Can functional traits explain phylogenetic signal in the composition of a plant community?

Daijiang Li[1*], Anthony R. Ives[2], Donald M. Waller[1]

[1]Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin, 53706

[2]Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin, 53706

* Correspondence: daijianglee@gmail.com; arrives@wisc.edu; dmwaller@wisc.edu

Authorship: DL and AI designed the study and performed the analyses. DL collected the vegetation and environmental data. DL and DW collected the functional traits data. DL wrote the first draft of the manuscript and all authors contributed substantially to revisions.

Running title: Functional traits and phylogeny

Key words: Functional traits, community assembly, phylogeny, phylogenetic linear mixed model.

Words in abstract: 223; Words in main text: 4998

Number of references: 43; Number of figures: 2, Number of tables: 4

Abstract:

Phylogeny-based and functional trait-based analyses are two principle ways to study community assembly and underlying ecological processes. In principle, knowing all information about species traits would make phylogenetic information redundant, at least that component of phylogenetic signal in the distribution of species among communities that is caused by phylogenetically related species sharing similar traits. In reality, phylogenies may contain more information than a set of singular, discretely measured traits because we cannot measure all species traits and may misjudge which are most important. The extent to which functional trait information makes phylogenetic information redundant, however, has not been explicitly studied with empirical data in community ecology. Here, we use phylogenetic linear mixed models to analyze community assembly of 55 understory plant species in 30 forest sites in central Wisconsin. These communities show strong phylogenetic attraction, yet variation among sites in 20 environmental variables could not account for this pattern. Most of the 15 functional traits we measured had strong phylogenetic signal, but only three varied strongly among sites in ways that affected species' abundances. These three traits explained only 19% of variation in phylogenetic patterns of species co-occurrence. Thus, phylogenies appear to provide considerably more information about community assembly than the functional traits measured in this study, demonstrating the value of phylogeny in studying of community assembly processes even with abundant functional traits.

Introduction

Functional traits, arising as innovations through evolution, can capture essential aspects of species' morphology, ecophysiology, and life-history strategy (McGill *et al.* 2006; Violle *et al.* 2007). Although closely related species can differ greatly in some functional traits due to rapid evolution or convergence (Losos, 2008, 2011), most functional traits show strong phylogenetic signal (Moles *et al.* 2005, Donoghue 2008, Freckleton *et al.* 2002; Webb *et al.* 2002). Functional traits, with or without phylogenetic signal, are known to influence the species composition of communities, thereby providing mechanistic links between fundamental ecological processes and community structure (McGill *et al.* 2006; Violle *et al.* 2007; Adler *et al.* 2013). Functional traits also provide a common currency that facilitates comparisons among species and across regions, allowing us to assess the generality of patterns and predictions in community ecology (McGill *et al.* 2006). This has lead to a proliferation of studies using functional traits to understand community structure and composition. Functional trait-based approaches, however, are limited by the fact that it is impossible to measure all potentially important functional traits affecting the distribution of species across communities.

Even in the absence of functional trait information, it is still possible to infer the effects of (unmeasured) functional traits on community composition by investigating phylogenetic patterns in community composition. Phylogenies play an important role in community ecology by giving information about evolutionary relationships among species (Graves & Gotelli, 1993; Losos 1996; Baum & Smith, 2012). Because phylogenetically related species often share similar functional trait values, we expect phylogenetically related species to co-occur more often in the same communities reflecting their shared environmental tolerances. Conversely, if

64 phylogenetically related species have similar traits that cause them to compete more, then closely
65 related species may be less likely to co-occur. These and other processes relating functional traits
66 to community composition likely lead to phylogenetic signatures in how species are distributed
67 among communities (Webb *et al.* 2002). However, in principle, if we have information for all
68 relevant functional traits, then we expect phylogeny to provide little additional information
69 relevant for community composition. That is, when all of the functional traits affecting
70 community composition are known, we do not expect the unexplained residual variation in the
71 occurrence of species to have phylogenetic signal (Ives & Helmus, 2011).

72 In practice, we cannot obtain information about all relevant functional traits. In addition,
73 phylogenetic signals in community composition may result from factors beyond functional traits
74 that generate phylogenetic signal, such as the biogeographical patterns generated as species
75 disperse across a landscape (Ricklefs *et al.* 1993; Moen *et al.* 2009). Therefore, even after
76 accounting for those functional traits whose measurements are available, we should expect
77 phylogenies to contain additional information about community composition (Vane-Wright *et al.*
78 1991; Cadotte *et al.* 2009). As far as we know, however, no study has explicitly assessed how
79 much information about community assembly provided by traits and phylogeny is overlapping.
80 Here, we ask how much of the phylogenetic signal in the composition of a plant community
81 assemblage can be explained by functional traits (Fig. 1).

82 To address this question, we take advantage of our data set of 30 understory communities
83 containing a total of 55 species from Wisconsin pine barrens for which we built a highly resolved
84 phylogeny, amassed a large database on 15 functional traits, and measured 20 environmental
85 variables at each community. We first investigate whether there is phylogenetic pattern in
86 community composition, using a phylogenetic community mixed model that tests for both

"phylogenetic attraction" (phylogenetically related species more likely to occur in the same communities) and "phylogenetic repulsion." If there is phylogenetic pattern, then it could be produced by measured functional traits that themselves have phylogenetic signal (Fig. 1, arrows 2, 4, and 7), unmeasured functional traits with phylogenetic signal (Fig. 1, arrows 2, 5, and 8), or phylogenetic processes unrelated to functional traits (Fig. 1, arrow 6). We then develop a phylogenetic community mixed model incorporating the functional traits we measured to ask whether there is phylogenetic signal in the residual variation in community composition after the effects of these functional traits are removed. This analysis tests the hypothesis that we can explain all the phylogenetic pattern in community composition using measured functional traits. Finally, we use a phylogenetic community mixed model to investigate whether phylogenetically related species respond similarly to environmental gradients across the communities. The motivation for this analysis is to indirectly identify possible unmeasured functional traits that might play a role in community composition. Cases where phylogenetically related species respond similarly to an environmental gradient suggest that they share traits that confer similar tolerances to, or preferences for, specific environmental conditions; this analysis therefore may suggest what additional functional traits to measure in order to explain patterns in community composition.

