Oecologia (2012) 170:501–506
DOI 10.1007/s00442-012-2318-8

COMMUNITY ECOLOGY - ORIGINAL RESEARCH

Computing diversity from dated phylogenies and taxonomic hierarchies: does it make a difference to the conclusions?

Carlo Ricotta · Giovanni Bacaro · Michela Marignani · Sandrine Godefroid · Stefano Mazzoleni

Received: 2 March 2011 / Accepted: 22 March 2012 / Published online: 17 April 2012

Abstract Recently, dated phylogenies have been increasingly used for ecological studies on community structure and conservation planning. There is, however, a major impediment to a systematic application of phylogenetic methods in ecology: reliable phylogenies with time-calibrated branch lengths are lacking for a large number of taxonomic groups and this condition is likely to continue for a long time. A solution for this problem consists in using undated phylogenies or taxonomic hierarchies as proxies for dated phylogenies. Nonetheless, little is known on the potential loss of information of these approaches compared to studies using dated phylogenies with time-calibrated branch lengths. The aim of this study is to ask how the use of undated phylogenies and taxonomic hierarchies biases a very simple measure of diversity, the mean pairwise phylogenetic distance between community species, compared to the diversity of dated phylogenies derived from the freely available software *Phylomatic*. This is illustrated with three sets of data on plant species sampled at different scales. Our results show that: (1) surprisingly, the diversity computed from dated phylogenies derived from *Phylomatic* is more strongly related to the diversity computed from taxonomic hierarchies than to the diversity computed from undated phylogenies, while (2) less surprisingly, the strength of this relationship increases if we consider only angiosperm species.

Keywords Dated phylogenies · Linnaean taxonomy · Phylogenetic distance · *Phylomatic* · Species diversity

Communicated by Melinda Smith.

C. Ricotta · M. Marignani
Department of Environmental Biology, University of Rome 'La Sapienza', Rome, Italy

G. Bacaro
BIOCONNET, Biodiversity and Conservation Network, Department of Environmental Sciences 'G Sarfatti', University of Siena, Siena, Italy

M. Marignani (✉)
Department of Life and Environmental Sciences, University of Cagliari, Viale S.Ignazio da Laconi 13, 09123 Cagliari, Italy
e-mail: marignani@unica.it

S. Godefroid
National Botanic Garden of Belgium, Meise, Belgium

S. Godefroid
Laboratory of Plant Biology and Nature Management (APNA), Vrije Universiteit Brussel, Brussels, Belgium

S. Godefroid
Laboratory of Systems Ecology and Resource Management, Université libre de Bruxelles, Brussels, Belgium

S. Mazzoleni
Department of Horticulture, Botany and Plant Pathology, University of Naples 'Federico II', Portici, Italy

> Counting the hair in the back of beetles became a dull activity, cool taxonomy being involved in counting stripes in electrophoretic probes, or triplets of CAGT in sequencing experiments.
>
> F. Boero (2010)

Introduction

In the last decade, community ecologists and conservation biologists have become increasingly interested in calculating the phylogenetic diversity of species assemblages (Rodrigues and Gaston 2002; Strauss et al. 2006; Forest

et al. 2007; Cavender-Bares et al. 2009; Ricotta et al. 2009; Kraft and Ackerly 2010). In conservation biology, phylogenetic diversity is used to capture the evolutionary history represented by one or more assemblages, and is thus a recommended criterion for the selection of biological reserves (Crozier 1997; Mace et al. 2003). In community ecology, phylogenetic diversity is relevant to classic perspectives on community assembly in which fundamental rules imposed by local environmental filters tend to constrain species traits within certain limits, and the principle of competitive exclusion prevents coexisting species from being too similar functionally (Cavender-Bares et al. 2009; Thompson et al. 2010). Therefore, given the strong link between phylogeny and variation in functional traits, phylogenetic diversity represents an important proxy for the functional diversity of a species assemblage. For instance, if functional traits exhibit phylogenetic signal, then the phylogenetic dispersion of a species assemblage should reflect its functional dispersion (Swenson and Enquist 2009).

