Abstract. How closely does variability in ecologically important traits reflect evolutionary divergence? The use of phylogenetic diversity (PD) to predict biodiversity effects on ecosystem functioning, and more generally the use of phylogenetic information in community ecology, depends in part on the answer to this question. However, comparisons of the predictive power of phylogenetic diversity and functional diversity (FD) have not been conducted across a range of experiments. To address how phylogenetic diversity and functional trait variation control biodiversity effects on biomass production, we summarized the results of 29 grassland plant experiments where both the phylogeny of plant species used in the experiments is well described and where extensive trait data are available. Functional trait variation was only partially related to phylogenetic distances between species, and the resulting FD values therefore correlate only partially with PD. Despite these differences, FD and PD predicted biodiversity effects across all experiments with similar strength, including in subsets that excluded plots with legumes and that focused on fertilization experiments. Two- and three-trait combinations of the five traits used here (leaf nitrogen percentage, height, specific root length, leaf mass per unit area, and nitrogen fixation) resulted in the FD values with the greatest predictive power. Both PD and FD can be valuable predictors of the effect of biodiversity on ecosystem functioning, which suggests that a focus on both community trait diversity and evolutionary history can improve understanding of the consequences of biodiversity loss.

Key words: biodiversity–ecosystem functioning; diversity metrics; meta-analysis; model comparison; niche conservation; phylogenetics.

INTRODUCTION

Substantial experimental evidence exists for the positive influence of biodiversity on ecosystem functioning, especially in grasslands, with a focus on aboveground plant biomass production (Balvanera et al. 2006, Duffy 2009). However, which facets of biodiversity most strongly influence ecosystem functioning remains a subject of debate. Recent studies have suggested that phylogenetic diversity (PD, the distinct evolutionary history in a community) can be used as a proxy for measures of functional diversity (FD, the functional trait distinctiveness in a community); this relationship between PD and FD is premised on the reasonable assumption that evolutionary diversification has generated trait diversification, which in turn may result in greater niche complementarity. This theory has been supported by a meta-analysis of biodiversity–ecosystem-functioning studies, finding that phylogenetic diversity (PD) predicted plant biomass accumulation stronger than species richness or functional group richness (Cadotte et al. 2008).

Two issues arise in the use of phylogenetic diversity to predict ecosystem functioning, one important to community ecology in general and one specific to grassland biodiversity–ecosystem-function research. First, the use of PD to predict ecosystem functioning assumes phylogeny represents functional differences relevant to a particular ecosystem function (Maherali and Klironomos 2007). This assumption will hold if there is a strong phylogenetic signal in the traits important for determining ecosystem functioning, or in other words that traits driving community interactions are conserved over the phylogeny, an assumption central to much recent work at the intersection of evolutionary biology and community ecology (e.g., Cavender-Bares et al. 2009). However, while ample evidence for this premise exists for certain traits (e.g., wood density; Chave et al. 2006), a recent study found little correlation between changes in mammal body size variation and changes in phylogenetic diversity (Fritz and Purvis 2010), and phylogeny does not always influence competition (Cahill et al. 2008) or niche structure (Silvertown et al. 2006) in
plants. Among the traits that drive grassland plant biomass accumulation, coevolved relationships with N-fixing bacteria or with pathogens exhibit strong phylogenetic signals, but such signals cannot be assumed for all traits. Directly testing for phylogenetic signal in functional trait variation in the context of ecosystem functioning is crucial for determining if PD can be an effective proxy for FD. In addition, it is worth noting that detection of phylogenetic signal does not necessarily imply phylogenetic niche conservatism (Losos 2008).

Second, since knowledge of which traits are important to ecosystem functioning and access to high-quality trait data are lacking for most species and ecosystem functions of interest, PD would be quite valuable as a proxy for FD. Grassland biodiversity–ecosystem-functioning experiments (see Plate 1) represent the best case for using plant traits to predict aboveground biomass production. Data on grassland plant ecophysiology and life history are copious, although compiling such data remains a challenge. Research in grassland communities has underscored the importance of leaf traits such as leaf mass per unit area (Garnier et al. 2004) and leaf percent nitrogen (Kahmen et al. 2006), belowground traits such as root thickness (Craine et al. 2002) and nitrogen fixation (Lee et al. 2007), and whole-plant traits such as root thickness (Craine et al. 2002), belowground traits such as root thickness (Craine et al. 2002), and nitrogen fixation (Lee et al. 2007) in controlling ecosystem processes. Thus, FD and PD should be directly comparable in predicting biodiversity effects, and how functional differences map onto phylogenies should be examined. The results will show how leveraging information on the functional traits and the evolutionary history of species may provide enhanced means to predict the ecosystem-level consequences of extinctions.