Methods

Data

Community composition. – We sampled 30 pine barrens forest sites in the central Wisconsin sand plains in 2012 using 50 1-m^2 quadrats placed along five transects at each site. Within each quadrat, we recorded the presence/absence of all vascular plant species (see Li & Waller 2015 for details). Across all sites, we recorded 152 species. For the analyses other than the initial exploration of phylogenetic patterns in community composition, we focused on the 55 species that occurred in three or more communities. This was done for two reasons. First, we did not have values for all functional traits for many of the rare species. Second, we wanted to limit the number of zeros in the data set, especially in analyses of responses of species to environmental variables in which one or two observations of a species give little information.

Functional traits. – For the 55 focal species, we measured 11 continuous and four categorical functional traits on at least 12 individuals from at least three populations using standard protocols (Pérez-Harguindeguy *et al.* 2013). Continuous traits include seed mass (*g*/seed), plant height (*cm*), specific leaf area (SLA, m^2/kg), leaf dry matter content (LDMC, %), leaf circularity (dimensionless), leaf length (*cm*), leaf width (*cm*), leaf thickness (*mm*), leaf carbon concentration (%), leaf nitrogen concentration (%), and stem dry matter content (SDMC, %). We aggregated categories of each categorical trait into two levels: growth form (woody vs. non-woody), life cycle (annual vs. non-annual), and pollination mode (biotic vs. abiotic). We divided seed dispersal mode into three binary variables (wind dispersed vs. not, animal dispersed vs. not, and unassisted vs. assisted dispersal). Collectively, these functional traits, covering the leaf-height-seed (LHS) plant ecology strategy (Westoby, 1998), represent multidimensional functions of

127 plants associated with resource use, competitive ability, dispersal ability, etc. For analyses, trait
128 values were Z-transformed to have means of zero and standard deviations of one, allowing
129 coefficients in the mixed models to be interpreted as effects sizes.

130 *Phylogeny*. – The phylogeny used in this study is a subset of a phylogeny for all vascular plants
131 in Wisconsin (Cameron *et al.* unpublished data[1]). Briefly, Cameron *et al.* used two plastid DNA
132 barcode loci *rbcL* and *matK* to build the phylogeny using maximum likelihood (ML) in the
133 program R_{AXML} (Stamatakis, 2014). The phylogeny was then time-calibrated using the branch
134 length adjuster (*bladj*) available in the program phylocom (Webb *et al.* 2008).

135 *Environmental data*. – At each site, we collected six soil samples and then pooled the samples
136 together to measure the soil properties listed in Table 4. We also took six vertical fish-eye
137 photographic images at each site to measure canopy cover. To characterize climatic conditions,
138 we extracted daily precipitation and minimum temperature for each site from interpolated values
139 estimated by Kucharik *et al*. (2010) from 2002 to 2006 (data after 2006 were not available). All
140 environmental variables were also scaled to have means of zero and standard deviations of one.

141 Phylogenetic community composition

142 We first tested for phylogenetic community structure without including environmental or
143 functional trait information. We used traditional metrics and randomization tests (i.e., null
144 models) to identify whether there was phylogenetic pattern (phylogenetic attraction or repulsion)
145 in the composition of our 30 communities. Specifically, we measured the phylogenetic structure
146 of species abundances at each site using phylogenetic species evenness (PSE, Helmus *et al.*

[1] Cameron, K., R. Kriebel, M. Pace, D. Spalink, P. Li, and K. Sytsma. *In prep*. A complete molecular community phylogeny for the flora of Wisconsin based on the universal plant DNA barcode.

2007) and mean phylogenetic distance (MPD, Webb, 2000). For a site with m individuals from n species, if we let **M** to be a $n \times 1$ column vector containing abundance of all species and let variance-covariance matrix **C** describing phylogenetic relationship of these species, we can calculate PSE as $\frac{m\mathrm{diag}(\boldsymbol{C})'\boldsymbol{M} - \boldsymbol{M}'\boldsymbol{C}\boldsymbol{M}}{m^2(1-1/n)}$ (the prime denotes transpose) and calculate MPD as $\frac{\sum_i^n \sum_j^n \boldsymbol{C}_{[i,j]}\boldsymbol{M}_i\boldsymbol{M}_j}{\sum_i^n \sum_j^n \boldsymbol{M}_i\boldsymbol{M}_j}$, $i \neq j$. PSE is scaled from zero to one, with one occurring when **C** is the identity matrix and all species have equal abundance. We then calculated the mean of these 30 PSE ($\overline{\mathrm{PSE}}_{\mathrm{obs}}$) and 30 MPD values ($\overline{\mathrm{MPD}}_{\mathrm{obs}}$). To test for phylogenetic pattern using these metrics, we permuted species randomly among sites (SIM2 in Gotelli, 2000) 4999 times and then calculated metrics base on each permutation data set. This null permutation model retains the prevalence of each species across sites, but allows sites to change in species richness. If $\overline{\mathrm{PSE}}_{\mathrm{obs}}$ or $\overline{\mathrm{MPD}}_{\mathrm{obs}}$ falls above (or below) 97.5% of the permutation values, then we infer a statistically significant phylogenetic repulsion (attraction). Using this null model where sites can vary in species richness is justified because under the null hypothesis of no phylogenetic signal, the values of PSE and MPD are independent of species richness at the sites.

In addition to testing the significance of the observed PSE and MPD values via null models, we fit a phylogenetic linear mixed model (PLMM) to test for phylogenetic community patterns in species abundances. A PLMM establishes a flexible statistical base to subsequently incorporate functional trait and environmental variables. Furthermore, PLMMs tend to have greater statistical power than other metrics examined using permutation tests (Ives & Helmus, 2011). To build the PLMM, let n be the number of species distributed among m sites. Letting Y be the $mn \times 1$ vector containing the abundance of species j ($j = 1, \ldots, n$) at site s ($s = 1, \ldots, m$), the PLMM is

$$\log(Y + 1) = \alpha + a_{\mathrm{spp}[i]} + b_{\mathrm{spp}[i]} + c_i + d_{\mathrm{site}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{S}_{\text{spp}})$$

$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \mathbf{S}_{\text{nested}}))$$