However, while time-calibrated phylogenies provide a meaningful metric of evolutionary relatedness among species, their general applicability is limited by the small proportion of taxa for which such phylogenies have been constructed (Crozier et al. 2005). Given that the lack of reliable dated phylogenies for many taxonomic groups is likely to continue, some researchers use instead just the branching topology of a phylogeny, estimating species relatedness by using the number of nodes between species on some undated phylogenetic tree (Webb 2000; Gerhold et al. 2008). At the same time, Crozier et al. (2005) proposed using classical Linnaean taxonomic hierarchy as a reasonable surrogate for phylogeny. According to Crozier et al. (2005), "systematists generally try to make the arrangement of species into taxa mirror the topology of an inferred evolutionary tree, and the various classificatory levels similarly reflect the systematist's judgment as to the degree of difference". Therefore, whenever dated phylogenies are missing, taxonomic hierarchy can be applied in community ecology as a reasonable surrogate for phylogenetic trees.

Although taxonomic hierarchies and undated phylogenies have allowed for evolutionary analyses of species assemblages with little phylogenetic information, it remains unclear how much information is lost in these studies compared to those using a dated phylogenetic tree. The aim of this paper is thus to focus on the issue of calculating phylogenetic diversity from undated phylogenies and taxonomic hierarchies as surrogates for dated phylogenies derived from the software *Phylomatic* (Webb and Donoghue 2005). Phylogenetic trees derived from *Phylomatic* are not the fully-resolved and precisely-calibrated trees that ideally should be used for this analysis. Nonetheless, under current circumstances, *Phylomatic* is the best possible tool for the construction of (time calibrated) phylogenetic relationships among the plants in large species assemblages. Accordingly, in this paper, we will address the following questions: (1) if *Phylomatic* trees are used as the default source for phylogenetic information, how similar are values of phylogenetic diversity obtained using two different measures of phylogenetic distance, (a) branch lengths based on fossil-dated nodes versus (b) node counts (branch topology) in lieu of branch lengths? (2) If a taxonomic hierarchy is used to calculate phylogenetic diversity using taxonomic ranks as a measure of branch length, how closely does this approximate the phylogenetic diversity calculated from a *Phylomatic* tree with branch lengths based on fossil-dated nodes?

Materials and methods

In this study, we first quantify the phylogenetic diversity of species assemblages using dated phylogenies derived from the freely available software *Phylomatic*. We then compare these values with those obtained for the same communities using either the branching topology of the same phylogeny or the corresponding Linnaean taxonomy. To summarize the strength of correlation between the phylogenetic diversity values generated using dated phylogenies and those generated using undated phylogenies or taxonomic hierarchies, we used three sets of data on plant species sampled at different scales: (1) floras of 86 urban green areas in Brussels (Belgium) sampled from 1992 to 1994 (490 species; Godefroid 2001); (2) data from Mediterranean forests of Monte Circeo, Central Italy, sampled in 2002 (98 circular plots of 3 m^2, 156 species; Marignani, unpublished); and (3) data from the Natura 2000 Network of the Siena Province, Italy (215 square plots of 100 m^2, 770 species; Chiarucci et al. 2008a, b). The Brussels data included samples taken in urban forest patches (12 sites), wetlands (5), parks (31), wastelands (19), railway stations (2) and cemeteries (17). For incorporating phylogeny into measures of biological diversity, rather than adopting the rigorous standpoint of a systematic biologist, we assumed the perspective of a community ecologist without deep knowledge of plant systematics, thus using available phylogenies depicting 'established' relationships among the major lineages within the seed plant tree. For every set of species, we constructed a phylogenetic tree using the *Phylomatic* software (http:\\www.phylodiversity.net/phylomatic). *Phylomatic* uses the base tree of the Angiosperm Phylogeny Group (APG) at APweb (http:\\www.mobot.org/MOBOT/research/APweb) as the backbone in combination with recently published family phylogenies to form

its reference tree. All monophyletic families in APG III (APG 2009) are included in the reference tree. Branch lengths were then assigned to the phylogenetic tree based on dated nodes (Myr) reported by Wikstrom et al. (2001) from fossil data. Nodes in the phylogenetic tree for which age estimates were available were fixed, while all remaining nodes were spaced evenly between dated nodes to minimize variance in branch length. The major shortcoming of *Phylomatic* is that the output phylogenetic tree contains many polytomies below the family level. On the other hand, *Phylomatic* is virtually the sole freely available tool for the construction of phylogenetic relationships among plant species. Accordingly, due to its simplicity, we consider it an appropriate operational tool for assembling into a phylogenetic tree all taxa in large species assemblages.

