained our assumption that communities are randomly assembled with respect to phylogeny. Using the power-law scaling between PD and species richness, we found (Appendix S7) that the expected fraction of PD shared between two communities, each spanning an area A, and separated by geographic distance d is given by:

$$\chi_{PD}(A,d) \sim 2 - (2 - \chi(A,d))^{z^*}, \quad (5)$$

where $\chi(A,d)$ is the expected Sorensen index of similarity. This equation confirms, as expected intuitively, that communities share a greater fraction of PD than species ($\chi_{PD}(A,d) \geq \chi(A,d)$).

To further formalize the scaling between phylogenetic similarity and geographic distance, we assumed a logarithmic model for the species-based distance–decay relationship of the form $\chi(A,d) = \alpha + \beta \log_{10}(d)$. We chose the logarithmic model because it provided a good fit to our data (Fig. 4). The logarithmic model has been observed in tropical forest communities, and has the additional value of being the predicted beta-diversity pattern under the neutral theory of biodiversity (Chave & Leigh 2002; Condit *et al.* 2002). With this model, and under the random assembly hypothesis, the expected shape of the phylogeny-based distance–decay relationship may also be described by a logarithmic function (Appendix S7):

$$\chi_{PD}(A,d) \sim \alpha_{PD} + \beta_{PD} \log_{10}(d) \quad (6)$$

with $\alpha_{PD} = 2 - (2 - \alpha)^{z^*}$ and $\beta_{PD} = \beta \frac{z^*}{(2-\alpha)^{1-z^*}}$.

Equation 6 provides an expectation for the fraction of PD shared between two communities spanning an area A and separated by a distance d. Although deviations from this equation occurred (e.g. in the kwongan and fynbos; Figs 4 and S11), the equation yielded a good description of the data in the matorral and chaparral. Equation 6 suggests that the rate of decay in phylogenetic similarity (β_{PD}) is less than the rate of decay in species similarity (β). This suggests that, within reserve networks, a greater spatial separation between protected sites will be required to preserve PD relative to the spatial extent required to preserve species richness.

Across Mediterranean-type ecosystems, no species were shared. The ecosystems that have been historically connected by landmasses and/or share geological attributes (e.g. California–Chile, Australia–South Africa and Chile–South Africa) were more phylogenetically similar (respective χ_{PD} values obtained by pulling all species within each dataset: 0.28, 0.26, 0.20) than Mediterranean-type systems that have been separated by oceans for longer time periods and/or are geologically very distinct (e.g. Australia–Chile, Australia–California and California–South Africa, χ_{PD} value $\sim$ 0.18 for all three pairs). When no species are shared and under the random model of community assembly, eqn 4 suggests that the phylogenetic similarity between the two communities equals $2 - 2^{z^*}$. The phylogenetic similarity between datasets was much lower than this expectation, reflecting dispersal limitation across continents acting over evolutionary time scales.

DISCUSSION

Although there has been an explosion of community phylogenetics papers in the last few years, no study has clearly identified the

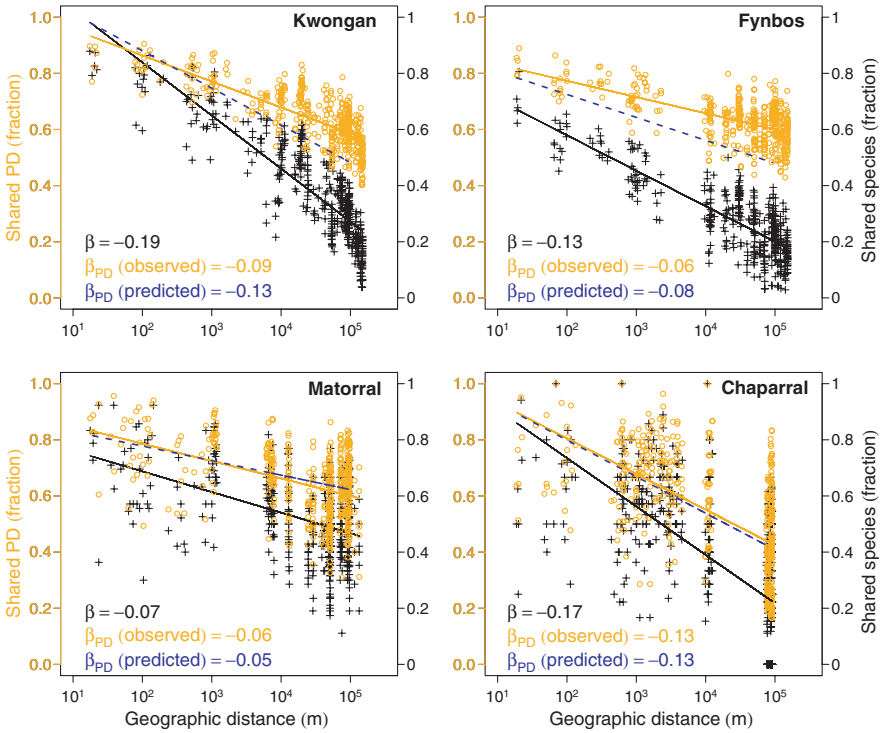


Figure 4 Decay in phylogenetic similarity with geographic distance within Mediterranean-type ecosystems. The observed phylogenetic distance–decay relationship (in orange: circles, data; line: logarithmic fit) can be approximated by expectations (eqn 6, in blue) obtained by simple transformation of the classical distance–decay relationship for species turnover (in black: crosses, data; line: logarithmic fit). The rate of decay in phylogenetic similarity (β_{PD}) is significant in all four datasets (mantel test, $P < 0.001$). This rate is lower than the rate of decay in taxonomic similarity (β), and the difference is the largest in floras where species are the least phylogenetically distinct (i.e. in the kwongan and fynbos).

mathematical form of spatial PD patterns. In this article, we provide theoretical predictions for the increase of PD with area and the decay in phylogenetic similarity with geographic distance under a model of random assembly from the regional species pool. These predictions have implications for conservation and for our understanding of how communities assemble.

In the future, conservation planners will likely leverage spatial models of PD to inform policy. The PD–area relationship, for example, can be used to estimate the potential loss of PD following habitat loss. Phylogenetically informed conservation research has primarily been focused on global-scale PD loss (Nee & May 1997), but the loss of PD at smaller spatial scales is of equal concern (e.g. Rodrigues & Gaston 2002; Forest *et al.* 2007; Winter *et al.* 2009; Devictor *et al.* 2010). For example, conservation strategies are often implemented at the level of geopolitical units interested in preserving regional evolutionary heritage and associated biological attributes of ethical, medical or economic value (Mooers & Atkins 2003; Purvis *et al.* 2005; Soutullo *et al.* 2005). Losing PD at any scale can lead to a reduced potential for communities to respond to changing environmental conditions, through a reduction of genetic diversity (Purvis *et al.* 2005).

Our derivation of the PD–area relationship shows that diversity depends on habitat area less strongly when measured as total phylogenetic branch-length vs. species richness. Although this may seem intuitive, a study by Rauch & Bar-Yam (2005), carried out in the context of population genetics, suggested the opposite pattern. This discrepancy is explained by the implicit assumption in Rauch and Bar-Yam's study that a genealogy remaining in a preserved area following habitat loss evolved solely in the preserved area. In contrast, our derivations acknowledge that a phylogeny observed after habitat loss is a sample of a phylogeny evolved in a larger area. Our derivations will thus provide more realistic estimates of PD loss with habitat loss.

Patterns of phylogenetic beta-diversity also have implications for conservation (Ferrier *et al.* 2007; Winter *et al.* 2009; Devictor *et al.* 2010). Communities share a greater fraction of PD than species (eqn 5). This suggests, as expected intuitively, that a single isolated area is more efficient in preserving PD than species richness. On the other hand, the phylogenetic similarity between communities decays with geographic distance at a slower pace than the similarity in species composition (eqns 5 and 6), such that larger distances between protected sites are needed to preserve PD relative to species diversity. In practice, as habitat degradation proceeds, conservation planners might have to choose between protecting distant but degraded sites vs. proximate but pristine ones. If degraded sites have lost their phylogenetic uniqueness, as can result from invasions (Winter *et al.* 2009), the beneficial effect of separating sites spatially needs to be compared with the beneficial effect of preserving the most unique species in pristine areas.

To make predictions about spatial PD patterns, we used species–PD curves. In our data, we found that species–PD curves were accurately modelled by power laws. This was not expected *a priori*: previous research predicted a logarithmic species–PD curve (equation 1 in Nee & May 1997). The logarithmic curve was not supported by our data, and there are multiple reasons to expect that it will not characterize empirical phylogenies. The logarithmic species–PD curve arises from Hey's model of cladogenesis, which is known to produce phylogenies with much shorter terminal branches than empirical phylogenies (Hey 1992). As terminal branches get longer than expected under Hey's model, species–PD curves become steeper than the logarithm and they tend toward a power-law function. Many phylogenies in nature have long terminal branches, as suggested by the preponderance of empirical phylogenies with negative values of the gamma statistic (negative gamma values reflect long terminal branches; Pybus & Harvey 2000). In addition, sampled phylogenies (e.g. continental or regional phylogenies) have fewer nodes towards the present than global-scale phylogenies, resulting in longer terminal branches (Pybus & Harvey 2000). Hence, the power-law approximation may be general to species–PD curves for a variety of taxonomic groups, sampled at a variety of spatial scales.

Our empirical evidence for power-law species–PD curves, rather than a logarithmic function, is relevant to seminal work linking species extinction and the loss of evolutionary history (Nee & May 1997; Heard & Mooers 2000). Nee & May (1997) suggested that PD is highly robust to random extinctions, based on the logarithmic shape of species–PD curves. This study has been criticized on the basis that extinctions are not random with respect to phylogeny (Heard & Mooers 2000; Purvis *et al.* 2000). However an even greater source of bias may come from the assumed shape for species–PD curves. The power-law shape observed in this study suggests that PD is not robust to extinctions, even under random loss. Intuitively, this increased loss of PD with extinction stems from the fact that species are much more evolutionarily distinct than expected under Hey's model.

In addition to assuming a power law species–PD curve, we assumed a random community assembly model. Within Mediterranean-type ecosystems, our data did not depart from this model. This absence of phylogenetic structure was likely a consequence of sampling in relatively homogeneous floras and environments, and at relatively small spatial scales. Deviations from the random assembly model are common in nature (Cavender-Bares *et al.* 2009; Vamosi *et al.* 2009) and have been reported in Mediterranean-type ecosystems (Proches *et al.* 2006; Forest *et al.* 2007).

A wide array of processes can lead to deviations from phylogenetic patterns predicted under the random assembly model. In turn, these deviations might offer insight into ecological and evolutionary processes. Within scales where species are not limited by their capacity to disperse, and under the hypothesis of trait conservatism, communities often switch from phylogenetic overdispersion at the smallest spatial scales (i.e. co-occurring species are distantly related) to phylogenetic clustering (i.e. co-occurring species are closely related) at larger spatial scales (Cavender-Bares *et al.* 2009; Kraft & Ackerly 2010). This happens, for example, when the competitive exclusion of closely related species, or the facilitation of distantly related ones, operates at smaller spatial scales than the filtering of closely related species by the environment. This scenario would increase PD values relative to the random assembly model at small scales, and decrease them at large scales, leading to a decrease of the slope of the observed PD–area curve compared with the null pattern. At spatial scales where dispersal limitation is a major driving force, evolutionary forces causing sister species to co-occur, such as *in situ* speciation, would result in a stronger signal of clustering compared with the null as spatial scale decreases. This situation would result in a steeper PD–area curve relative to the null.