$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \quad (1)$$

We use the convention of multilevel models here (Gelman & Hill, 2007), with fixed and random effects given by Greek and Latin letters, respectively. The function spp[i] maps the observation i in vector Y to the identity of the species (Gelman & Hill, 2007, p251-252), so i takes value from 1 to mn. The intercept α estimates the overall average log abundance of species across all sites. The following three random variables $a_{\text{spp}[i]}$, $b_{\text{spp}[i]}$ and c_i incorporate variation in abundance among plant species. Specifically, the n values of $a_{\text{spp}[i]}$ give differences among species in mean log abundance across all sites and are assumed to be drawn independently from a Gaussian distribution with mean 0 and variance σ_a^2. The n values of $b_{\text{spp}[i]}$ also give differences in mean log abundance across sites but are assumed to be drawn from a multivariate Gaussian distribution with covariance matrix $\sigma_b^2 \mathbf{S}_{\text{spp}}$, where the $n \times n$ matrix $\mathbf{S}_{\text{spp}}$ is derived from the phylogeny (see next paragraph below), and the scalar σ_b^2 dictates the overall strength of the phylogenetic signal. Thus, $a_{\text{spp}[i]}$ and $b_{\text{spp}[i]}$ together capture variation in mean species log abundances that is either unrelated to phylogeny or has phylogenetic signal. The random variable c_i accounts for covariance in the log abundances of plant species nested within sites (using the Kronecker product, ⊗). Specifically, c_i assesses whether phylogenetically related plant species are more or less likely to co-occur at the same sites. Hence, c_i is used to measure either phylogenetic

attraction or phylogenetic repulsion; because σ_c^2 dictates the overall strength of these phylogenetic patterns, it is the key term we are interested in. Random effect $d_{\text{site}[i]}$ is assumed to contain m values, one for each site, that are distributed by a Gaussian distribution with variance σ_d^2 to account for differences in the average log abundances of species from site to site. Finally, e_i captures residual variance σ_e^2.

We base the phylogenetic covariance matrix $\mathbf{S}_{\text{spp}}$, on the assumption of Brownian motion evolution. If a continuous-valued trait evolves up a phylogenetic tree with a constant probability of increasing or decreasing slightly, the covariance in trait values between two species will be proportional to the length of shared evolution, given by the distance on the phylogenetic tree between the root and the species' most recent common ancestor (Martins & Hanson 1997). This gives a direct way to convert the phylogeny into a hypothesis about the covariance matrix. For the assessment of phylogenetic attraction within sites, c_i, we use $\mathbf{S}_{\text{nested}} = \mathbf{S}_{\text{spp}}$. For phylogenetic repulsion, we use the matrix inverse of $\mathbf{S}_{\text{spp}}$, $\mathbf{S}_{\text{nested}} = (\mathbf{S}_{\text{spp}})^{-1}$. Theoretical justification for $\mathbf{S}_{\text{nested}} = (\mathbf{S}_{\text{spp}})^{-1}$ comes from a model of competition among community members (Ives & Helmus 2011, Appendix A). Briefly, if the strength of competition between species is given by $\mathbf{S}_{\text{spp}}$, as might be the case if closely related species are more likely to share common resources, then the relative abundances of species will have covariance matrix $(\mathbf{S}_{\text{spp}})^{-1}$.

Equation 1 is the same as model I in Ives & Helmus (2011), except model I includes variation among species in mean log abundance across sites as fixed effects rather than two random effects, $a_{\text{spp}[i]}$ and $b_{\text{spp}[i]}$. This change allows us to align equation 1 with equation 3 below that includes variation in the relationship between trait values and log abundance within sites as random effects. In our analyses, treating variation among species in mean log abundance as fixed effects (results not presented) led to almost identical estimates of phylogenetic signal (estimates

of σ_c^2), and therefore our treatment of $a_{spp[i]}$ and $b_{spp[i]}$ as random effects does not change the conclusions.

We fit the PLMM with maximum likelihood using function communityPGLMM in the pez library of R (R Core Team, 2015). Statistical significance of the variance estimates σ^2 was determined using a likelihood ratio test. Because the null hypothesis $\sigma^2 = 0$ is on the boundary of the parameter space (because σ^2 cannot be negative), we used the $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution of Self & Liang (1987) for significance tests. The distribution of χ_0^2 represents a distribution with a point mass at 0, and the *p*-values given by the constrained likelihood ratio test are one-half the values that would be calculated from a standard likelihood ratio test using χ_1^2. Simulations suggest that *p*-values calculated in this way are more conservative than those from a parametric bootstrap (Appendix Text S1).

Our data set contained many zeros, raising the question of the validity of applying a linear model to transformed data. Nonetheless, transforming data and applying a linear analysis is robust when assessing the significance of regression parameters (Ives, 2015). To check this robustness for our results, we repeated all analyses using phylogenetic generalized linear mixed models (PGLMM) after converting our abundance data to presence/absence data. PGLMMs yielded qualitatively similar results in all of the analyses we present in the main text; these PGLMM analyses are presented in Appendix. We present the PLMM results in the main text, because including abundance data in phylogenetic community structure analyses can provide more information about community assembly (Freilich & Connolly, in press).

Can functional traits explain phylogenetic community composition?

To quantify how much of the variation in phylogenetic patterns can be explained by measured functional traits, we estimated PLMMs with and without functional traits, and then compared the strength of phylogenetic signal in the residual variation: if functional traits alone serve to explain phylogenetic community composition, then as functional traits are included, the strength of the phylogenetic signal in the residuals should decrease. We selected functional traits one by one based on the two conditions necessary for them to generate phylogenetic signal in community composition. First, a functional trait must show phylogenetic signal among species, because in the absence of phylogenetic signal among species, a trait could not produce phylogenetic signal in species' abundances. Second, there must be variation among sites in which species are selected according to their functional traits; if a trait has phylogenetic signal but there is no variation of relationships between plant functional trait values and abundances among sites, then species with similar trait values will also have similar overall abundance. This will contribute to the overall phylogenetic signal of species abundance and will be captured by $b_{\mathrm{spp}[i]}$ in equation 1. Therefore, we only investigate traits that exhibit both strong phylogenetic signal *and* variation among sites in the apparent advantages the traits give to species.

Phylogenetic signal of functional trait was tested with model-based methods. Each continuous trait was tested with Pagel's λ (Pagel, 1999) using phylolm (Ho & Ané, 2014). For the binary traits, we applied phylogenetic logistic regression (Ives & Garland, 2010) as implemented by phyloglm (Ho & Ané, 2014). We also tested phylogenetic signal of functional traits via Blomberg's K (Blomberg *et al*. 2003) with picante (Kembel *et al*. 2010).