We computed the phylogenetic diversity of each plot in the simplest manner possible, namely as the mean pairwise phylogenetic distance between species in the assemblages (see Webb 2000; Ricotta et al. 2008a, b). For an ultrametric tree, like the dated phylogeny produced by *Phylomatic*, this will be twice the time since divergence from the most recent common ancestor (branch length from species 1 to the most recent common ancestor plus branch length from the most recent common ancestor to species 2); for undated phylogenies and taxonomic dendrograms, this distance is the number of nodes connecting species pairs. Specifically, for the Linnaean taxonomic hierarchy, if two species belong to the same genus, their distance is 2 (there are two nodes separating both species); if two species belong to different genera but the same family, their distance is 4, etc. Unlike alternative metrics of phylogenetic diversity as, e.g., the Faith Index (1992), our measure is mathematically independent of species number, and is thus adequate for comparing plots of different richness.

First, we calculated the phylogenetic diversity of all plots using dated branch lengths. Next, using simple linear regressions, we compared those diversities with those calculated using (1) the sole branching topology of the phylogeny, or (2) the classical Linnaean taxonomic hierarchy. For constructing the Linnaean taxonomic tree, we used the following hierarchies: species, genus, family, order, subclass, class, subphylum, and phylum according to the APG III nomenclature (APG 2009). While datasets A and B are composed exclusively of angiosperm species, for dataset C, the same exercise is repeated using angiosperms only (760 species). All diversity analyses were run with the programs *Phylocom* (Webb et al. 2008) and *TreeCreeper*, freely available for download at: http:\\www.phylodiversity.net/phylocom and http:\\www.ecoap.unina.it/doc/publications.htm (or from the direct link http:\\www.ecoap.unina.it/download/TreeCreeper.zip), respectively.

Results and discussion

Measures of diversity based on phylogenetic trees or on surrogates, like taxonomic hierarchies or undated phylogenies, have been extensively used for analyzing community structure (Warwick and Clarke 1994, 2001; Webb et al. 2002; Cavender-Bares et al. 2009; Ricotta et al. 2010) and for prioritizing natural areas for conservation actions (Barker 2002; Rodrigues and Gaston 2002; Forest et al. 2007). However, in many cases, these studies face two main problems (Crozier et al. 2005; Swenson 2009): the phylogenies being used contain pervasive polytomies, while lacking information on branch length. While the effect of polytomies on the calculation of phylogenetic diversity was extensively analyzed by Swenson (2009), the present study is aimed at focusing on how much information is lost in using undated phylogenies and taxonomic hierarchies as proxies for available dated phylogenies in the calculation of a simple measure of biological diversity.

The linear regressions between the phylogenetic diversity values calculated using time-calibrated branch lengths and the diversity values calculated from the undated phylogenies and from the taxonomic hierarchies are shown in Fig. 1. For all datasets, the diversity values calculated from the dated phylogenies are strongly related to the diversity values calculated from the taxonomic hierarchy, whereas their relationship with the diversity values calculated from the undated phylogenies is weak to very weak. Also, for dataset C, both relationships are much stronger considering the angiosperm species only (Fig. 1g, h).

The weak correlation between the dated and undated phylogenies is possibly due to the fact that the structure of the phylogenetic tree changes as a function of the species included. Therefore, the number of species within a clade will influence its branching topology, and hence the degree of species relatedness, while the topological distance between two species in the taxonomic hierarchy is not affected by the number of species. This can lead to the paradox that, in the undated phylogeny, two congeneric species within a species-rich genus like *Carex*, *Trifolium*, or *Vicia* are likely to appear as distant or more distant than two species that belong to the same family, but to different genera (Webb 2000).

By contrast, in the Linnaean hierarchy, the ranks are nested within one another and, therefore, the more inclusive ranks should be generally older than less inclusive ranks. That is, orders are generally older than families, while families are generally older than genera, thus providing a rough representation of evolutionary relatedness among species. However, dealing with taxonomic hierarchies, we are assuming some kind of evolutionary equivalence between taxa at the same hierarchical level, such that a family in the *Fabales* and a family in the *Poales*