Deviations from null phylogenetic beta-diversity patterns have been reported in the past, in particular for communities sampled along strong environmental gradients (Hardy & Senterre 2007; Bryant *et al.* 2008), or across sites separated by strong barriers to dispersal (e.g.

mountain ranges, oceans or large geographic distances; Forest *et al.* 2007; Chave *et al.* 2007; Graham *et al.* 2009). We observed deviations from the random assembly hypothesis when comparing communities across Mediterranean-type ecosystems, reflecting the presence of distinct floras in regions that have been geographically separated over evolutionary time scales. The strength of the deviation corresponded to the degree of historical isolation and geological differences between regions. More generally, deviations from the random decay in phylogenetic similarity with geographic distance are likely to happen if geographic distance is associated with strong barriers to dispersal, or if species traits are evolutionarily conserved and geographic distance is strongly associated with environmental distance. In these cases, the spatial turnover of lineages will be faster than expected from species turnover alone, steepening the slope of the decay in phylogenetic similarity with geographic distance compared with the null.

In conclusion, we used information on the spatial distribution of species and a random sampling of phylogenies to develop the first sampling theory for spatial patterns of PD. This framework offers the promise of using, in future research, well-studied macro-evolutionary models of cladogenesis to understand how phylogenies map on ecological communities and the landscape. This may ultimately improve our ability to conserve biodiversity.

ACKNOWLEDGEMENTS

The authors thank S. Kembel, J. Plotkin, J. Chave, C. Webb, J. O'Dwyer and E. Perry for discussions; Arne Mooers, David Ackerly and several anonymous referees for thoughtful comments on a previous version of the manuscript; A. Burns for help with data organization. J. Clines, E. D. Haaksma, N. Helme L. Husted, F. Salinas, R. Turner and staff at the Compton Herbarium, Kirstenbosch, for plant identifications; K. Thiele and the staff of Western Australian Herbarium (WA Department of Environment and Conservation) for access to collections and Florabase; C. Garin, J. Hinds, Y. Hussei, L. Husted, F. Salinas, I. Wright, botanists and volunteers in the four continents for help in the field and/or discussions. This project was supported by NSF grants DEB 0743885 awarded to J.L.G and B.J.M.B. and MCB 0500124 awarded to J.L.G.; P.A.M. acknowledges funding from FONDAP-1501-0001, ICM P05-002 and PFB-23.

REFERENCES

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008). Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc. Natl. Acad. Sci. USA*, 105, 11505–11511.

Cadotte, M.W., Davies, T.J., Regetz, J., Kembel, S.W., Cleland, E. & Oakley, T.H. (2010). Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol. Lett.*, 13, 96–105.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, 12, 693–715.

Chave, J. & Leigh, E.G. (2002). A spatially explicit neutral model of beta-diversity in tropical forests. *Theor. Popul. Biol.*, 62, 153–168.

Chave, J., Muller-Landau, H.C. & Levin, S. (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *Am. Nat.*, 159, 1–23.

Chave, J., Chust, G. & Thébaud, C. (2007). The importance of phylogenetic structure in biodiversity studies. In: *Scaling Biodiversity* (eds Storch, D., Marquet, P. & Brown, J.H.). Institute Editions, Santa Fe, pp. 151–167.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B. *et al.* (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669.

Crozier, R.H. (1997). Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Ann. Rev. Ecol. Syst.*, 28, 243–268.

Crozier, R.H., Dunnett, L.J. & Agapow, P.M. (2005). Phylogenetic biodiversity assessment based on systematic nomenclature. *Evol. Bioinform.*, 1, 11–36.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W. & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol. Lett.*, 13, 1030–1040.

Diniz-Filho, J.A. (2004). Phylogenetic autocorrelation analysis of extinction risks and the loss of evolutionary history in Felidae (Carnivora: Mammalia). *Evol. Ecol.*, 18, 273–282.

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, 61, 1–10.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.*, 13, 252–264.

Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P. *et al.* (2007). Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, 445, 757–760.

Graham, C.H. & Fine, P.V.A. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecol. Lett.*, 11, 1265–1277.

Graham, C.H., Parra, J.L., Rahbek, C. & McGuire, J.A. (2009). Phylogenetic structure in tropical hummingbird communities. *Proc. Natl. Acad. Sci. USA*, 106, 19673–19678.

Guilhaumon, F., Gimenez, O., Gaston, K.J. & Mouillot, D. (2008). Taxonomic and regional uncertainty in species–area relationships and the identification of richness hotspots. *Proc. Natl. Acad. Sci. USA*, 105, 15458.

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

Heard, S.B. & Cox, G.H. (2007). The shapes of phylogenetic trees of clades, faunas, and local assemblages: exploring spatial pattern in differential diversification. *Am. Nat.*, 169, 107–118.

Heard, S.B. & Mooers, A.O. (2000). Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proc. R. Soc. Lond. B*, 267, 613–620.

Hey, J. (1992). Using phylogenetic trees to study speciation and extinction. *Evolution*, 46, 627–640.

Jabot, F. & Chave, J. (2009). Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.*, 12, 239–248.

Kembel, S.W., Cowan, P., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D. *et al.* (2010). Picante: tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.

Kraft, N.J.B. & Ackerly, D.D. (2010). Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. *Ecol. Monogr.*, 80, 401–422.

Lozupone, C. & Knight, R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol. Rev.*, 32, 557–578.

MacArthur, R.H. & Wilson, E.O. (2001). *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Mooers, A. & Atkins, R.A. (2003). Indonesia's threatened birds: over 500 million years of evolutionary heritage at risk. *Anim. Conserv.*, 6, 183–188.

Morlon, H., Chuyong, G., Condit, R., Hubbell, S.P., Kenfack, D., Thomas, D. *et al.* (2008). A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.*, 11, 904–917.

Morlon, H., Potts, M.D. & Plotkin, J.B. (2010). Inferring the dynamics of diversification: a coalescent approach. *PLoS Biol.*, 8, e1000493.

Nee, S. (2006). Birth–death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.*, 37, 1–17.

Nee, S. & May, R.M. (1997). Extinction and the loss of evolutionary history. *Science*, 278, 692–694.

Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *J. Biogeogr.*, 26, 867–878.

Pimm, S. & Askins, R. (1995). Forest losses predict bird extinctions in eastern North America. *Proc. Natl. Acad. Sci. USA*, 92, 9343–9347.

Preston, F.W. (1962). The canonical distribution of commonness and rarity. *Ecology*, 43, 185–432.
Proches, S., Wilson, J.R.U. & Cowling, R.M. (2006). How much evolutionary history in a 10 × 10 m plot? *Proc. R. Soc. Lond. B*, 273, 1143–1148.
Purvis, A., Agapow, P.M., Gittleman, J.L. & Mace, G.M. (2000). Nonrandom extinction and the loss of evolutionary history. *Science*, 288, 328–330.
Purvis, A., Gittleman, J.L. & Brooks, T.M. (2005). *Phylogeny and Conservation*. Cambridge University Press, Cambridge.
Pybus, O.G. & Harvey, P.H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond. B*, 267, 2267–2272.
Rauch, E.M. & Bar-Yam, Y. (2005). Estimating the total genetic diversity of a spatial field population from a sample and implications of its dependence on habitat area. *Proc. Natl. Acad. Sci. USA*, 102, 9826–9829.
Richardson, J.E., Weitz, F.M., Fay, M.F., Cronk, Q.C.B., Linder, H.P., Reeves, G. *et al.* (2001). Rapid and recent origin of species richness in the cape flora of South Africa. *Nature*, 412, 181–183.
Rodrigues, A.S.L. & Gaston, K.J. (2002). Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol. Conserv.*, 105, 103–111.
Rosenzweig, M.L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
Soutullo, A., Dodsworth, S., Heard, S.B. & Mooers, A. (2005). Distribution and correlates of carnivore phylogenetic diversity across the Americas. *Anim. Conserv.*, 8, 249–258.
Vamosi, S.M., Heard, S.B., Vamosi, J.C. & Webb, C.O. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.*, 18, 572–592.
Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.*, 55, 235–254.
Webb, C.O. & Donoghue, M.J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes*, 5, 181–183.
Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.
Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24, 2098–2100.
Wikström, N., Savolainen, V. & Chase, M.W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B*, 268, 2211–2220.
Winter, M., Schweigera, O., Klotza, S., Nentwigc, W., Andriopoulosd, P., Arianoutsoud, M. *et al.* (2009). Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proc. Natl. Acad. Sci. USA*, 106, 21721–21725.

SUPPORTING INFORMATION

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

Appendix S1 Mediterranean flora data and phylogeny.
Appendix S2 Random community assembly.
Appendix S3 Species–PD relationship of the combined phylogeny and sensitivity analysis.
Appendix S4 Statistical tests relevant to spatial phylogenetic diversity patterns and predictions.
Appendix S5 Potential loss of PD with habitat loss in Mediterranean-type ecosystems.
Appendix S6 A general relationship between the species–PD curve, the species–area curve and the PD–area curve.
Appendix S7 The decay of phylogenetic similarity with geographic distance.
Appendix S8 Specific phylogenetic resolutions.

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

Editor, Arne Mooers
Manuscript received 1 February 2010
First decision made 4 March 2010
Second decision made 15 June 2010
Third decision made 1 October 2010
Fourth decision made 29 October 2010
Manuscript accepted 3 November 2010

Spatial patterns of phylogenetic diversity

H. Morlon, D.W. Schwilk, J.A. Bryant, P.A. Marquet,
A.G. Rebelo,C. Tauss, B.J.M. Bohannan, J.L. Green

This document comprises the following items:

- **Appendix S1:** Mediterranean flora data and phylogeny
- **Appendix S2:** Random community assembly
- **Appendix S3:** Species-PD relationship of the combined phylogeny and sensitivity analysis
- **Appendix S4:** Statistical tests relevant to spatial phylogenetic diversity patterns and predictions
- **Appendix S5:** Potential loss of PD with habitat loss in Mediterranean-type ecosystems
- **Appendix S6:** A general relationship between the species-PD curve, the species-area curve, and the PD-area curve
- **Appendix S7:** The decay of phylogenetic similarity with geographic distance
- **Appendix S8:** Specific phylogenetic resolutions

Appendix S1: Mediterranean flora data and phylogeny

Data Presence/absence data for woody angiosperms in the mediterranean climate zone of Australia, California, Chile and South-Africa were recorded between April and December 2006 (Fig. S1). On each continent, thirty nested quadrats were sampled at the 2.5 x 2.5 m, 7.5 x 7.5 m and 20 x 20 m scales (120 quadrats total). Sampled quadrats were laid out along transects ranging between $(30°42'S, 115°31'E)$ and $(29°16'S, 115°06'E)$ in Australia, $(36°26'N, 118°44'W)$ and $(37°06'N, 119°25'W)$ in California, $(34°22'S, 71°18'W)$ and $(33°05'S, 71°09'W)$ in Chile, and $(33°55'S, 19°11'E)$ and $(32°27'S, 18°53'E)$ in South-Africa. Quadrats were separated by geographic distances ranging from 20 m (adjacent) to 170 km. Within each quadrat, presence/absence data were recorded at the 2.5 x 2.5 m, 7.5 x 7.5 m and 20 x 20 m scales (nested sampling). Data were recorded only at the 20 x 20 m scale in California. A Google Earth File comprising all our sampling sites is available in the online Supplementary Information.