METHODS


Phylogenetic and functional diversity

We calculated PD from the molecular phylogeny of Cadotte et al. (2008), which covered 110 of the species in the meta-analysis, using data for congeners in several cases. In addition, we also calculated PD from a phylogeny extracted from the supertree of Davies et al. (2004) using Phylomatic (Webb and Donoghue 2005; available online), which covered all 121 of the species in the meta-analysis, but with lower phylogenetic resolution. We used the phylogenetic diversity measure PD used by Cadotte et al. (2008), which is the sum of the branch lengths for the species present in a community. This metric is based on the PD developed by Faith (1992), which differs from the present index by always including the root node. For the supertree-based phylogeny, branch lengths were based on the angiosperm-wide divergence dates, interpolated for undated nodes using the branch length adjustment algorithm in the software Phylomatic (Webb et al. 2008). The PD values calculated from these two phylogenies were highly correlated ($R^2 = 0.964$), but yielded different model comparison results.

Calculating functional diversity requires several key decisions. We used the metric FD proposed by Petchey and Gaston (2002) because it exactly parallels PD, accommodates a variety of data types, and has been widely applied as a measure of functional diversity. Which and how many traits are used to calculate FD are the most critical questions in this analysis. We selected a small number of traits known to be important for biomass production in grasslands and for which data are widely available. These traits were leaf mass per unit area (LMA), plant height, leaf nitrogen percentage (%N), specific root length (SRL, a measure of root thickness), and whether the plant supports root nodules capable of biological nitrogen fixation (Table 1). Continuous data were rescaled to center on 0 with a

<table>
<thead>
<tr>
<th>Trait</th>
<th>Sources (n)</th>
<th>Total n</th>
<th>Functional significance</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant height (cm)</td>
<td>8 (46), 2 (52), 3 (4), 4 (11), 5 (7)</td>
<td>120</td>
<td>rate of resource capture; light competition strategy; competitive ability</td>
<td>2.5 (0.5–5.2)</td>
</tr>
<tr>
<td>SRL (g/cm)</td>
<td>6</td>
<td>24</td>
<td>investment belowground, root lifespan</td>
<td>96.6 (22.9–288.4)</td>
</tr>
<tr>
<td>N fixation (binary)</td>
<td>7</td>
<td>121</td>
<td>competitive ability in N-poor soil</td>
<td>0:103; 1:18</td>
</tr>
</tbody>
</table>

standard deviation of 1. We calculated FD values from all 26 combinations of two to five traits for each polyculture, focusing the results on the FD with the best predictive power for a given analysis and the FD with all five traits.

FD requires calculating the multivariate distance between each pair of species based on their functional traits; we used Gower distances to accommodate both the continuous (LMA, %N, height, SRL) and binary data (N fixation; Podani and Schmera 2006). Clustering was performed using the unweighted pair group method with arithmetic means, which gave the highest cophenetic correlation with the original trait distances (0.89) of many clustering algorithms. Trait data came from individual studies (e.g., Craine et al. 2001), published compilations (de Faria et al. 1989, Wright et al. 2004), the LEDA database (Kleyer et al. 2008), reference texts (Grime et al. 1988, Gleason and Cronquist 1991), and unpublished data compilations (D. Bunker, unpublished data).

**Analysis**

For each polyculture, we calculated the net biodiversity effect on aboveground biomass production as the log ratio of the biomass in polyculture (\(y_p\)) to the mean biomass of the constituent species grown in monoculture (\(y_m\)): \(\text{LR}_{\text{mean}} = \ln(y_p/y_m)\) (Cardinale et al. 2006). Since not all experiments had every species grown in monoculture, \(\text{LR}_{\text{mean}}\) could only be calculated for 1433 of the polycultures (see Appendix: Table A1 for data summary). When using PD values calculated from the molecular phylogeny, additional plots were excluded because this phylogeny did not cover all species, yielding 1088 plots.