We tested variation of relationships between trait values and log abundances with the LMM

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + (\beta + b_{\text{site}[i]})t_{\text{spp}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2\mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2\mathbf{I}_{mn}) \qquad (2)$$

where $t_{\text{spp}[i]}$ is the focal functional trait value of the species corresponding to observation i, and σ_b^2 gives the variation among sites of the relationship between species trait values and log abundances. If $\sigma_b^2 > 0$, we conclude that different sites select species differently based on the tested trait. Significance of σ_b^2 was tested with a likelihood ratio test.

We quantified the contribution of a trait to the observed phylogenetic pattern using the model

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]} + (\beta + f_{\text{site}[i]})t_{\text{spp}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2\mathbf{S}_{\text{spp}})$$

$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2\mathbf{S}_{\text{nested}}))$$

$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2\mathbf{I}_m)$$

$$f \sim \text{Gaussian}(\mathbf{0}, \sigma_f^2\mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2\mathbf{I}_{mn}) \qquad (3)$$

This model is the same as equation 1 used to assess phylogenetic patterns in community composition, except that it includes functional trait values $t_{\text{spp}[i]}$. Random variables $a_{\text{spp}[i]}$, $b_{\text{spp}[i]}$,

c_i, and $d_{\text{site}[i]}$ are defined as in equation 1. The proportion of phylogenetic signal in species composition (measured by σ_c^2) that this trait can explain is then assessed by comparing σ_c^2 between models with and without this trait. Finally, to evaluate the overall contribution of functional traits to the observed phylogenetic patterns, we built a multivariate version of equation 3 which included all traits that have both phylogenetic signal *and* strong variation among sites.

Does any environmental variable drive phylogenetic pattern?

If phylogenetic patterns in community composition are observed, yet no functional traits can explain the patterns, how could we identify additional functional traits that might be responsible? Phylogenetically related species usually are assumed to be ecologically similar due to niche conservatism (Wiens *et al.* 2010). Therefore, related species will tend to have similar responses to environmental variable. If these environmental variables are strong enough to drive phylogenetic patterns in community composition, then functional traits that are associated with tolerance or sensitivity to these environmental variables will likely be important in explaining community composition. Thus, we investigated phylogenetic patterns in the responses of species to environmental variables to suggest additional, unmeasured functional traits that might be important to explain phylogenetic patterns in community composition.

Given the large number of environmental variables we have, we first selected variables for which there is significant variation among species' responses without accounting for their phylogenetic relationships, using linear mixed model (LMM) structured as:

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + (\beta + b_{\text{spp}[i]})x_{\text{site}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{a}}\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{b}}\mathbf{I}_n)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{e}}\mathbf{I}_{mn}) \quad (4)$$

As in model 1, spp[i] gives the species identity of datum i while site[i] gives the site identity of datum i. This makes $x_{\text{site}[i]}$ the value of the focal environmental variable at the site corresponding to observation i in the data. Variation among species in their response to the environmental variable x is given by the σ^2_{b} term here. If there is significant variation in species' responses to this variable ($\sigma^2_{\text{b}} > 0$), we then further tested whether phylogenetically related species respond to environmental variables in a similar way using the PLMM model

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + (\beta + c_{\text{spp}[i]} + d_{\text{spp}[i]})x_{\text{site}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{a}}\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{b}}\mathbf{S}_{\text{spp}})$$

$$c \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{c}}\mathbf{I}_n)$$

$$d \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{d}}\mathbf{S}_{\text{spp}})$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{\text{e}}\mathbf{I}_{mn}) \quad (5)$$

Here, $c_{\text{spp}[i]}$ and $d_{\text{spp}[i]}$ represent non-phylogenetic and phylogenetic variation among species in their response to environmental variable x (see model II in Ives & Helmus, 2011). The key parameter of interest is σ^2_{d}, which we tested using a likelihood ratio test. If $\sigma^2_{\text{d}} > 0$, phylogenetically related species respond to environmental variable x in similar ways, suggesting the existence of an unmeasured phylogenetically inherited trait that is associated with species tolerances or sensitivities to x.

Results

Phylogenetic community composition

Phylogenetically related species co-occur more often than expected by chance in pine barrens communities in central Wisconsin (Fig. 2). Permutation tests including all 152 species show that closely related species are likely to co-occur within communities, as judged by either phylogenetic species evenness ($\overline{\mathrm{PSE}}_{\mathrm{obs}}$ = 0.32, p = 0.03) or mean phylogenetic distance ($\overline{\mathrm{MPD}}_{\mathrm{obs}}$ = 338, p = 0.01). Confining the analyses to the 55 focal species that occurred in at least three communities, the permutation tests did not statistically significant phylogenetic patterns ($\overline{\mathrm{PSE}}_{\mathrm{obs}}$ = 0.27, p = 0.29; $\overline{\mathrm{MPD}}_{\mathrm{obs}}$ = 286, p = 0.17), although the PLMM gave statistically significant phylogenetic patterns (p = 0.008; Table 1). Converting the data to presence/absence among communities, the PGLMM also confirmed phylogenetic attraction pattern in these communities (Appendix Table S1). The stronger statistical results for the PLMM and PGLMM are not unexpected, given that model-based approaches are generally statistically more powerful that metric-based approaches (Ives & Helmus 2011).

Can functional traits explain phylogenetic community composition?

Most functional traits show strong phylogenetic signal (Table 2). Three traits – SLA, leaf circularity, and leaf thickness – significantly affected plant species' abundances among sites ($\sigma_b^2 >$ 0, equation 2, Table 2). In other words, different sites selected different species based on these three functional traits. Individually, SLA, leaf circularity, and leaf thickness reduced the phylogenetic variance by 6%, 2%, and 6% measured by the reduction in σ_c^2 with the inclusion of these functional traits (equation 3). Including all three traits in the final model, the phylogenetic

335 variation σ_c^2 decreased 19%. In other words, the many functional traits we measured in this study
336 can only explain 19% of the phylogenetic signal in species composition among these
337 communities. Converting the data to presence/absence and using the PGLMM, σ_c^2 was reduced
338 by 53% (Appendix, Table S2). Thus, functional traits explained more of the phylogenetic
339 patterns in community composition in the presence/absence of species from communities than in
340 their log abundance, although this reduction still only explains about one-half of the phylogenetic
341 pattern in community composition.

342 Does any environmental variable drive phylogenetic pattern?

343 There is significant variation among species in their responses to most environmental variables,
344 including most soil properties we measured, canopy shade, precipitation and minimum
345 temperature (Table 4). However, none of these variables show phylogenetic signal (last column
346 in Table 4). Therefore, no environmental variable we measured can explain the observed
347 phylogenetic pattern in community composition. When using PGLMM with presence/absence
348 data, we found that minimum temperature and soil pH, Ca, and Mn concentration show
349 phylogenetic patterns, with related species showing similar patterns in presence/absence in
350 response to these environmental variables (Appendix Table S3).