All woody angiosperms were collected, with no size cut-off. Specimens were identified by expert botanists in each region. Sub-species were lumped, resulting in a total of 538 species encompassing 254 genera and 71 families. In Australia, species were identified with reference to specimens held by the WA Herbarium and Florabase (the online database of the Western Australia Herbarium, http://florabase.calm.wa.gov.au/). In California, we used the Jepson manual (Jepson, 1993). In Chile, we used the Flora Silvestre de Chile (Hoffman, 2005). In South-Africa, species were identified with reference to specimens held by the Compton Herbarium (http://posa.sanbi.org/searchspp.php); records were checked against the latest synonyms in the National Herbarium Pretoria Computerised Information System (PRECIS). Species from the Restionaceae and Bromeliaceae are not woody; nonetheless, several genera from these families include species which fill an ecological sub-shrub niche as persistent, shrubby perennials. Therefore, puya species (Bromeliaceae) were included in Chile. Due to the ambiguity in categorizing species from the Restionaceae, these species were collected by the South-African field

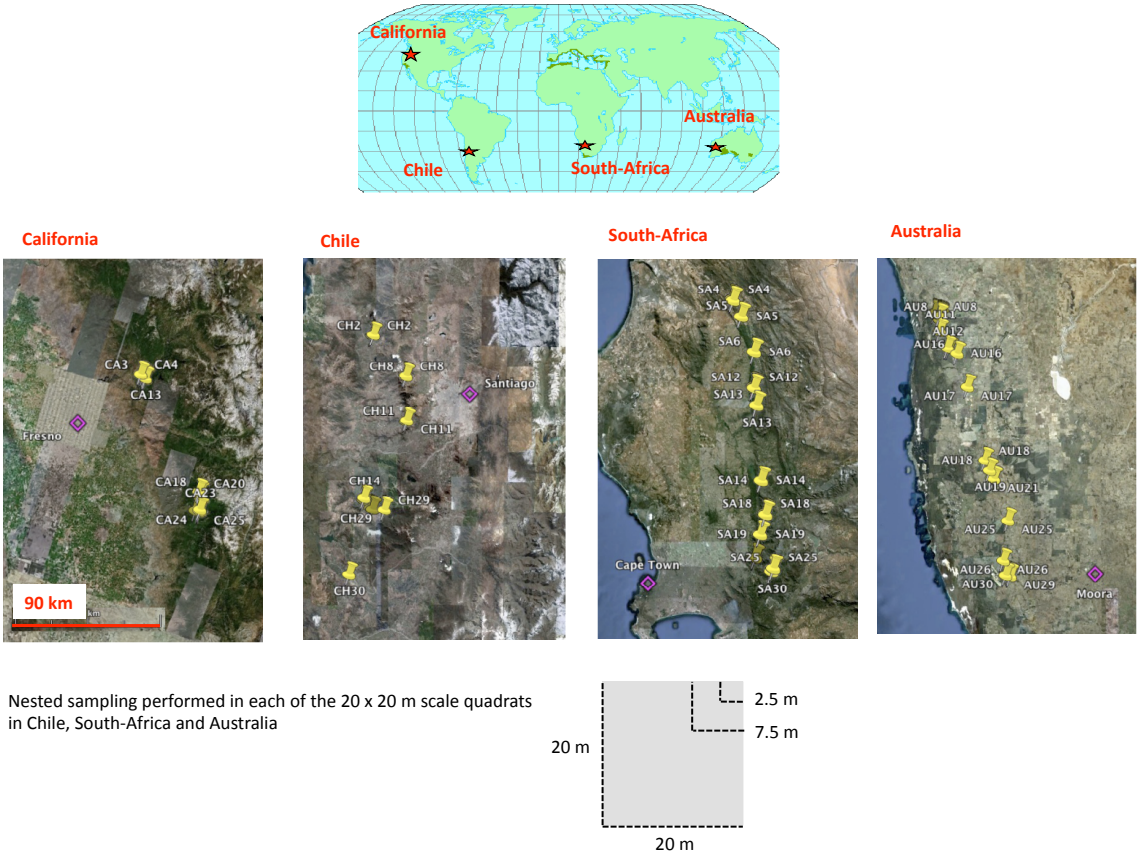
crew, but not by the Australian crew. Hence, the analyses in the paper include species from the Restionaceae in South-Africa, but not in Australia.

Phylogeny The phylogeny of the 538 species collected was constructed as specified in the main text. Thereafter, we term the phylogeny of all 538 species the "combined phylogeny", and we term the phylogenies of the species present in each dataset the "regional phylogenies". Phylogenetic data added to (or differing from) data given by the Phylomatic2 repository as of March 2010 are provided at the end of this document. A visual representation of the combined and regional phylogenies is shown in Fig. S2.

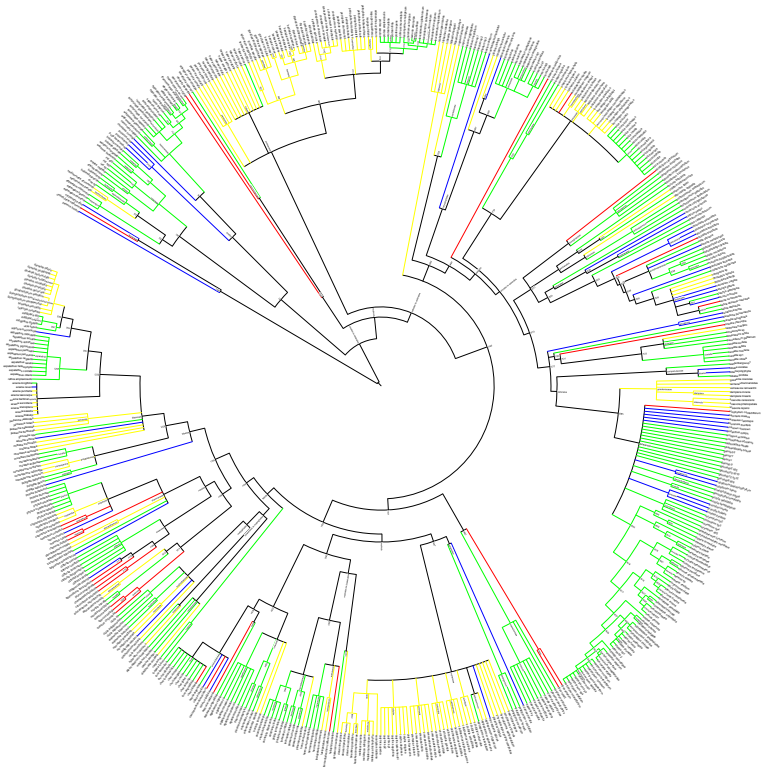
Appendix S2: Random community assembly

In Australia, Chile and South-Africa, we tested for potential deviations from the random assembly hypothesis in all 30 samples (90 samples total) at the 2.5 x 2.5 m, 7.5 x 7.5 m, and 20 x 20 m scales. In California, we only tested for deviations from the random assembly hypothesis at the scale where data was available (i.e. the 20 x 20 m scale). Following Webb *et al.* (2002), we ranked the PD observed in a sample containing S species within the PD of 1000 communities assembled by randomly sampling S species in each regional phylogeny. The significance of the deviation from the random assembly model was then obtained by dividing the rank of the observed PD by the number of observations (1001). A relative rank lower than 0.05 indicates that communities are significantly less phylogenetically diverse than expected by chance given their species richness (clustering). A relative rank greater than 0.95 indicates that communities are significantly more phylogenetically diverse than expected by chance given their species richness (overdispersion). With this level of significance, only few communities deviated significantly from the random assembly hypothesis (Fig. S3). There was a tendency for clustering in communities from the kwongan at the 7.5 m and 20 m scales, and a tendency for overdispersion in

Supplementary Figure 1: Overview of the location and spread of sampling sites. From left to right: mediterranean climate zone of Australia, California, Chile and South-Africa. Below: illustration of the nested sampling performed in each quadrat and each Mediterranean-type region except California.



Supplementary Figure 2: Combined phylogeny (i.e phylogeny of all 538 species combined). In yellow: species collected in the kwongan (Australia). In red: species collected in the chaparral (California). In blue: species collected in the matorral (Chile). In green: species collected in the fynbos (South-Africa). Phylogeny plotted using iTOL (http://itol.embl.de/index.shtml).

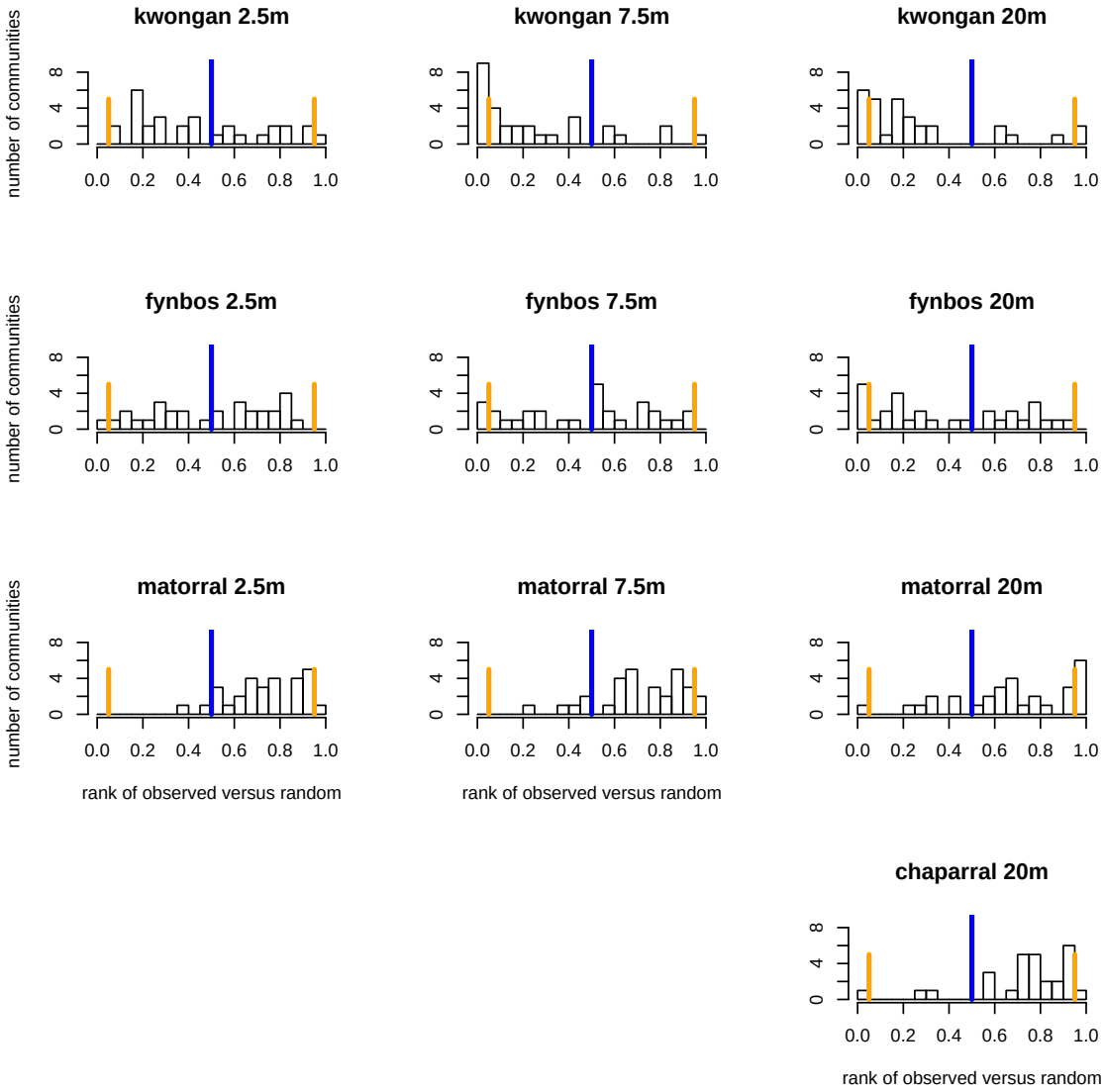


communities from the matorral and chaparral. These tendencies did not cause major deviations of the observed PD-area relationship from that expected under the random assembly hypothesis (Figure 3 from the main text, see also Appendix S4). Results obtained using other phylogenetic diversity metrics, namely the mean pairwise distance (MPD), which measures the mean phylogenetic distance among all pairs of species in the community and the mean nearest neighbor distance (MNND), which measures the mean phylogenetic distance to the nearest relative for all species in the community (Webb *et al.*, 2002), were qualitatively similar (results not shown).