We assessed the relative importance of each diversity metric in predicting \(\text{LR}_{\text{mean}}\) using single-variable mixed effects models. We further assessed the predictive power of the best functional diversity metric in combination with phylogenetic diversity, to test whether the two types of diversity in combination would yield greater predictive power than either alone. Model parameters were
Table 2. Model comparison results of linear mixed models predicting the log response ratio of biomass production for all plots, including with and without legumes and fertilized experimental plots.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Using PD from molecular phylogeny (110 species)</th>
<th>Using PD from angiosperm supertree (121 species)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>( w_i )</td>
</tr>
<tr>
<td>All plots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>0.196</td>
<td>0.989</td>
</tr>
<tr>
<td>FD[N, height, N-fixation]</td>
<td>0.181</td>
<td>4.8 \times 10^{-5}</td>
</tr>
<tr>
<td>PD + FD[N, height, N-fixation]</td>
<td>0.177</td>
<td>5.5 \times 10^{-6}</td>
</tr>
<tr>
<td>FGR</td>
<td>0.170</td>
<td>7.5 \times 10^{-9}</td>
</tr>
<tr>
<td>No legumes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>0.105</td>
<td>0.48</td>
</tr>
<tr>
<td>FD[N, height]</td>
<td>0.096</td>
<td>0.146</td>
</tr>
<tr>
<td>PD + FD[N, height]</td>
<td>0.107</td>
<td>0.064</td>
</tr>
<tr>
<td>FGR</td>
<td>0.074</td>
<td>3.3 \times 10^{-6}</td>
</tr>
<tr>
<td>Fertilized plots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>0.186</td>
<td>0.117</td>
</tr>
<tr>
<td>FD[N, height, N-fixation]</td>
<td>0.172</td>
<td>0.216</td>
</tr>
<tr>
<td>PD + FD[N, height, N-fixation]</td>
<td>0.188</td>
<td>0.024</td>
</tr>
<tr>
<td>FGR</td>
<td>0.123</td>
<td>6.7 \times 10^{-5}</td>
</tr>
</tbody>
</table>

**Notes:** Predictors are ranked by Akaike weight; best-fit models are shown in boldface type. Comparisons were performed between 26 trait combinations for functional diversity (FD), phylogenetic diversity (PD), species richness (S), and functional group richness (FGR), and a multivariate model combining PD and the best FD. Subscripts for FD indicate the traits used. Results are shown from PD based on the molecular phylogeny of Cadotte et al. (2008), which covers 110 of the 121 species used in these plots, as well as from PD based on the angiosperm supertree. Cadotte et al. (2008) created a phylogeny of 145 species, of which 121 are present in plots where \( LR_{\text{mean}} \) can be calculated. Key to abbreviations: \( N \), number of experimental units in this subset; \( w_i \), Akaike weight.

estimated by restricted likelihood estimation, and compared by Akaike weights. Goodness-of-fit for these models was assessed by \( R^2 \) of observed and model-fitted \( LR_{\text{mean}} \) values. Fourteen outliers identified from a Bonferroni two-sided test on Studentized residuals were removed. We examined two subsets of the data set, separately examining the diversity metrics in experimental units that (1) did not include legumes, and (2) were experimentally fertilized. Legume presence is an important factor in many grassland biodiversity experiments (e.g., Marquard et al. 2009), and biodiversity-ecosystem-function relationships can vary depending on soil fertility conditions (Reich et al. 2001, Lanta and Leps 2007), so these subsets allowed us to compare these different aspects of biodiversity under different conditions.

In order to account for the covariations among the alternative measures of biodiversity, we also employed structural equation modeling (SEM). Both the PD and FD metrics used here are highly dependent on species richness. The models tested reflect this dependency, and are constructed to test how PD and/or FD mediate the effect of species richness on the biodiversity effect (\( LR_{\text{mean}} \)). Alternative pathways included direct effects of \( S \) on the ecosystem function, the inclusion of functional group richness, and correlations between FD and PD (Appendix: Fig. A1). SEMs were implemented using the R package sem (Fox 2006).