351 Discussion

352 We found phylogenetic patterns in community composition, in which phylogenetically related
353 species were more likely to occur in the same communities, yet we could not explain this pattern
354 completely using information among species' functional traits. When functional traits that
355 themselves showed phylogenetic signal were included in the phylogenetic linear mixed model

356 (PLMM) for log abundances of species in communities, that component of the residual variance
357 having phylogenetic covariances decreased by only 19%. The decrease in the phylogenetic
358 component of residual variation decreased 53% in the analyses of presence/absence data, yet
359 even this leaves considerable phylogenetic pattern in the unexplained variation in the
360 presence/absence of species among communities. Thus, even though we measured 15 functional
361 traits, including most of the "standard" functional traits used for studies on community structure,
362 we could not explain the phylogenetic patterns in community composition. This suggests that
363 there are either important functional traits that we have not measured, or that there are
364 phylogenetic processes unrelated to functional traits that we have not identified.

365 *Phylogenetic community composition*

366 The permutation tests using PSE and MPD, and the PLMM and PGLMM, found phylogenetic
367 attraction in community composition for the plant communities of central Wisconsin when all
368 152 species were included. However, only the PLMM and PGLMM found statistically
369 significant phylogenetic patterns when using the subset of 55 species that occurred in three or
370 more communities for which we had complete information on functional traits. Ives & Helmus
371 (2011) found that phylogenetic mixed models have greater power than metrics such as PSE and
372 MPD used with permutation tests to detect phylogenetic community structure. Simulations
373 tailored for our plant community data (Appendix Text S1) showed that the PLMM analyses
374 tended to have, if anything, incorrectly low Type I error rates, implying that our PLMM results
375 were unlikely to be the result of false positives. Therefore, we can conclude that closely related
376 species are more likely to co-occur than expected by chance in these sand plain communities.

377 *Functional traits and phylogenetic patterns*

We used our extensive database of functional traits to answer a key question in trait-based and phylogeny-based community ecology: Can information about functional traits explain phylogenetic patterns in community composition? Incorporating our measured functional traits into a PLMM for log species' abundance only reduced by 19% the phylogenetic component of residual variation in species composition (Table 3). For presence/absence data and a PGLMM, this increased to 53%, although there still remained considerable phylogenetic covariances. A possible explanation for this residual phylogenetic variation is some unknown historical process (Fig. 1B, IV). However, all of our sites are located within 100 km with each other (Li & Waller, 2015), and therefore historical biogeographical processes are unlikely to affect plant community composition.

We think the main source of phylogenetic patterns that were not explained by our measured functional traits is additional functional traits that we did not measure. Further analyses of presence/absence data with PGLMMs suggested that soil pH, Ca and Mn concentrations, and minimum temperature are the potential driving variables for the residual phylogenetic patterns (Appendix Table S3). However, none of the functional traits we measured are likely to explain how plants respond to these environmental variables. Measurements of additional functional traits that control plant performance related to these environmental variables (e.g., root structure, micorrhizal associations, frost tolerance, etc.) might be able to explain more of the phylogenetic pattern in community composition.

Implications for other studies

Our study has several implications for community ecology. First, it is clear that studying community composition should incorporate analyses of both phylogenetic structure and variation in functional traits. These data clearly complement each other in allowing sophisticated analyses

401 that can partition the amount of phylogenetic signal in community assembly that is associated
402 with functional trait variation (Fig. 1). Our results provide empirical support for the argument
403 that phylogeny can provide more information than a set of singular, discretely measured traits
404 (Vane-Wright *et al.* 1991; Cadotte *et al.* 2009). Although functional traits are necessary in order
405 to accurately infer processes from phylogenetic patterns (Kraft *et al.* 2007; Cavender-Bares *et al.*
406 2009), functional traits alone might not give a complete picture of community assembly. This
407 implies that both phylogenetic and trait-based data are needed if we are to understand how
408 measured and unmeasured traits, along with biogeography, influence community composition.

409 Second, our finding also provides an explanation for different (sometimes even opposite)
410 conclusions about community assembly processes based on comparisons between dispersion
411 patterns of phylogeny and traits (e.g. Swenson & Enquist, 2009; Kraft & Ackerly, 2010; Graham
412 *et al.* 2012). As functional traits and phylogeny provide overlapped and complementary
413 information about species and communities, different conclusions can be possible depend on the
414 functional traits used in the analyses. If the functional traits used provide similar information
415 with phylogeny, then conclusion about community assembly processes may be the same and vice
416 versa. Instead of compare and contrast dispersion patterns of traits and phylogeny, we argue that
417 integrating both into a model-based method such as PLMM used here. Model-based methods are
418 being increasingly applied in ecology because they are more interpretable, flexible, and powerful
419 than either null models or conventional algorithmic multivariate analyses (Warton *et al.* 2014).
420 With phylogenetic linear mixed models (PLMM), we not only detected phylogenetic patterns in
421 community composition, but also assessed whether these could be explained by functional traits.
422 Thus, both phylogenies and functional traits could be incorporated into the same statistical

423 model, with PLMMs (and PGLMMs) providing an integrated and quantitative framework for
424 analyzing ecological communities and predicting abundance of one taxon from others.

425 Finally, we can use phylogenetic analyses to suggest possible unmeasured functional traits that
426 underlie patterns in community composition and that therefore should be measured. If species
427 respond differently to an environmental variable, and if these differences are phylogenetic (i.e.,
428 related species respond to the environmental variable in similar ways), then there is likely to be a
429 functional trait or traits that underlie the response of species to this environmental variable. In
430 our study, the phylogenetic patterns in species responses to edaphic conditions like soil
431 chemistry highlighted our lack of data on the specific functional traits related to roots or
432 water/nutrient uptake. While this reveals that our study is incomplete, it also provides a valuable
433 lesson and demonstrates the power of the integrated PLMM approach.

434

435 Acknowledgements

436 We thank K. Cameron, R. Kriebel, M. Pace, D. Spalink, P. Li, and K. Sytsma for building and
437 providing the phylogeny we used in this study. This project was funded by US-NSF grant DEB-
438 1046355 and DEB-1240804;

439

440

References:

Adler, P.B., Fajardo, A., Kleinhesselink, A.R. & Kraft, N.J.B. (2013). Trait-based tests of coexistence mechanisms. *Ecology Letters*, 16, 1294–1306.