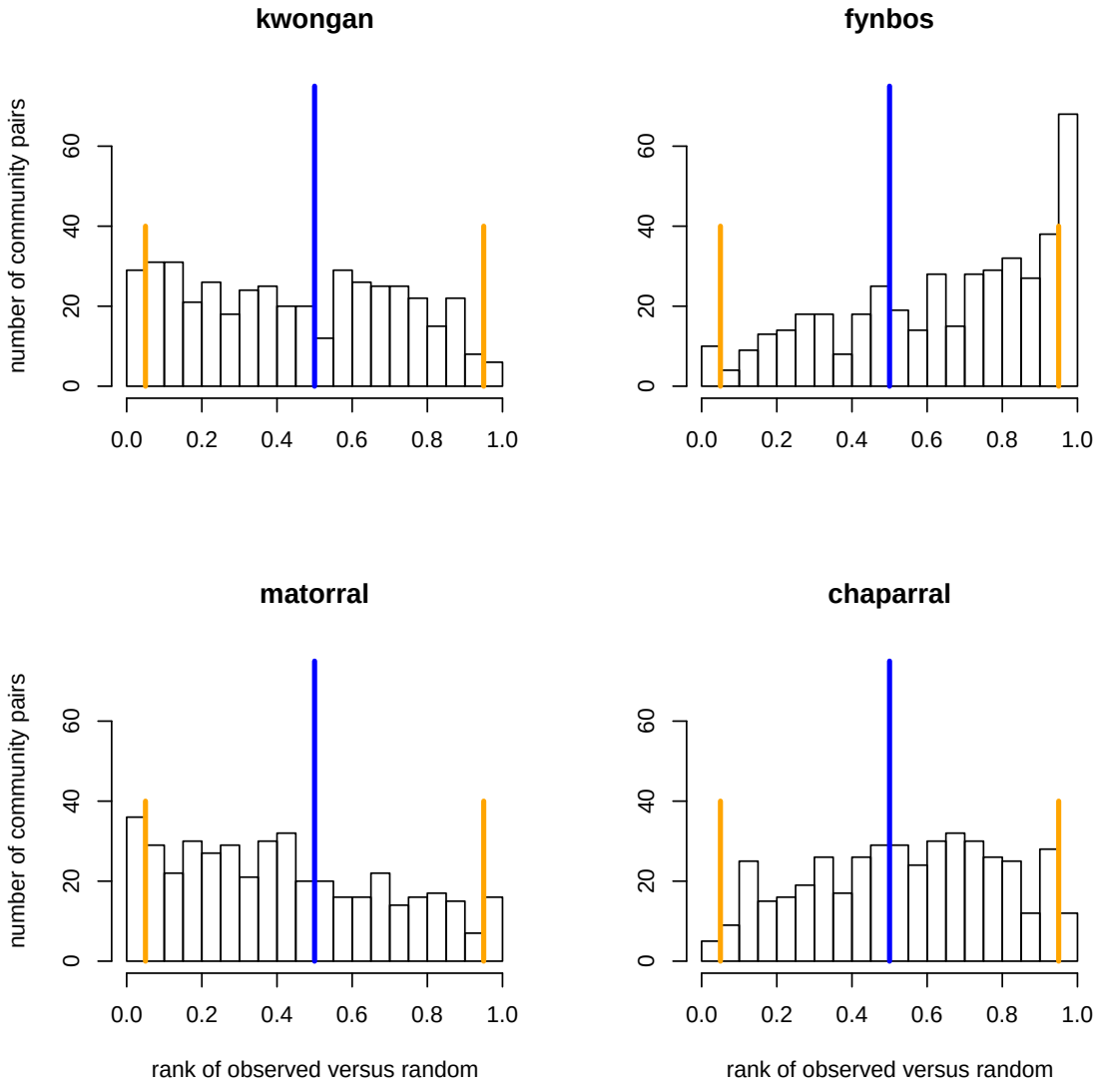
Using a similar approach, we tested for potential deviations from the random assembly hypothesis across pairs of samples (within region), at the sample size used in the paper (i.e. the 20 x 20 m scale). We ranked the χ_{PD} observed between two samples, one containing S_1 species, the other containing S_2 species, and the two sharing $S_{1\cap2}$ species within the χ_{PD} of 1000 community pairs assembled by randomly sampling S_1 and S_2 species in each regional phylogeny while keeping $S_{1\cap2}$ constant. The significance of the deviation from the random assembly model was then obtained by dividing the rank of the observed χ_{PD} by the number of observations (1001). A relative rank lower than 0.05 indicates communities that are significantly less phylogenetically similar than expected by chance given the number of species present within each, and shared between, the two communities. A relative rank greater than 0.95 indicate communities that are significantly more phylogenetically similar than expected by chance given the number of species present within each, and shared between, the two communities. With this level of significance, only a few communities deviated significantly from the random assembly hypothesis, except in the fynbos, where there was a marked tendency for pairs of communities to be more similar than expected by chance (Fig. S4). This tendency caused the observed decay in phylogenetic similarity to lie above (i.e. have greater similarity values) the one expected under the random assembly hypothesis (Figure 4 from the main text, see also Appendix S4).

Our result that the PD supported by communities, and shared across communities, is most

Supplementary Figure 3: Random assembly within communities in the Mediterranean data. Histograms report the number of communities falling in a given rank class (relative rank as defined above). Communities falling on the left of the blue line are less phylogenetically diverse (i.e. have a smaller PD) than expected by chance given their species richness, and significantly so when they fall on the left of the first orange line. Communities falling on the right of the blue line are more phylogenetically diverse (i.e. have a higher PD) than expected by chance given their species richness, and significantly so when they fall on the right of the second orange line. From left to right: data collected at the 2.5 x 2.5 m, 7.5 x 7.5 m, and 20 x 20 m scales.



Supplementary Figure 4: Random assembly across communities in the Mediterranean data. Histograms report the number of community pairs falling in a given rank class (relative rank as defined above). Communities falling on the left of the blue line are less phylogenetically similar than expected by chance given their species richness and turnover, and significantly so when they fall on the left of the first orange line. Communities falling on the right of the blue line are more phylogenetically similar than expected by chance given their species richness and turnover, and significantly so when they fall on the right of the second orange line.



often not significantly different from expected under the random assembly hypothesis is conservative. Applying a Bonferroni correction in order to account for multiple testing (per continent, we performed 30 tests within communities, and 435 tests across communities tests) would reduce the number of communities or pairs of communities deviating significantly from this hypothesis.

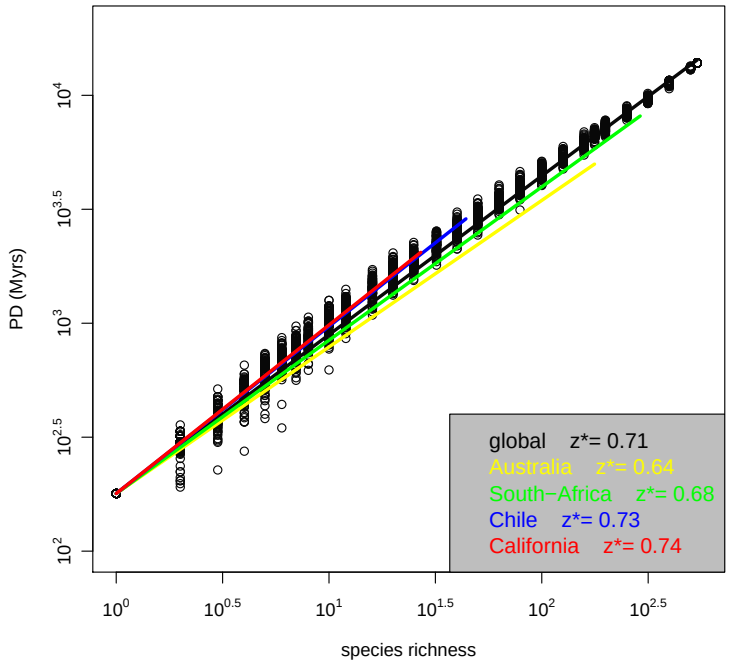
Appendix S3: Species-PD relationship of the combined phylogeny and sensitivity analysis

Species-PD relationship of the combined phylogeny Fig. S5 illustrates the species-PD curve of the phylogeny of the 538 species combined, and how it compares to regional species-PD curves. The figure shows that the scale invariant species-PD curve holds on the combined phylogeny, with a z* exponent similar to exponents observed in regional phylogenies.

Sensitivity analysis To test the robustness of the species-PD curves and related exponents to uncertainty in the phylogeny, we separately explored the effect of polytomies and of the node age assignment algorithm. Code for these analyses is available at www.schwilk.org/research/data.html.

Polytomies An analysis of the effect of a lack of resolution on measurements of phylogenetic diversity, performed on simulated phylogenies, has shown that phylogenetic diversity is particularly sensitive to a lack of resolution basally (Swenson, 2009). Here, we were interested in the effect of missing resolution in our specific data, and on the specific patterns investigated in the paper (in particular the species-PD curve). To explore the effect of polytomies, we conducted the following procedure for each phylogeny (the combined phylogeny and the four regional phylogenies): 1) we created a set of 1000 alternative versions of the full unpruned phylogeny (i.e. angiosperm backbone tree + 538 species) and then ran the modified BLADJ algorithm on each of these to assign branch-lengths by dating undated nodes. For each of these

Supplementary Figure 5: Species-PD relationship of the combined phylogeny (in black) and comparison with the species-PD curve of each regional phylogeny. Data points (black circles) are the results of 100 simulated random samplings across the tips of the combined phylogeny. Lines are power-law fits across the data (data points corresponding to regional phylogenies not shown for clarity). The species-PD curve of the combined phylogeny is well approximated by a power-law curve with an exponent similar to those observed in regional phylogenies.



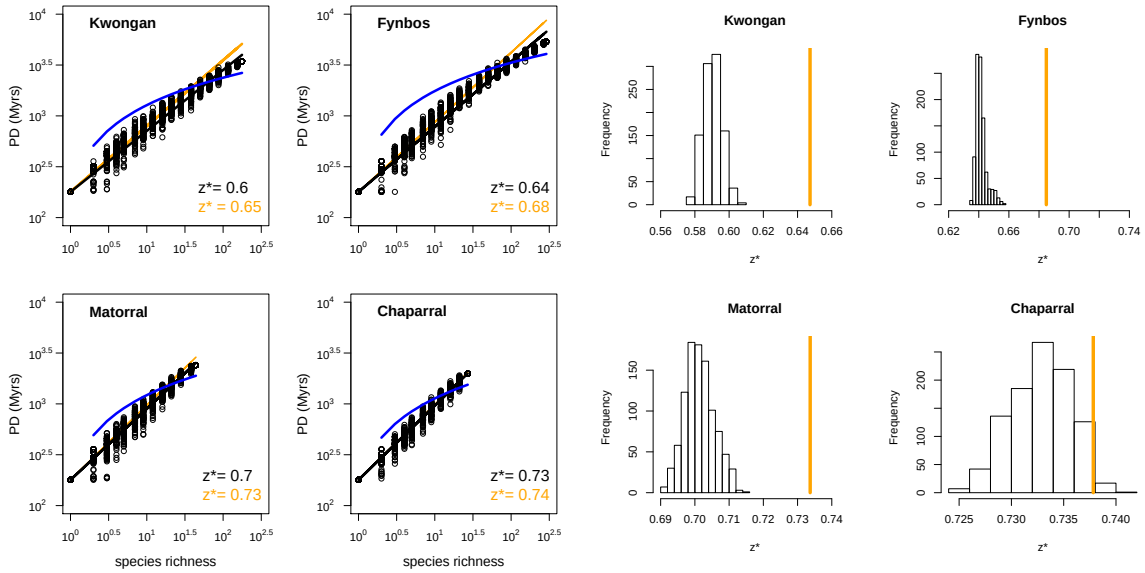
randomizations, we conducted a rarefaction analyses as described in the main text with 100 random draws for each species richness value.