We assessed the phylogenetic signal in the functional traits at three levels. First, we compared the relationship between PD and FD. Second, we compared the distances between species based on functional traits with distances based on phylogeny; these distances are the foundation for the diversity metrics. We tested the degree of phylogenetic signal in each trait using the K statistic (Blomberg et al. 2003), as implemented in the R package picante (Kembel et al. 2010). All analyses used the statistical programming software R 2.11.0 (available online).\(^4\)

**Results**

PD as calculated from the molecular phylogeny was the most parsimonious predictor of the biodiversity effects on aboveground biomass (Fig. 1), followed closely by the combination of PD and the FD calculated from leaf %N, mean plant height, and N-fixation ability (FD[N, height, N-fixation]), and then by FD[N, height, N-fixation] alone. In the most inclusive comparison, using 1419 plots and the PD based on the angiosperm supertree, FD[N, height, N-fixation] was the best predictor of the effect of plant biodiversity on aboveground biomass production (Table 2).

Thus, the model comparison yielded a different order of predictors depending on which phylogeny was used, although the amount of variance explained by both PD and FD (18–23%) was similar within each subset of the data (Table 2). The use of the more highly resolved molecular phylogeny resulted in PD being the more parsimonious predictor overall, and for the plots that

\(^4\) (http://www.r-project.org)
did not include legumes. Examining only experiments where N fertilizer was added, FD[height, N-fixation] was the most parsimonious predictor, regardless of which phylogeny was used in calculating PD. In every case, FGR was the weakest predictor of biodiversity effects. Combining PD with the best FD resulted in greater variance explained for the biodiversity effect on aboveground biomass, but was not the most parsimonious model in any case.

Despite the similar power for FD and PD to predict biodiversity effects in grassland experiments, the relationship between the indices results almost entirely from the correlation of each with S. Both PD and FD increase with S, but while PD increases nearly linearly with S, a large range of FD values was found at all levels of S (Fig. 2), resulting in a modest relationship between FD and PD (e.g., FD[N, height, N-fixation] and PD, $R^2 = 0.237$). Correspondence between FD and PD is not strong for any given number of species (Appendix: Fig. A2), resulting in a minimal relationship between PD and FD when the effect of S is removed (residuals of FD[N, height, N-fixation] and PD against S, $R^2 = 0.024$; Appendix: Fig. A3).

Comparison of competing structural equation models demonstrated that the best-fit model required including both PD and FD as predictors of the biodiversity effect for all sets of the data. Including the correlation between PD and FD improved the model fit for subsets excluding legume-containing plots or unfertilized plots, but not for the other sets of data (Table 3). However, in all cases when the correlation was included, the value was small (e.g., Fig. 3). The strength of the predictive power of PD and FD in the SEMs largely corroborated the results of the linear mixed models.

The only trait showing a significant phylogenetic signal was N-fixing ability (Table 4, Appendix: Fig. A4). When using the angiosperm supertree, with a complete coverage of species but only genus-level resolution,
significant phylogenetic signal was detected for LMA, height, and N-fixing ability, indicating that close relatives were more likely to have similar trait values than would be expected by chance.

**Discussion**

Our analyses demonstrate that measures of functional and phylogenetic diversity have similar abilities to predict biodiversity effects and that functional group richness has the weakest predictive power in nearly all cases. The similar predictive power of FD and PD (Fig. 1) is surprising because the two indices are based on mostly different information: ecophysiological traits for FD vs. time since evolutionary divergence for PD. There is robust evidence for phylogenetic signal in N fixation, unsurprisingly, and some evidence for phylogenetic signal in LMA and height, but the diversity metrics summarizing the functional and phylogenetic information do not correlate after the effect of species richness is removed. SEMs demonstrated small or zero correlation between the two diversity metrics when species richness was also included.

The lack of correlation between FD and PD values for communities of a given species richness suggests that

![Diagram](image)

**Fig. 3.** Best-fit structural equation model combining S, FD, and PD calculated from the molecular phylogeny, across all plots ($\chi^2 = 3.37$, df = 1, P = 0.067). The model shown, M8, includes a correlation between FD and PD. Model M3, the best model for three subsets of the data, excludes this correlation (see Table 3). Values give the standardized coefficients for the relationship between “upstream” and “downstream” variables; all coefficients are significant. Epsilons represent the error term for downstream variables. See the Appendix for the full set of models.
while the traits used in the FD calculations are important, additional axes of trait variation are captured in PD. These unmeasured traits may include pathogen tolerance (Petermann et al. 2008) or other coevolutionary relationships, and seem to be important in determining grassland ecosystem functioning. PD potentially captures all such additional axes, but is not informative for identifying what they might be. Identifying the traits that drive ecosystem functioning will spur better understanding of the consequences of species loss and the mechanisms driving ecosystem processes, such as niche complementarity and the selection effect, and will clarify how evolutionary history can be a good proxy for trait measurements.