Baum, D.A. & Smith, S.D. (2012). *Tree Thinking: An Introduction to Phylogenetic Biology*. 1st Edition edition. Roberts; Company Publishers, Greenwood Village, Colo.

Cadotte, M.W., Cavender-Bares, J., Tilman, D. & Oakley, T.H. (2009). Using Phylogenetic, Functional and Trait Diversity to Understand Patterns of Plant Community Productivity. *PLoS ONE*, 4, e5695.

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

Donoghue, M.J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences*, 105, 11549–11555.

Emerson, B.C. & Gillespie, R.G. (2008). Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, 23, 619–630.

Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *The American Naturalist*, 160, 712–726.

Freilich, M.A. & Connolly, S.R. (2015). Phylogenetic community structure when competition and environmental filtering determine abundances. *Global Ecology and Biogeography*, n/a–n/a.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gotelli, N.J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606–2621.

Graham, C.H., Parra, J.L., Tinoco, B.A., Stiles, F.G. & McGuire, J.A. (2012). Untangling the influence of ecological and evolutionary factors on trait variation across hummingbird assemblages. *Ecology*, 93, S99–S111.

Graves, G.R. & Gotelli, N.J. (1993). Assembly of avian mixed-species flocks in Amazonia. *Proceedings of the National Academy of Sciences*, 90, 1388–1391.

Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007). Phylogenetic Measures of Biodiversity. *The American Naturalist*, 169, E68–E83.

Ho, L.S.T. & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, syu005.

Ives, A.R. & Garland, T. (2010). Phylogenetic Logistic Regression for Binary Dependent Variables. *Systematic Biology*, 59, 9–26.

Ives, A.R. & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, 81, 511–525.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H. & Ackerly, D.D.*et al.* (2010). Picante: R 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. *Ecological Monographs*, 80, 401–422.

Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait Evolution, Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The American Naturalist*, 170, 271–283.

Kucharik, C.J., Serbin, S.P., Vavrus, S., Hopkins, E.J. & Motew, M.M. (2010). Patterns of Climate Change Across Wisconsin From 1950 to 2006. *Physical Geography*, 31, 1–28.

Li, D. & Waller, D. (2015). Drivers of observed biotic homogenization in pine barrens of central Wisconsin. *Ecology*, 96, 1030–1041.

Losos, J.B. (1996). Phylogenetic Perspectives on Community Ecology. *Ecology*, 77, 1344–1354.

Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11, 995–1003.

Losos, J.B. (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in Comparative Biology. *The American Naturalist*, 177, 709–727.

Mayfield, M.M. & Levine, J.M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities: Phylogeny and coexistence. *Ecology Letters*, 13, 1085–1093.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.

Moen, D.S., Smith, S.A. & Wiens, J.J. (2009). Community Assembly Through Evolutionary Diversification and Dispersal in Middle American Treefrogs. *Evolution*, 63, 3228–3247.

Moles, A.T., Ackerly, D.D., Webb, C.O., Tweddle, J.C., Dickie, J.B. & Westoby, M. (2005). A Brief History of Seed Size. *Science*, 307, 576–580.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.

Pakeman, R.J. & Quested, H.M. (2007). Sampling plant functional traits: What proportion of the species need to be measured? *Applied Vegetation Science*, 10, 91–96.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H. & Jaureguiberry, P.*et al.* (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, 61, 167–234.

Ricklefs, R.E., Schluter, D. & others. (1993). Species diversity in ecological communities: Historical and geographical perspectives. *Species diversity in ecological communities: historical and geographical perspectives.*

Self, S.G. & Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, 82, 605–610.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, btu033.

Swenson, N.G. & Enquist, B.J. (2009). Opposing assembly mechanisms in a Neotropical dry forest: Implications for phylogenetic and functional community ecology. *Ecology*, 90, 2161–2170.

Vamosi, S.M., Heard, S.B., Vamosi, J.C. & Webb, C.O. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, 18, 572–592.

Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991). What to protect?Systematics and the agony of choice. *Biological Conservation*, 55, 235–254.

Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C. & Hummel, I.*et al.* (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892.

Warton, D.I., Foster, S.D., Death, G., Stoklosa, J. & Dunstan, P.K. (2014). Model-based thinking for community ecology. *Plant Ecology*, 1–14.

Webb, C.O. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *The American Naturalist*, 156, 145–155.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505.

Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, 199, 213–227.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B. & Cornell, H.V.*et al.* (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, 13, 1310–1324.

Tables:

Table 1 Estimated variance of random effects for the PLMM (equation 1) used to detect phylogenetic patterns in community composition.

PLMM	σ^2_a	σ^2_b	σ^2_c	σ^2_d	σ^2_e	$p(\sigma^2_c = 0)$	AIC
Phylogenetic attraction: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c\mathbf{S}_{spp}))$	0.9789	1.265×10^{-7}	6.503×10^{-3}	7.712×10^{-10}	0.5154	0.008	3900
Phylogenetic repulsion: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c (\mathbf{S}_{spp})^{-1})$	0.9785	1.715×10^{-7}	1.227×10^{-9}	2.282×10^{-2}	0.5308	0.496	3906
Non-nested model: *c* removed	0.9838	2.287×10^{-2}	-	8.828×10^{-7}	0.5306	-	3904

Table 2 Phylogenetic signal and site variation for each functional trait. P-values for the null hypothesis $\sigma_b^2 = 0$ (equation 2) implying no difference among sites in the effects of trait values on log abundance are given in the column labeled $p(\sigma_b^2 = 0)$. Functional traits with strong phylogenetic signal and $p(\sigma_b^2 = 0) < 0.05$ are considered to be important in explaining phylogenetic patterns.