The random resolution of polytomies followed by BLADJ branch-length assignment tended to curve the power-law species-PD curve downward towards the end of the sampling procedure (i.e. when almost all species were included), and consistently lowered z* values (Fig. S6). Randomly resolving polytomies pushed undated nodes towards the tips of the phylogenies. Deviations from the pattern observed without randomly resolving polytomies were the lowest in the Californian dataset where all nodes were resolved, and the largest in the Australian and South-African datasets where many polytomies remained. In the global phylogeny, the z* value was 0.71 without resolution, and the mean over 100 random resolutions was 0.65. In all phylogenies, the power-law approximation remained relevant after random resolution. In particular, the power-law always provided a better fit than the previously proposed logarithm (Nee & May, 1997) (black versus blue fit in Fig. S.6). Deviations from z* values obtained without random resolutions were always less than 0.1 unit. Hence, the presence of polytomies in the phylogenies does not compromise the main approach and conclusions of our study.

BLADJ branch-length assignment To test the sensitivity of the relationships to the BLADJ evenly-spaced node age method, we generalized the node dating algorithm to allow undated nodes to be assigned dates according to any normalized age distribution. We explored two variations.

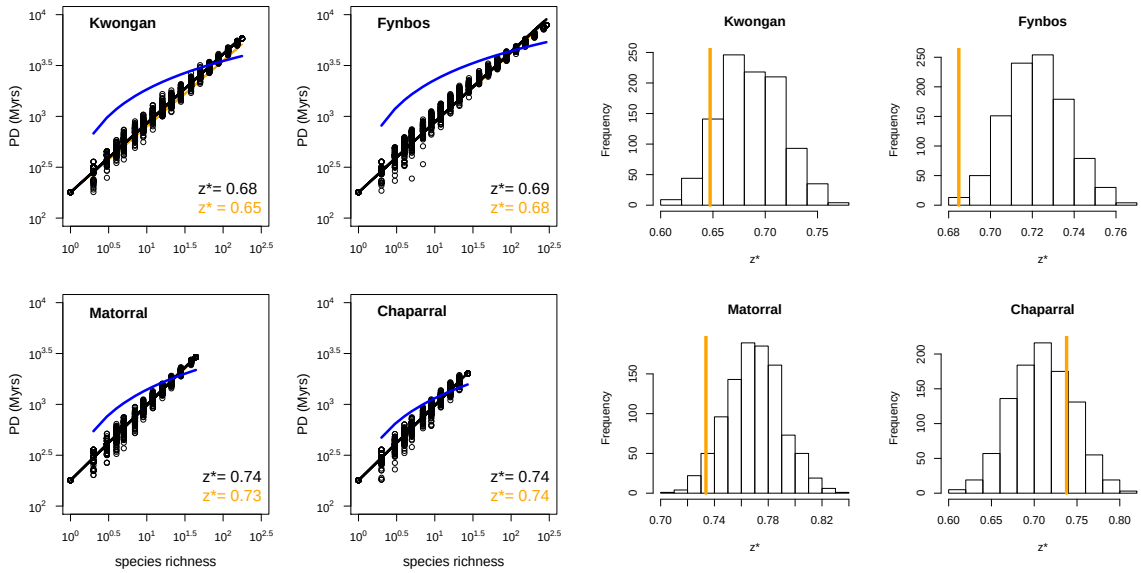
In the first branch-length sensitivity analysis, instead of assigning node ages deterministically and evenly, node ages were assigned from a uniform random distribution with bounds set by fixed ages of ancestors and descendants. Using this method, we explored a set of 1000 phylogenies that varied in branch-length assignment for each topology. As expected, using a

Supplementary Figure 6: Robustness of the power-law shape of species-PD curves, and sensitivity of z* values, to random resolutions of polytomies followed by BLADJ branch-length assignment. For each dataset, we constructed 1000 randomly resolved phylogenies. On the left: data points (black circles) are the results of 100 simulated random samplings across the tips of one (randomly chosen) of the 1000 randomly resolved phylogenies. The black line is the power-law fit across the data points. The orange line is the power-law fit corresponding to the original unresolved phylogeny (data points not shown for clarity). This line is barely visible in the matorral and chaparral, because randomly resolving polytomies in the corresponding phylogenies had very little effect on the species-PD curve. Note that deviations from the orange line in the kwongan and fynbos do not reflect deviations from the power-law (deviations from the black line would), but rather deviations from the species-PD curve obtained without randomly resolving the polytomies. The blue line is the best-fit logarithm, shown for comparison with previous literature (Nee & May, 1997). The power-law (black line) provides a much better fit than the logarithm (blue line). On the right: Distribution of z* values for the 1000 randomly resolved phylogenies. The red line indicates the z* value corresponding to the original unresolved phylogenies. Randomly resolving polytomies pushed undated nodes towards the tips of the phylogenies, tended to curve the power-law species-PD curve downward towards the end of the sampling procedure, and consistently lowered z* values.



uniform distribution of ages instead of an evenly-spaced distribution did not change the shape of curve; it increased the variance in z* values, but did not drastically change mean values (Fig. S7).

Supplementary Figure 7: Robustness of the power-law shape of species-PD curves, and sensitivity of z* values, to branch-length assignment using a random uniform distribution of node ages instead of evenly spacing nodes. For each dataset, we constructed 1000 phylogenies with undated nodes assigned ages from a uniform distribution. On the left: data points (black circles) are the results of 100 simulated random samplings across the tips of one (randomly chosen) of the 1000 random phylogenies. The black line is the power-law fit across the data points. The orange line, which represent the the power-law fit corresponding to the original phylogeny, can barely been seen due to the robustness of the species-PD curve to the method of branch-length assignment. The blue line is the best-fit logarithm, shown for comparison with previous literature(Nee & May, 1997). On the right: Distribution of z* values for the 1000 random node age phylogenies. The red line indicates the z* value corresponding to the original phylogenies. Uniformly distributing nodes does not significantly change the shape of the species-PD curve, and does not greatly influence mean z* values. Rather, this method increases the variance in z* values.



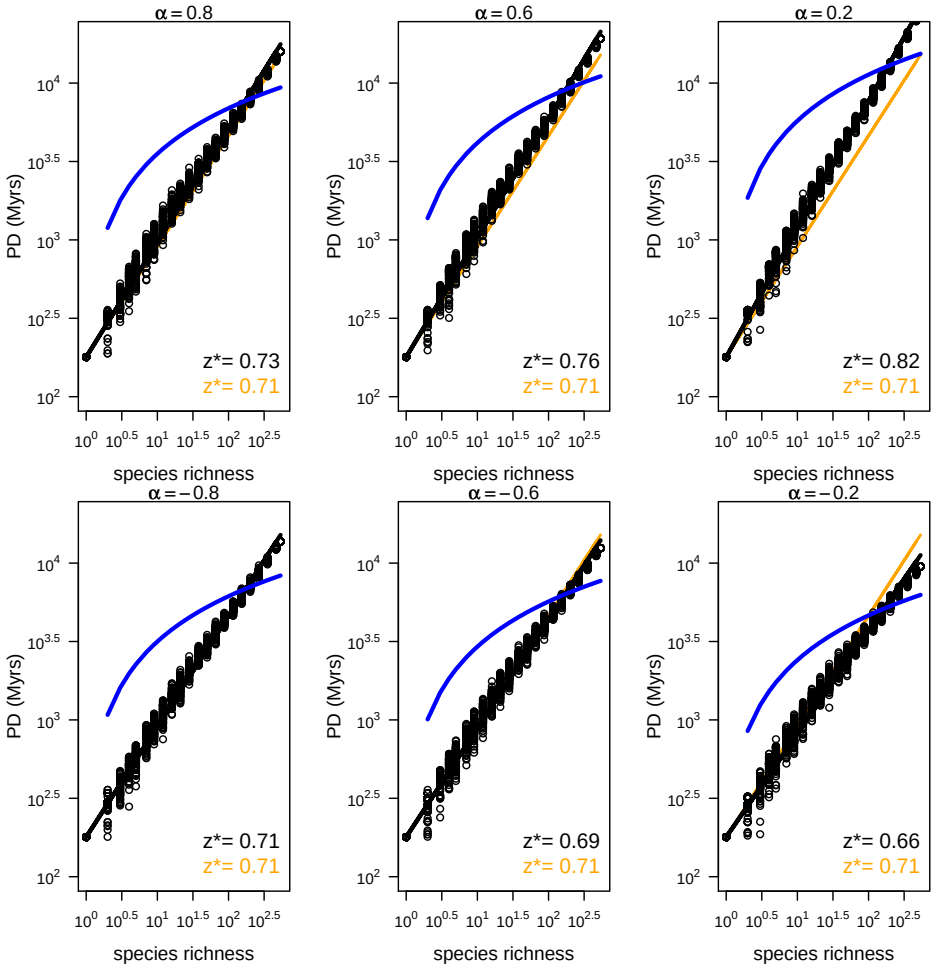
In the second branch-length sensitivity analysis, we explored node age distributions skewed towards either the root or the tips. For this analysis, node ages were drawn from a truncated

exponential distribution (again with bounds set by fixed ancestor and descendant ages). We defined a parameter, α, to control the amount of skewness. The truncated exponential parameter, λ, was then calculated from α: $\lambda = \frac{1}{\alpha}(A_a - A_d)$, where A_a and A_d are the ages of the fixed ancestor and descendant. Lower absolute values of alpha result in greater skew. We defined the algorithm such that λ can be relative to the ancestor age (positive alpha values, skewed toward root, resulting in longer branches at the tips) or the descendant age (negative alpha values, skewed toward tips, resulting in longer branches toward the root). We explored 1000 randomizations for each of six different truncated exponential distributions: three skewed towards the root with α values of 0.8, 0.6, and 0.2 and three skewed towards the tips with α values of -0.8, -0.6, and -0.2. We then ran the full rarefaction analysis on each of these phylogenies. The power-law shape of the curve was only affected for strong skews ($|\alpha|$=0.2) and more sensitive to a skew towards the tips than towards the root (Fig. S8). As expected, z* values decreased when nodes were distributed towards the tips, and decreased when nodes were distributed towards the root. Deviations from initial z* values were always less than 0.1 unit and did not compromise the main approach and conclusions of our study (Fig. S9).