We found that variation in leaf %N, height, and N fixation were consistently the most important traits for predicting biodiversity effects. Leaf N concentration relates to resource acquisition strategy, while height relates to partitioning of light resources in grasslands (Grime 2001). Differentiation in height and LMA was partially driven by phylogenetic relationships (Table 4). N fixation coincides completely with Fabaceae, and is the only trait with an overwhelming phylogenetic signal. However, PD was still an effective predictor of the biodiversity effect even when plots with legumes were excluded (Table 2). Thus, phylogenetic divergence can reflect functional differentiation, but this does not result in diversity metrics that correspond closely at a given level of species richness. The variation in the strength of the phylogenetic signal can reflect variation in the rate of evolution of different traits, which would result in a nonlinear relationship between character divergence and

<table>
<thead>
<tr>
<th>Trait</th>
<th>Using molecular phylogeny</th>
<th>Using angiosperm supertree</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMA</td>
<td>0.240 64</td>
<td>0.326 87</td>
</tr>
<tr>
<td>N</td>
<td>0.268 45</td>
<td>0.343 63</td>
</tr>
<tr>
<td>SRL</td>
<td>0.282 11</td>
<td>0.358 24</td>
</tr>
<tr>
<td>Height</td>
<td>0.273 82</td>
<td><strong>0.635</strong> 120</td>
</tr>
<tr>
<td>Nitrogen fixation</td>
<td><strong>6.197</strong> 83</td>
<td><strong>9.017</strong> 119</td>
</tr>
</tbody>
</table>

**Notes:** Key to abbreviations: LMA, leaf mass per unit area; n, number of species with trait data represented in the given phylogeny. Values in bold are statistically significant ($P \leq 0.05$).

**Plate 1.** A biodiversity–ecosystem-functioning experiment in Inner Mongolia, China, being carried out by researchers at the Institute of Botany in the Chinese Academy of Sciences. Cages for experimental enclosure of grasshoppers appear on a subset of plots. Photo credit: S. Naeem.
phylogenetic distance. A branch-summing metric like PD assumes linear relationships, while alternative measures of phylogenetic diversity (e.g., Helms et al. 2007) do not. Just as different measures of functional diversity capture different aspects of trait variation in communities (Petchey et al. 2009), researchers must consider how the details of a phylogenetic diversity calculation, including the variation in rates of character divergence, impact the answer to the research question.

Previous studies have evaluated the performance of different diversity metrics in predicting biodiversity–ecosystem-function relationships, notably Petchey et al. (2004), who demonstrated that FD was a stronger predictor of aboveground biomass production than S or FGR. Notably, Cadotte et al. (2009) assessed PD, several versions of FD, and other diversity metrics as predictors of the biodiversity effect in one of the studies included in this meta-analysis. They found that FD and PD were weakly correlated, but that PD and combinations of PD and other metrics were always superior predictors of ecosystem functioning. This contrasts with our results, but their study differed from the current study because they used a different set of traits, fewer species, and focused on a single biodiversity experiment. These contrasting results highlight the need for a mechanistic understanding of which traits are represented by PD.

Importantly, other studies have found that the traits of the dominant species can be more important than any aggregate measure of functional diversity in determining ecosystem processes (Mokany et al. 2008, Griffin et al. 2009). This highlights the need for further analyses of how plant traits control ecosystem processes to partition complementarity from selection effects, which we did not address here. In addition, trait data compilation remains a challenge, with a clear need for a central repository of functional trait data. We suggest that further progress in resolving these issues will require examining for which traits and to what extent evolutionary relationships closely match functional relationships, to address the question of how much phylogenetic niche conservatism there may be for the traits important for ecosystem functioning (Ackerly and Reich 1999).

Acknowledgments

We thank the authors of the original studies for generously sharing their data and M. Cadotte, B. Cardinale, and T. Oakley for sharing their molecular phylogeny and R code for calculating PD. We thank B. Schmid, members of the Naem lab, and two anonymous reviewers for substantial and constructive feedback.

Literature Cited


APPENDIX

Supporting information for model comparisons (Ecological Archives E092-132-A1).