Trait	Pagel's λ	*K*	$p(\sigma_b^2 = 0)$
Leaf specific area (SLA, *m2/kg*)	0.70**	0.26**	**0.002**
Leaf circularity (Dimensionless)	1.00***	0.71***	**0.001**
Leaf thickness (*mm*)	0.96***	1.80***	**0.032**
Life cycle (Annual or non-annual)	0.00	0.30	0.479
Growth habit (woody or non-woody)	1.08***	0.24**	0.500
Pollination mode (Biotic or abiotic)	0.00	0.08	0.500
Seed mass (*g/seed*)	1.00***	0.46	0.077
Leaf dry mass content (LDMC, %)	0.52	0.16	0.500
Stem dry mass content (SDMC, %)	0.00	0.14	0.500
Plant height (*cm*)	0.00	0.16	0.500
Leaf length (*cm*)	0.98***	0.66**	0.500
Leaf width (*cm*)	1.00***	0.57***	0.206
Leaf carbon content (%)	0.65***	0.26**	0.500
Leaf nitrogen content (%)	0.34	0.09	0.334
Wind dispersal (Yes or no)	1.15***	0.45***	0.196
Animal dispersal (Yes or no)	0.64***	0.28**	0.072
Unassisted dispersal (Yes or no)	0.00	0.15	0.500

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

571 Table 3 Reduction of the phylogenetic variance in community composition caused by the inclusion of

572 functional traits (equation 3).

Trait	σ_c^2 with traits	σ_c^2 without traits	$100 \times \sigma^2_{c\ (\text{with traits})}/\sigma^2_{c\ (\text{without traits})}$
SLA	0.006054	0.006461	6.30
Leaf circularity	0.006304	0.006415	1.73
Leaf thickness	0.006057	0.006459	6.22
SLA + circularity + thickness	0.005208	0.006437	19.09

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591 Table 4 Variation in the response of species abundances to environmental variables (equation 4).
592 Although 13/20 environmental variables generated variation in species composition among communities,
593 none of these showed phylogenetic signal in which related species responded more similarly to the
594 environmental variable.
595

Environmental variables	*P*-values σ_c^2 (no phylogenetic signal)	*P*-values for σ_d^2 (phylogenetic signal)
Minimum temperature	**< 0.001**	0.500
Precipitation	**< 0.001**	0.500
Canopy shade	**0.002**	0.500
Total exchange capacity	**0.002**	0.500
Organic matter	**0.001**	0.500
pH	**< 0.001**	0.500
N	**< 0.001**	0.500
P	**0.039**	0.500
Mg	**0.030**	0.500
K	**0.007**	0.500
Na	**< 0.001**	0.500
Mn	**< 0.001**	0.354
Ca	**< 0.001**	0.122
Clay	0.110	-
Silt	0.070	-
Sand	0.117	-
Fe	0.500	-
S	0.458	-
Zn	0.500	-
Al	0.500	-

596

Figures:

Figure 1: Schematic diagram of the conceptual framework of the study. (A) Evolution is the ultimate source of all traits, although only some traits have phylogenetic signal that reflects phylogenetic history (arrows 2, 4 and 5). Other traits do not (arrows 1 and 3), possibly because these traits evolve rapidly or experience convergent evolution. Community composition is determined by unmeasured and measured traits, and also by additional processes that could generate phylogenetic signal, such as biogeographical patterns in the distribution of species. Phylogenetic patterns in community composition can be generated from measured and unmeasured traits with phylogenetic signal (arrows 7 and 8), and by other phylogenetic processes (arrow 6). The question we address is how much of the phylogenetic signal in community composition can be explained by measured functional traits, and whether after accounting for these traits there is residual phylogenetic signal that could have been generated by unmeasured traits or other phylogenetic processes. (B) Traits and phylogeny contain overlapping and complementary information about how communities are assembled. Here, we focus on estimating the proportion of this overlapping information that the phylogeny contains (i.e., $\frac{i}{i+ii+iii}$). Note that we do not try to explain the proportion of overlapping information that functional traits contain (i.e., $\frac{I}{I+II+III}$) due to our inability to estimate the amount of information provided by unmeasured traits and hence estimate ($I + II + III$). Note: panel B is just a heuristic diagram.

Figure 2: Phylogeny and relative abundance of common plant species found in the pine barrens of central Wisconsin in 2012. The area of dots is proportional to abundances within each site.

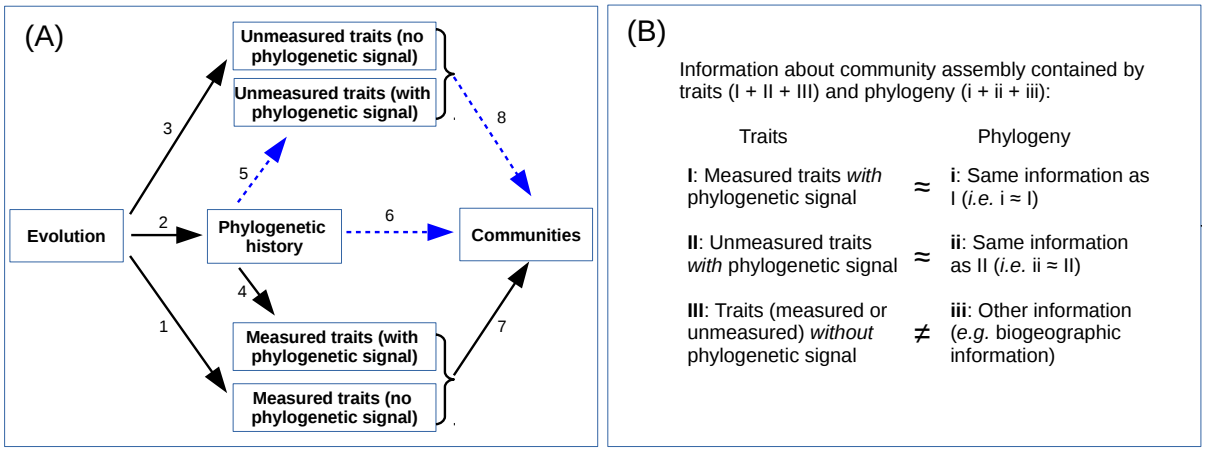
(A)
Unmeasured traits (no phylogenetic signal)
Unmeasured traits (with phylogenetic signal)
Evolution
Phylogenetic history
Communities
Measured traits (with phylogenetic signal)
Measured traits (no phylogenetic signal)
1
2
3
4
5
6
7
8
(B)
Information about community assembly contained by traits (I + II + III) and phylogeny (i + ii + iii):
Traits
Phylogeny
I: Measured traits with phylogenetic signal
≈
i: Same information as I (i.e. i ≈ I)
II: Unmeasured traits with phylogenetic signal
≈
ii: Same information as II (i.e. ii ≈ II)
III: Traits (measured or unmeasured) without phylogenetic signal
≠
iii: Other information (e.g. biogeographic information)