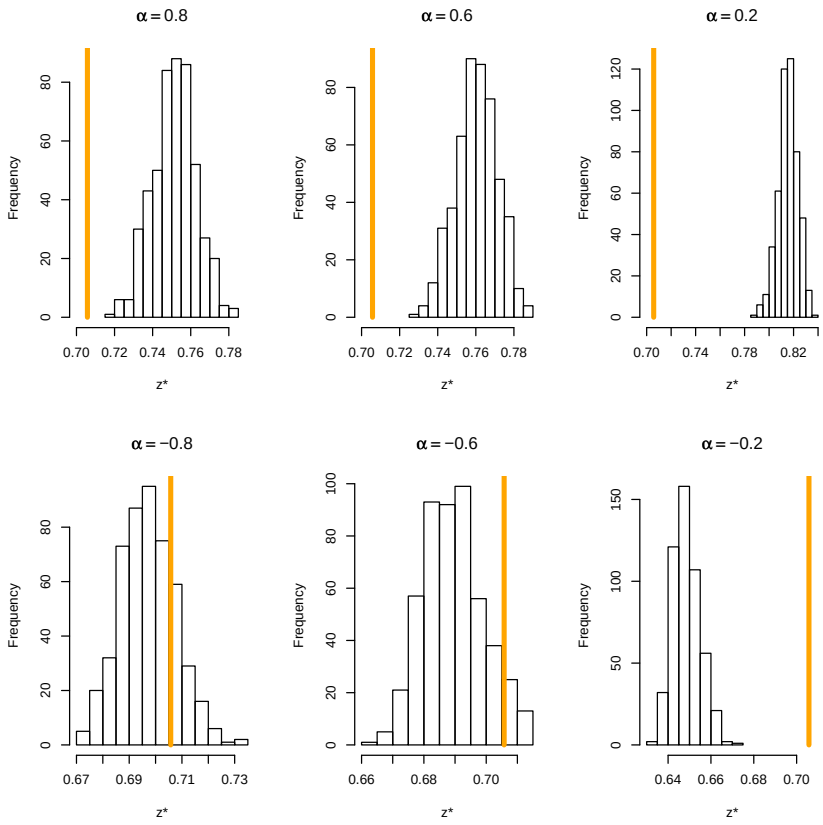
Appendix S4: Statistical tests relevant to spatial phylogenetic diversity patterns and predictions

This section describes the statistical tests that we used to: 1) test the ability of the random assembly process to reproduce observed spatial phylogenetic diversity patterns (thereafter referred to as Test 1), and 2) test the accuracy of our spatial phylogenetic diversity theory predictions (Test 2). Test 1 is different from the tests performed in Appendix S2. In Appendix S2, we tested for potential deviations from the random assembly model at the level of individual communities (Fig. S3) or pairs of communities (Fig. S4). Here, we test for potential deviations from the random assembly model across communities, to assess if the observed spatial patterns deviate from

Supplementary Figure 8: Robustness of the power-law shape of species-PD curves to branch-length assignment using a distribution of node ages skewed towards the root or the tips instead of evenly spacing nodes. α values determine the direction and strength of the skew. Top row: distribution skewed toward the root (i.e. undated nodes are pushed towards the root). Bottom row: distribution skewed towards the tips (i.e. undated nodes are pushed towards the tips). From left to right: increasing skew. For each α value, we constructed 1000 phylogenies in which node ages were drawn from a truncated exponential distribution. The power-law shape is robust. Results - shown here for the combined phylogeny - were similar for regional phylogenies.



Supplementary Figure 9: Sensitivity of z* values to branch-length assignment using a distribution of node ages skewed towards the root or the tips instead of evenly spacing nodes. α values determine the direction and strength of the skew. Top row: distribution skewed toward the root (i.e. undated nodes are pushed towards the root). Bottom row: distribution skewed towards the tips (i.e. undated nodes are pushed towards the tips). From left to right: increasing skew. For each α value, we constructed 1000 phylogenies in which node ages were drawn from a truncated exponential distribution. As expected, phylogenies with longer terminal branch-lengths (top) have higher z* values, and phylogenies with shorter terminal branch-lengths (bottom) have higher z* values. Results - shown here for the combined phylogeny - were similar for regional phylogenies.



those expected under random community assembly. To perform Test 1, we compared observed curves to the 95% confidence envelopes of the curves obtained by simulations of the random assembly process (see below). To perform Test 2, we compared the predicted curves to these same confidence envelopes.

Statistical tests related to PD-area curves

To construct the 95% confidence envelope of the PD-area curve in a given Mediterranean region, we performed the three following steps: 1) we ran 1000 simulations of the random assembly process (keeping species richness constant) across all 30 samples at all scales, 2) we constructed the PD-area curve corresponding to each simulation by averaging, at each spatial scale, PD values across the 30 samples, and 3) we excluded, at each spatial scale, the 5% most extreme values.

To test the ability of the random assembly process to reproduce the observed PD-area curve (Test 1), we compared this observed curve (obtained by averaging across the 30 samples at each scale) to the 95% confidence envelope of the PD-area curve. The observed PD-area curve (orange line in Fig. S10) fell within the 95% confidence envelope of the curve obtained under random assembly (black lines, Fig. S10). In agreement with the results found in Appendix S2 (Fig. S3), the observed PD tended to be lower than expected under random assembly in the kwongan (specially at the largest spatial scales) and fynbos (at all spatial scaes), and higher than expected under random assembly in the chaparral. None of these tendencies were significant.

To test the accuracy of the PD-area predictions (Test 2), we compared the predicted curves (Equations 2 and 4 from the main text) to the 95% confidence envelope. The predictions from Equation 2 do not make the assumption that the species-area curve is power-law, whereas the predictions from Equation 4 do. The predictions from Equation 4 (blue dashed lines, Fig. S 10) were in good agreement with the predictions from Equation 2 (blue circles, Fig. S10),

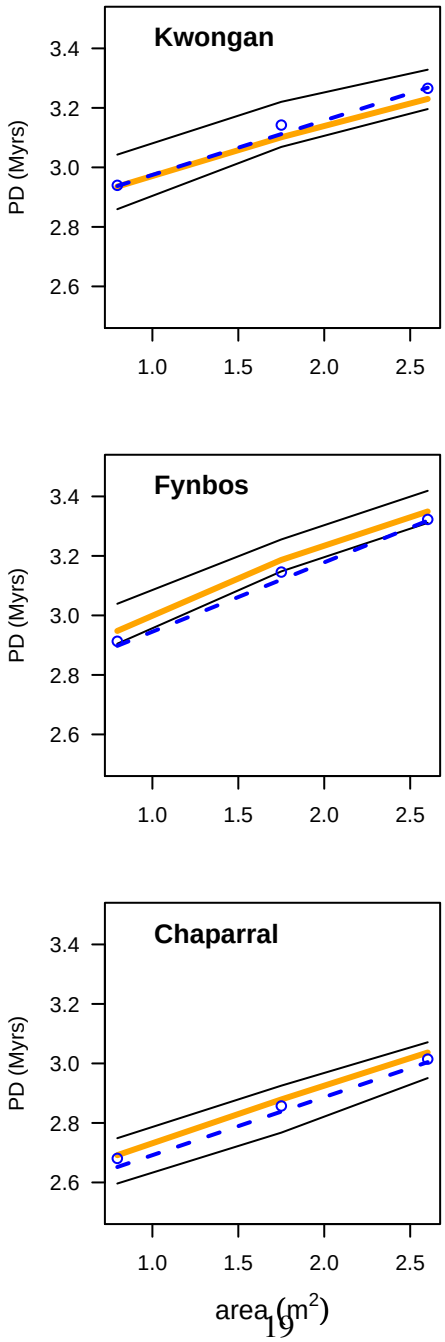
demonstrating the validity of using the power-law to describe the species-area curve in our data. The predicted curves (blue circles and blue dashed lines) fell within the 95% confidence envelope obtained under random assembly (black lines, Fig. S10) in the kwongan and chaparral. In the fynbos, the predicted curves were close to the lower bound of the 95% confidence interval. This deviation was due in part to the fact that PD values were slightly lower than expected by chance, and it was also due to deviations from the power-law assumption for the species-PD curve in the fynbos. Regardless of these deviations, the predicted curves yielded a reasonable quantitative description of the data.

Statistical tests related to the decay in phylogenetic similarity with geographic distance

To construct the 95% confidence envelope of the phylogenetic distance-decay curve, we performed the three following steps: 1) we ran 1000 simulations of the random assembly process (keeping species richness and turnover constant) across all 435 sample pairs at the 20 x 20 m scale 2) we constructed the phylogenetic distance-decay curve corresponding to each simulation by pulling data points falling in 0.2 distance bins (on a log scale) 3) we excluded, in each bin, the 5% most extreme values.

To test the ability of the random assembly process to reproduce the observed phylogenetic distance-decay curve (Test 1), we compared this observed curve (obtained by pulling data points falling in 0.2 distance bins, orange line in Fig. S11) to the 95% confidence envelope of the curve obtained under random assembly (black lines in Fig. S11). This comparison shows that the random assembly process tends to overestimate similarity values in the kwongan, and underestimate them in the fynbos, although this discrepancy is only significant at the smallest spatial separation in the kwongan. The fact that observed similarity values tend to be higher in the fynbos than expected under random assembly is in agreement with the results found in Appendix S2 (Fig. S4).

Supplementary Figure 10: Test of theory predictions for the PD-area relationship. The observed relationship (orange line) is in good agreement with relationships obtained under random assembly (95% confidence interval represented by black lines). The predictions from Equation 2 (blue circles) fail in the fynbos, mainly due to deviations from the species-PD curve power-law assumption. The predictions from Equation 4 (blue dashed lines) fail in the fynbos due to the previous failure of Equation 2. PD-area curves were obtained by averaging the data over the 30 samples at each spatial scale.

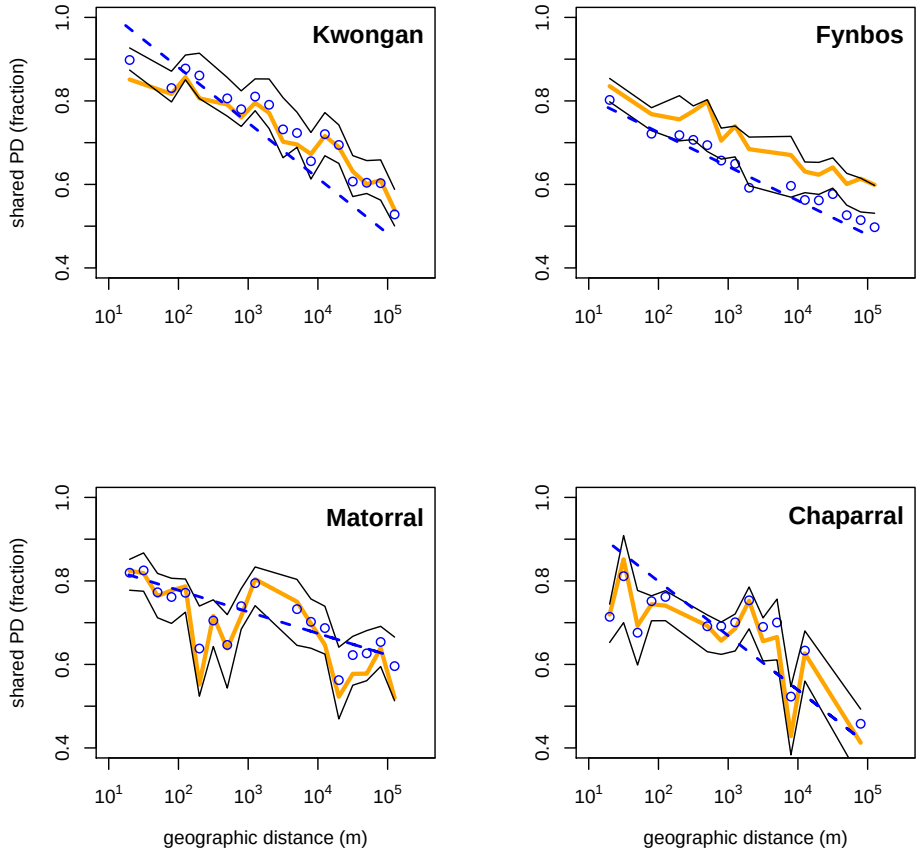


Deviations between the observed and predicted (Test 2, Equation 6 from the main text) phylogenetic distance-decay curves occurred in the kwongan and fynbos (Figure 4 from the main document). These deviations from the observed phylogenetic distance-decay curves have three potential origins: 1) deviations linked to the random assembly hypothesis 2) deviations linked to the power-law approximation for species-PD curves (resulting in Equation 5 from the main text) 3) deviations linked to fitting a logarithmic curve to the species distance-decay relationship (resulting in Equation 6 from the main text). The first source of error has been discussed in the previous paragraph, and explains part of the deviations in the fynbos. To evaluate the second source of error, we compared predictions given by Equation 5 (blue circles in Fig. S11) to the 95% confidence envelope of the curve obtained under random assembly (black lines). This comparison reveals that a good part of the deviations observed in the fynbos is linked to the power-law assumption for the species-PD curve, which results in a consistent underestimation of phylogenetic similarity values. To evaluate the third source of error, we compared predictions given by Equation 6 (blue dashed lines; this equation assumes a logarithmic fit to the species distance-decay relationship) to predictions given by Equation 5 (blue circles; this equation does not make any assumption on the shape of the species distance-decay relationship). This comparison reveals that a good part of the deviations observed in the kwongan is linked to the logarithmic assumption for the species distance-decay relationship. Regardless of deviations from the theory predictions, these predictions yielded a reasonable quantitative description of the data.