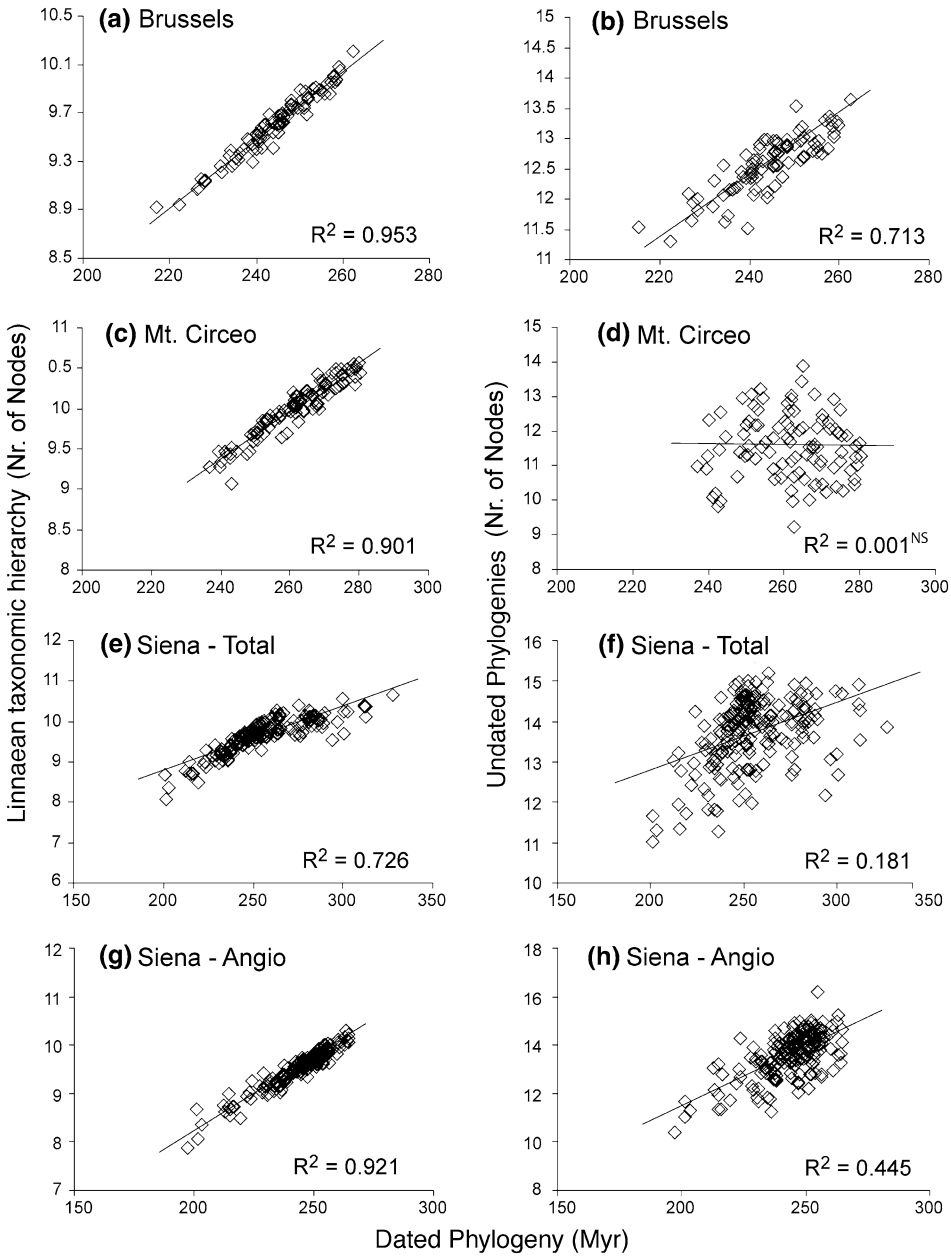

Fig. 1 Linear regressions between a measure of phylogenetic diversity calculated from dated phylogenies and two measures of phylogenetic diversity calculated from undated phylogenies for three datasets: **a**, **b** plant species occurring in green areas in Brussels ($n = 86$ sites); **c**, **d** Mediterranean forests of Mt Circeo ($n = 98$); **e**, **f** all plant species in the plots of the Natura 2000 Network of the Province of Siena ($n = 215$); **g**, **h** angiosperm species only in the Natura 2000 Network of Siena ($n = 215$). The diversity values from the undated phylogenies are calculated using (1) the Linnaean taxonomic hierarchy (*left column*, **a**, **c**, **e**, **g**), and (2) the phylogenetic tree derived from Phylomatic (*right column*, **b**, **d**, **f**, **h**). Each point in the scatter diagrams represents a single community sample. The shown coefficients of determination for the linear regressions between the phylogenetic diversity values are calculated using dated phylogenies versus Linnaean taxonomic hierarchy (*left column*) and dated phylogenies versus undated phylogenies (*right column*). Only Mt Circeo was not significant at the $p = 0.05$ level (NS); values are not Bonferroni-adjusted. The coefficients of determination for the linear regressions between the phylogenetic diversity values calculated using undated branch lengths versus the corresponding Linnaean taxonomic hierarchy are $R^2 = 0.809$ for the Brussels dataset, $R^2 = 0.024$ for Mt Circeo ($p > 0.05$), $R^2 = 0.486$ for Siena-Total, and $R^2 = 0.604$ for Siena-Angio (data not shown)