Appendix S5: Potential loss of PD with habitat loss in Mediterranean-type ecosystems

We investigated the potential loss of PD within each Mediterranean-type hotspot. This loss represents the loss of PD in the region under study: PD may be preserved elsewhere on Earth

Supplementary Figure 11: Test of theory predictions for the decay in phylogenetic similarity with geographic distance. The observed relationship (orange line) is in general in good agreement with relationships obtained under random assembly (95% confidence interval represented by black lines), except in the fynbos were communities tend to be more phylogenetically similar than expected under random assembly. The predictions from Equation 5 (blue circles) fail in the fynbos, due to deviations from the species-PD curve power-law assumption. The predictions from Equation 6 (blue dashed lines) fail in the fynbos due to the previous failure of Equation 5, and in the kwongan due to deviations from the logarithmic curve for the species distance-decay relationship. Distance-decay curves were obtained by pulling data points in 0.2 distance bins (on a log scale).



due to the presence of closely related species outside of the region considered. However, it is important to preserve PD at all spatial scales (see main text), and thus to investigate the potential regional loss of PD irrespective of what is preserved outside of the region.

By analogy with the classical species-area relationship (how species richness increases with area, Rosenzweig (1995)) which is commonly used to estimate the number of species threatened by habitat loss (Pimm *et al.*, 1995), we used the PD-area relationship to estimate the amount of PD threatened by habitat loss. Using the PD-area relationship to estimate PD loss presents several serious drawbacks in line with the drawbacks associated with using the species-area relationship to estimate species loss (Seabloom *et al.*, 2002). When full censuses, knowledge on species habitat, and information on habitat loss are available, more elaborate methods than species-area based methods exist for estimating diversity loss (Seabloom *et al.*, 2002; Faith, 2008). Area selection algorithms may also be used to select a set of protected sites maximizing the amount of PD preserved (Rodrigues & Gaston, 2002; Faith, 2006; Forest *et al.*, 2007). However, species-area based methods are useful when data on the systems to protect are incomplete, which explains that species-area curves are still used to derive estimates of species loss (e.g. Hubbell *et al.* (2008)).

In addition to limitations associated with using the species-area relationships, our estimates of PD-area slopes relied on small samples (i.e. the extrapolation is large), and they were based on data collected in one flora only, not the full hotspot. With these limitations in mind, we estimated how much floristic PD is potentially threatened in each Mediterranean-type hotspot as a result of habitat that has already been lost (i.e. due to extinctions that have already occurred or to an extinction debt), and how much is protected in current conservation areas. These calculations require an estimation of: 1) the scaling of PD with area, 2) the fraction of habitat lost relative to the total area of the hotspot, and 3) the fraction of habitat protected relative to the total area of the hotspot. The total area of the hotspot is defined as the sum of the biogeographic

region encompassing the current flora and the area that has been converted to human use.

To estimate the scaling between PD and area, we used the empirical data (Figure 3 in the main text). The slope of the power-law relationship between PD and area estimated in each continent was: 0.16 in Australia, 0.20 in Chile, and 0.23 in South-Africa. Due to a lack of nested data in California, we assumed the canonical value $z = 0.25$ for the power-law exponent of the species-area curve (Rosenzweig, 1995), and the empirical value $z^* = 0.74$ for the scaling of phylogenetic diversity with species richness.

To estimate the fraction of habitat lost and the fraction of habitat protected, we used recent estimates for the total, current and protected (defined as IUCN category I-IV) areas in each hotspot. For each Mediterranean-type region, these estimates were as follows (in the order: total, current, protected): Australia (D. Shepherd, personal communication updated from Beeston *et al.* (2006)): 297 928 km^2, 120 258 km^2 (40% of original), 37 844 km^2 (13% of original); California Mittermeier *et al.* (2005): 293 804 km^2, 73 451 km^2 (25% of original), 30 002 km^2 (10.2% of original); Chile (Wilson *et al.*, 2007; Mittermeier *et al.*, 2005): 148 383 km^2, 10 000 km^2 (7% of original), 1 332 km^2 (0.9% of original); South-Africa (Rouget *et al.*, 2006): 83 946 km^2, 57 923 km^2 (69% of original), 8 395 km^2 (10% of original). In the four regions combined, the estimated original, current and protected habitats span 824 061 km^2, 261 632 km^2 and 77 573 km^2, respectively.

Based on the slopes of the PD-area relationship and the fraction of area lost and protected, we estimated from Equation 4 (in the main text) the percentage of PD lost as a result of habitat loss in each Mediterranean-type region, yielding 14% in Australia, 23% in California, 42% in Chile, and 8.2% in South-Africa. The percentage of PD protected in conservation areas in each Mediterranean-type region was also predicted from Equation 4, yielding: 72% in Australia, 65% in California, 39% in Chile and 59% in South-Africa. To obtain similar estimates in the four regions combined, we used the average slope of the species-area relationships across con-

tinents (0.28), and the z* value for the combined species-PD curve ($z^* = 0.71$), resulting in a slope for the PD-area relationship of 0.20. If current habitat loss leads to extinctions predicted by the species-area relationship, then 20% of PD has been lost (or will likely be lost) in the Mediterranean-type ecosystems (excluding the Mediterranean Bassin), and 62% is protected in current conservation areas.

Appendix S6: A general relationship between the species-PD curve, the species-area curve, and the PD-area curve

In the main text, we used the power-law relationship to provide a simple characterization of the species-PD and species-area curves. The approach used, however, may be generalized to any functional characterization of the curves. Suppose that the species-PD curve follows any functional form f:

$$PD(S) = f(S) \tag{1}$$

Suppose that the species-area relationship follows any functional form g:

$$S(A) = g(A) \tag{2}$$

Under the hypothesis that communities are randomly assembled at each spatial scale, the expected PD contained in an area A ($PD(A)$) is the expected PD of the expected number of species contained in an area A. In other words:

$$PD(A) = PD(S(A)) = f(g(A)) = fog(A) \tag{3}$$

Hence, the functional form of the PD-area relationship is simply given by the composition of the species-PD curve with the species-area curve. Equation 3 may be used to predict, under the random assembly hypothesis, the shape of the PD-area relationship for any functional form of the species-PD and species-area relationships.

Appendix S7: The decay of phylogenetic similarity with geographic distance

The Sorensen index of similarity between two communities 1 and 2 is given by:

$$\chi = \frac{S_1 + S_2 - S_{1\cup 2}}{\frac{1}{2}(S_1 + S_2)} \tag{4}$$

where S_1 and S_2 represent species richness in community 1 and 2, respectively, and $S_{1\cup 2}$ represents species richness in community 1 and 2 combined. The expected similarity between two sampled communities spanning area A and separated by distance d is given by:

$$\chi(A,d) \sim \frac{2S(A) - S_{1\cup 2}(A,d)}{S(A)} = 2 - \frac{S_{1\cup 2}(A,d)}{S(A)} \tag{5}$$

A phylogenetic equivalent of the Sorensen index, measuring the fraction of phylogenetic branch-length shared between two communities 1 and 2, is given by (see Material and Methods in the main text):

$$\chi_{PD} = \frac{PD_1 + PD_2 - PD_{1\cup 2}}{\frac{1}{2}(PD_1 + PD_2)} \tag{6}$$

where PD_1 and PD_2 represent evolutionary history in community 1 and 2, respectively, and $PD_{1\cup 2}$ represents evolutionary history in community 1 and 2 combined. The expected phylogenetic similarity between two sampled communities spanning area A and separated by distance d is approximated by:

$$\chi_{PD}(A,d) \sim \frac{2PD(A) - PD_{1\cup 2}(A,d)}{PD(A)} = 2 - \frac{PD_{1\cup 2}(A,d)}{PD(A)} \tag{7}$$

Assuming that species are randomly assembled with respect to phylogeny and using the power-law scaling given by Equation 3 (from the main document) yields:

$$\chi_{PD}(A,d) \sim 2 - \Big(\frac{S_{1\cup 2}(A,d)}{S(A)}\Big)^{z^*} \tag{8}$$

Finally, combining Equations 5 and 8 yields Equation 5 in the main document. Note that Equation 8, and thus also Equation 5 in the main document, only holds when the area sampled are of the same size.

Under the logarithmic model $\chi(A,d) = \alpha + \beta \log_{10}(d)$, Equation 6 (in the main document) yields:

$$\chi_{PD}(A,d) \sim 2 - \Big(2 - \alpha + \beta \log_{10}(d)\Big)^{z^*} \tag{9}$$

$$\chi_{PD}(A,d) \sim 2 - (2-\alpha)^{z^*}\Big(1 + \frac{\beta}{2-\alpha}\log_{10}(d)\Big)^{z^*} \tag{10}$$

For $\frac{\beta}{2-\alpha}\log_{10}(d) << 1$, we obtain:

$$\chi_{PD}(A,d) \sim 2 - (2-\alpha)^{z^*} + \beta\frac{z^*}{(2-\alpha)^{1-z^*}}\log_{10}(d) \tag{11}$$

Appendix S8: Specific phylogenetic resolutions

Phylogenetic data added to, or differing from data given by the Phylomatic2 repository as of March 2010.

Asparagales Resolution as in apweb 2005, not apweb 2009, with Asphodelaceae sister to Xanthorrhoeaceae.

(orchidaceae,boryaceae,(blandfordiaceae,(lanariaceae,(asteliaceae,hypoxidaceae))),((ixioliriaceae, tecophilaeaceae),(doryanthaceae,(iridaceae,(xeronemataceae,((hemerocallidaceae,(xanthorrhoeaceae, asphodelaceae)),((alliaceae,(amaryllidaceae,agapanthaceae)),((hesperocallidaceae,aphyllanthaceae, (hyacinthaceae,themidaceae),agavaceae),(laxmanniaceae,(asparagaceae,ruscaceae)))))))));

Malvales Resolution as in apweb 2005, not apweb 2009, since apweb 2005 includes within family resolution for malvaceae, thymelaeaceae and dipterocarpaceae whereas apweb 2009 does not.

(neuradaceae,((gonystylus,(aquilaria,(daphne,(phaleria,dirca))))thymelaeaceae,(sphaerosepalaceae, ((bixaceae,diegodendraceae),cochlospermaceae),(cistaceae,(sarcolaenaceae,((dipterocarpus, (dryobalanops,(hopea,shorea,parashorea))),(vatica,cotylelobium),(upuna,vateria),anisoptera)dipterocarpaceae)), muntingiaceae,((((grewia,luehea),apeiba),(kleinhovia,byttneria)),((neesia,durio),(pentace,(heretiera,sterculia, scaphium),(tilia,pterospermum),ceiba)))malvaceae)));

Asteraceae Resolution as specified in Forest *et al.* (2007).

(corymbium, (chrysanthemoides, ((oedera, (((anaxeton, syncarpha), helichrysum), ((elytropappus, stoebe), metalasia))), (euryops, (othonna, senecio))), ((berkheya, cullumia), heterolepis)));

Fabaceae Resolution supplemented by resolutions specified in Forest *et al.* (2007).

((bauhinia, cercis)cercideae, (((((berlinia, brachystegia, oddoniodendron), brownea, cynometra, amherstia), ((hymenaea, guibourtia, peltogyne), tessmannia)), (barnebydendron, goniorrhachis), schotia, (colophospermum, prioria))detarieae, (((((dialium, martiodendron), petalostylis), apuleia), poeppigia)dialiinae, (((arcoa, ceratonia, gymnocladus, gleditsia)umtiza_clade, diptychandra, (((chamaecrista, cassia, senna)cassiinae, (((hoffmannseggia, zuccagnia), (caesalpinia, cenostigma, pomaria, poincianella, guilandia, stuhlmannia, haematoxylum, erythrostemon))caesalpinia_group, pterogyne)pterogyne_group), tachigali, ((conzattia, parkinsonia, peltophorum)core_peltophorum_group, ((mora, dimorphandra, erythrophleum)dimorphandra_group, (dinizia, pentaclethra, mimozyganthus, ((amblygonocarpus, adenanthera, tetrapleura, xylia, pseudoprosopis, calpocalyx)adenanthera_group,(piptadeniastrum, (entada, (plathymenia, ((neptunia, prosopis, prosopidastrum)prosopis_group, (desmanthus, leucaena)leucaenae_group, (dichrostachys, gagnebina)dichrostachys_group, (parkia, (microlobius, parapiptadenia, stryphnodendron, anadenanthera, pseudopiptadenia, adenopodia, piptadenia, mimosa)piptadenia_group, (acacia, ((faidherbia, zapoteca), lysiloma, enterolobium, albizia, ((chloroleucon, leucochloron,

blanchetiodendron)chloroleucon_alliance, (abarema, pararchidendron)abarema_alliance, (samanea, pseudosamanea)samanea_alliance, (havardia, ebenopsis, pithecellobium)pithecellobium_alliance, (calliandra, cojoba, zygia, macrosamanea, cedrelinga, archidendron, inga)inga_alliance)) ingeae))))))))mimosoids)))), ((((bobgunnia, bocoa, candolleodendron, swartzia), ((ateleia, cyathostegia), trischidium))swartzieae, ((((myrocarpus, myroxylon, myrospermum), dussia), amburana), ((dipteryx, pterodon), taralea)), angylocalyx, (styphnolobium, pickeringia, cladrastis), (((uribea, calia), (holocalyx, lecointea, zollernia)), (luetzelburgia, sweetia), (((((((((arachis, stylosanthes), chapmannia), fiebrigiella), (brya, cranocarpus), platymiscium, grazielodendron, (cascaronia, geoffroea), (centrolobium, ramorinoa, inocarpus, tipuana, maraniona, pterocarpus, platypodium)), (discolobium, riedeliella)), (aeschynomene, machaerium, dalbergia, kotschya, (diphysa, (ormocarpopsis, ormocarpum), zygocarpum, pictetia), weberbauerella)),(adesmia, ((chaetocalyx, nissolia), (poiretia, (amicia, zornia)))))dalbergioids, (((marina, dalea), psorothamnus), (apoplanesia, ((amorpha, parryella), (eysenhardtia, errazurizia)))))amorpheae, (andira, hymenolobium), (vatairea, vataireopsis), ((((((((((genista, ulex), spartium), cytisus), (lupinus, anarthrophyllum)), dichilus), (crotalaria, lebeckia)), calpurnia), (piptanthus, (baptisia, thermopsis))),(((((templetonia, hovea), lamprolobium), (plagiocarpus, brongniartia)), harpalyce), cyclolobium, poecilanthe)), (bolusanthus, dicraeopetalum), (sophora, ammodendron, maackia), (diplotropis,(bowdichia, acosmium)), clathrotropis), ormosia)genistoids, (baphia, (((((aotus, gastrolobium), isotropis), gompholobium, daviesia, bossiaea)mirbelieae, hypocalyptus), ((((((((((macroptilium, mysanthus), (strophostyles, dolichopsis)), (oxyrhynchus, ramirezella), phaseolus, vigna, physostigma), (dipogon, lablab)), (spathionema, vatovaea), (((dolichos, nesphostylis), macrotyloma), sphenostylis), wajira), (cologania, (pseudovigna, neorautanenia), pueraria, amphicarpaea, (ophrestia, glycine), ((otholobium, psoralea), (rupertia, psoralidium, pediomelum)))), (erythrina, psophocarpus)), butea), ((campylotropis, desmodium), apios))phaseoloids, (abrus, (galactia, (philenoptera, (piscidia, ((lonchocarpus, dahlstedtia, deguelia, be-

haimia, bergeronia), (apurimacia, mundulea, tephrosia), (derris, paraderris), neodunnia, brachypterum, (millettia, pongamiopsis)))))), fordia, austrosteenisia, dalbergiella, xeroderris, platycyamus)millettioids, (phylloxylon, (indigofera, (cyamopsis, microcharis)))indigofereae), ((((anthyllis, hammatolobium, lotus, ornithopus), (coronilla, hippocrepis))loteae, (((((genistidium, peteria), coursetia), olneya, poissonia, sphinctospermum, robinia), (poitea, gliricidia)), (hebestigma, lennea))robinieae, (sesbania, (glottidium, daubentonia))sesbanieae), (wisteria, callerya, glycyrrhiza, ((((caragana, halimodendron), (alhagi, (hedysarum, onobrychis))), (oxytropis, astragalus, sphaerophysa, colutea, (lessertia, sutherlandia), (swainsona, (carmichaelia, clianthus)))), (parochetus, (galega, (cicer, ((((melilotus, trigonella), medicago), ononis),(trifolium, (vicia, (lens, (pisum, lathyrus))))))))))irlc)))))), (liparia, (rafnia, aspalathus)))))));

Iridaceae Resolution as specified in Forest *et al.* (2007).

((watsonia, bobartia), aristea, irid);

Lamiaceae Resolution as specified in Forest *et al.* (2007).

((stachys, teucrium), salvia);

Poaceae Elegia placed in the Restionaceae instead of the Poaceae, based on Forest *et al.* (2007).

(flagellaria, baloskion, (joinvillea, ((anomochloa, streptochaeta), (pharus, ((guaduella, puelia), ((((streptogyna, (ehrharta, (oryza, leersia))), ((pseudosasa, chusquea), (buergersiochloa, ((lithachne, olyra), (eremitis, pariana))))), (brachyelytrum, ((lygeum, nardus), ((melica, glyceria), (((diarrhena, (brachypodium, (avena, (bromus, triticum)))), ((phaenosperma, anisopogon), (ampelodesmos, (piptatherum, (stipa, nassella)))))))))bep, (micraira, (((chasmanthium, (thysanolaena, zeugites)), (gynerium, (danthoniopsis, ((miscanthus, zea), (panicum, pennisetum))))), (eriachne, (((aristida, stipagrostis), (merxmuelleraa, (danthonia, (karoochloa, austrodanthonia

))))), (((molinia, phragmites), (amphiopogon, arundo)), ((merxmuellerab, centropodia), ((pappophorum, (eragrostis, uniola)), (distichlis, (zoysia, (spartina, sporobolus))))))))))pacc))))));

Proteaceae Resolution as specified in Forest *et al.* (2007)

(bellendena, (placospermum, toronia), ((agastachys, symphionema), (eidothea, cenarrhenes, (stirlingia, (conospermum, synaphea)), franklandia, (aulax, petrophile), beauprea, (isopogon, (adenanthos, (leucadendron, protea)))), (carnarvonia, sphalmium, knightia, triunia, (neorites, orites), (helicia, hollandaea), lomatia, stenocarpus, (telopea, alloxylon, embothrium), (opisthiolepis, (buckinghamia, grevillea)), (banksia, (austromuellera, musgravea)), roupala, (lambertia, xylomelum), (macadamia, (brabejum, panopsis)), (cardwellia, (euplassa, gevuina)))));

Restionaceae Resolution as specified in Forest *et al.* (2007)

(willdenowia, (ischyrolepis, ((hypodiscus, elegia, thamnochortus), (staberoha, restio))));

Rutaceace Resolution as specified in Forest *et al.* (2007)

(((acronychia, flindersia), zanthoxlum), ((murraya, poncirus), ruta), agathosma, (adenandra, diosma));

Scrophulariaceae Resolution as specified in Forest *et al.* (2007)

(((pseudoselago, selago), microdon), oftia);

Thymelaeaceae Resolution as specified in Forest *et al.* (2007)

(gonystylus, (aquilaria, (daphne, (phaleria, dirca))), (gnidia, struthiola), (passerina, lachnaea));

Quercus Resolution as specified in Manos *et al.* (1999)

((quercus_kellogii, quercus_wislizeni), quercus_chrysolepis);

References

Beeston, G. R., Hopkins, A. J. M. & Shepherd, D. P. (2006). Land-use and vegetation in Western Australia. Tech. Rep. 249, Department of Agriculture, Western Australia.

Faith, D. P. (2006). *Actions for the 2010 biodiversity target in Europe - How does research contribute to halting biodiversity loss?* Oxford University Press, Oxford, pp. 70–71.

Faith, D. P. (2008). *Conservation biology: evolution in action*. pp. 99–115.

Forest, F., Grenyer, R., Rouget, M., Davies, T. J., Cowling, R. M., Faith, D. P., Balmford, A., Manning, J. C., Proches, S., van der Bank, M. *et al.* (2007). Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, 445, 757–760.

Hoffman, A., J. (2005). *Flora Silvestre de Chile, Zona Central: Una guia para la identificacin de las especies vegetales ms frecuentes*. Cuarta Edicion, Princeton.

Hubbell, S. P., He, F., Condit, R., Borda-de Agua, L., Kellner, J. & ter Steege, H. (2008). How many tree species are there in the Amazon and how many of them will go extinct? *Proc. Natl. Acad. Sci.*, 105, 11498–11504.

Jepson (1993). *The Jepson manual: higher plants of California*. University of California Press, Berkeley.

Manos, P. S., Doyle, J. J. & Nixon, K. C. (1999). Phylogeny, biogeography, and processes of molecular differentiation in Quercus subgenus Quercus (Fagaceae). *Mol. Phyl. Evol.*, 12, 333–349.

Mittermeier, R. A., Gil, P. R., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J. & Da Fonseca, G. A. B. (2005). *Hotspots revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. The University of Chicago Press, Chicago.

Nee, S. & May, R. M. (1997). Extinction and the loss of evolutionary history. *Science*, 278, 692–692.

Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. (1995). The future of biodiversity. *Science*, 269, 347–350.

Rodrigues, A. & Gaston, K. (2002). Maximising phylogenetic diversity in the selection of networks of conservation areas. *Biol. Conserv.*, 105, 103–111.

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press, Cambridge.

Rouget, M., Jonas, Z., Cowling, R. M., Desmet, P. G., Driver, A., Mohamed, B., Mucina, L., Rutherford, M. C. & Powrie, L. W. (2006). *The Vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity Institute Pretoria, Strelitzia.

Seabloom, E. W., Dobson, A. P. & Stoms, D. M. (2002). Extinction rates under nonrandom patterns of habitat loss. *Proc. Natl. Acad. Sci.*, 99, 11229–11234.

Swenson, N. (2009). Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLoS ONE*, 4, e4390.

Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annu. Rev. Ecol. Syst.*, 475–505.

Wilson, K. A., Underwood, E. C., Morrison, S. A., Klausmeyer, K. R., Murdoch, W. W., Reyers, B., Wardell-Johnson, G., Marquet, P. A., Rundel, P. W., McBride, M. F. *et al.* (2007). Conserving biodiversity efficiently: what to do, where, and when. *PLoS. Biol.*, 5, 1850–1861.