should represent the same level in the hierarchy. However, it is widely recognized that present systematic classification can be inconsistent in the way it combines lower-level taxonomic ranks into higher-level taxa (Harper and Hawksworth 1995). It is sufficient to look at the complexity of the microspecies in the genera *Hieracium* or *Rubus* to

recognize these problems. Accordingly, while the phylogenetic diversity calculated from dated phylogenies should be roughly correlated to the diversity calculated from taxonomic hierarchies, what is surprising is the strength of the observed relationship.

The second result is that this latter relationship increases if we consider angiosperm species only. That is, the measures of species relatedness obtained from the Linnaean taxonomic hierarchy seem unable to properly summarize the large evolutionary distance between angiosperms and the remaining seed plants. This is because in the Linnaean hierarchy both angiosperms and the remaining lineages of gymnosperms are located at the same hierarchical level; therefore rank differences between co-occurring members of these clades are not going to adequately reflect the differences in evolutionary time separating them. For example, while some debate still remains as to the origin of angiosperms (Triassic, Jurassic, or Cretaceous), we know that cycads, ginkgos, and conifers all originated in the Permian, perhaps even in the Carboniferous (Stewart and Rothwell 1993; Soltis and Soltis 2004). So, even though they share the same rank, they certainly do not reflect the same amount of evolutionary time, and this leads to a biased estimate of phylogenetic diversity measured by the Linnaean hierarchy approach.

Hence, from the above observations, we can drive the following conclusions:

1. because *Phylomatic* uses a backbone tree that is consistent with current level of higher-level taxonomic relationships (e.g., APG 2009), measures of phylogenetic diversity calculated from dated phylogenies are superior to measures obtained from taxonomic hierarchies in depicting evolutionary relationships among distant groups, like angiosperms versus gymnosperms.
2. On the other hand, Phylomatic trees are too coarse-grained to pick up the subtle nuances of phylogenetic diversity among closely related species (see, e.g., Kress et al. 2009). Thus, when dealing with closely related species, a *Phylomatic* tree is no better than a taxonomic hierarchy, irrespective of plot size, lineage composition, or biogeographic region (i.e., Belgium vs. Italy)
3. While *Phylomatic* is only a 'suboptimal' tool for generating comprehensive phylogenies for large species assemblages, at the moment, it is the best available option for incorporating evolutionary relationships into ecological work. For instance, none of the more sophisticated methods of phylogenetic reconstruction are now able to provide dated phylogenies for large and taxonomically heterogeneous numbers of species; rather, they have been usually applied to selected taxonomic groups (e.g., Buerki et al. 2011) or for selected growth forms like trees (e.g., Kress et al. 2010).

As stressed by one anonymous reviewer, the need of relying on *Phylomatic* is due to the general failure of the systematic community to provide a useable synthesis of current phylogenetic knowledge. In this framework, we see *Phylomatic* as a pragmatic tool for easily assembling plant phylogenies that is open to external contributions. As more resolved plant phylogenies are included (or as new tools for phylogeny assembly are developed), the resolution of the output phylogeny will increase, enabling for a more accurate integration of phylogenetic information into studies of community ecology.

Acknowledgments We thank the anonymous referees for their very constructive comments on a previous version of our paper. Data collection in Brussels (dataset A) was financed by the Brussels Institute for Environment Management (IBGE-BIM) in the framework of the research project 'Information and survey network on the biodiversity in Brussels'. G.B. is grateful to Alessandro Chiarucci for providing useful ideas and support during the MOBISIC data collection phase (dataset C). G.B. also warmly acknowledges the valuable support of Elisa Baragatti, Giulia Bennati, Domenico Bernardini, Marta Chincarini, Alessia Delli Bove, Francesco Geri, Sara Ghisleni, Sara Landi, Lia Pignotti, Duccio Rocchini, Elisa Santi, Mauro Taormina, Emanuele Vallone and Arianna Vannini during data collection and plant identification.

References

APG (2009) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 161:105–121

Barker GM (2002) Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. Biol J Linn Soc 76:165–194

Boero F (2010) The study of species in the era of biodiversity: a tale of stupidity. Diversity 2:115–126

Buerki S, Forest F, Salamin N, Alvarez N (2011) Comparative performance of supertree algorithms in large data sets using the soapberry family (Sapindaceae) as a case study. Syst Biol 60:32–44

Cavender-Bares J, Hozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. Ecol Lett 12:693–715

Chiarucci A, Bacaro G, Rocchini D (2008a) Quantifying plant species diversity in a Natura 2000 network: old ideas and new proposals. Biol Conserv 141:2608–2618

Chiarucci A, Bacaro G, Vannini A, Rocchini D (2008b) Quantifying species richness at multiple spatial scales in a Natura 2000 network. Community Ecol 9:185–192

Crozier RH (1997) Preserving the information content of species: genetic diversity, phylogeny and conservation worth. Annu Rev Ecol Syst 24:243–268

Crozier RH, Dunnett LJ, Agapow PM (2005) Phylogenetic biodiversity assessment based on systematic nomenclature. Evol Bioinform Online 1:11–36

Faith DP (1992) Conservation evaluation and phylogenic diversity. Biol Conserv 61:1–10

Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, Balmford A, Manning JC, Proches S, van der Bank M, Reeves G, Hedderson TAJ, Savolainen V (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445:757–760

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

Godefroid S (2001) Temporal analysis of the Brussels flora as indicator for changing environmental quality. Landsc Urban Plan 52:203–224

Harper JL, Hawksworth DL (1995) Preface. In: Hawksworth DL (ed) Biodiversity: measurements and estimation. Chapman and Hall, London, pp 5–12

Kraft NJB, Ackerly DD (2010) Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. Ecol Monogr 80:401–422

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci USA 106:18621–18626

Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. PLoS ONE 5:e15409

Mace GM, Gittleman JL, Purvis A (2003) Preserving the tree of life. Science 300:1707–1709

Ricotta C, Di Nepi M, Guglietta D, Celesti-Grapow L (2008a) Exploring taxonomic filtering in urban environments. J Veg Sci 19:229–238

Ricotta C, Godefroid S, Celesti-Grapow L (2008b) Common species have lower taxonomic diversity: evidence from the urban floras of Brussels and Rome. Divers Distrib 14:530–537

Ricotta C, LaSorte FA, Pyšek P, Rapson GL, Celesti-Grapow L, Thompson K (2009) Phyloecology of urban alien floras. J Ecol 97:1243–1251

Ricotta C, Godefroid S, Rocchini D (2010) Invasiveness of alien plants in Brussels is related to their phylogenetic similarity to native species. Divers Distrib 16:655–662

Rodrigues ASL, Gaston KJ (2002) Maximising phylogenetic diversity in the selection of networks of conservation areas. Biol Conserv 105:103–111

Soltis PS, Soltis DE (2004) The origin and diversification of angiosperms. Am J Bot 91:1614–1626

Stewart WN, Rothwell GW (1993) Paleobotany and the evolution of plants. Cambridge University Press, Cambridge

Strauss SY, Webb CO, Salamin N (2006) Exotic taxa less related to native species are more invasive. Proc Natl Acad Sci USA 103:5841–5845

Swenson NG (2009) Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. PLoS ONE 4:e4390

Swenson NG, Enquist BJ (2009) Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. Ecology 90:2161–2170

Thompson K, Petchey OL, Askew AP, Dunnett NP, Beckerman AP, Willis AJ (2010) Little evidence for limiting similarity in a long-term study of a roadside plant community. J Ecol 98:480–487

Warwick RM, Clarke KR (1994) New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Mar Ecol Prog Ser 129:301–305

Warwick RM, Clarke KR (2001) Practical measures of marine biodiversity based on relatedness of species. Oceanogr Mar Biol Annu Rev 39:207–231

Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. Am Nat 156:145–155

Webb CO, Donoghue MJ (2005) Phylomatic: tree assembly for applied phylogenetics. Mol Ecol Notes 5:181–183

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. Annu Rev Ecol Syst 33:475–505

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24:2098–2100

Wikstrom N, Savolainen V, Chase MW (2001) Evolution of angiosperms: calibrating the family tree. Proc R Soc Lond B 268:2211–2220