Figure 1

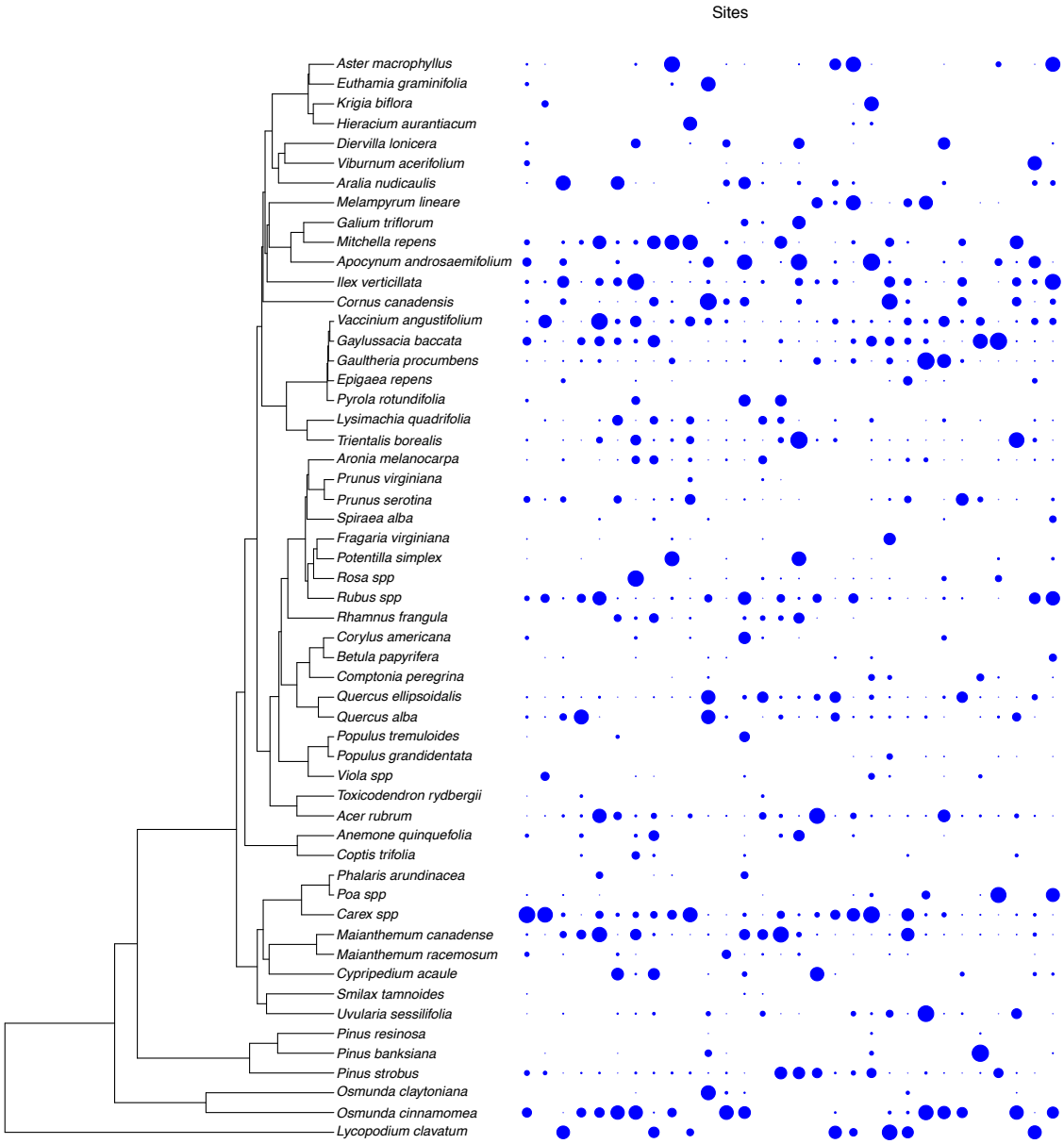
Sites
Aster macrophyllus
Euthamia graminifolia
Krigia biflora
Hieracium aurantiacum
Diervilla lonicera
Viburnum acerifolium
Aralia nudicaulis
Melampyrum lineare
Galium triflorum
Mitchella repens
Apocynum androsaemifolium
Ilex verticillata
Cornus canadensis
Vaccinium angustifolium
Gaylussacia baccata
Gaultheria procumbens
Epigaea repens
Pyrola rotundifolia
Lysimachia quadrifolia
Trientalis borealis
Aronia melanocarpa
Prunus virginiana
Prunus serotina
Spiraea alba
Fragaria virginiana
Potentilla simplex
Rosa spp
Rubus spp
Rhamnus frangula
Corylus americana
Betula papyrifera
Comptonia peregrina
Quercus ellipsoidalis
Quercus alba
Populus tremuloides
Populus grandidentata
Viola spp
Toxicodendron rydbergii
Acer rubrum
Anemone quinquefolia
Coptis trifolia
Phalaris arundinacea
Poa spp
Carex spp
Maianthemum canadense
Maianthemum racemosum
Cypripedium acaule
Smilax tamnoides
Uvularia sessilifolia
Pinus resinosa
Pinus banksiana
Pinus strobus
Osmunda claytoniana
Osmunda cinnamomea
Lycopodium clavatum

Figure 2

Appendix

Table S1 Estimated variance of random effects within phylogenetic generalized linear mixed models used to detect phylogenetic patterns (phylogenetic attraction and phylogenetic repulsion, estimated by σ_c^2).

Phylogenetic linear mixed models	σ_a^2	σ_b^2	σ_c^2	σ_d^2	$p(\sigma_c^2 = 0)$
Phylogenetic attraction:					
$\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha + a_{spp[i]} + b_{spp[i]} + c_i + d_{site[i]})$	2.8416	6.738×10^{-4}	0.0452	0.0110	<0.001
$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2\mathbf{S}_{spp}))$					
Phylogenetic repulsion:					
$\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha + a_{spp[i]} + b_{spp[i]} + c_i + d_{site[i]})$	3.1011	1.026×10^{-4}	0.0011	0.1936	0.5
$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 (\mathbf{S}_{spp})^{-1})$					
$\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha + a_{spp[i]} + b_{spp[i]} + d_{site[i]})$	2.8323	1.454×10^{-5}	-	0.1796	-

Table S2 Proportion of phylogenetic signal of species composition in communities explained by individual functional trait and multiple functional traits. With selected multiple functional traits, only about 19% percent of phylogenetic variation was explained, suggesting that phylogenies can provide additional information about community assembly beyond measured functional traits. See equation 3 in the Methods section for details about models.

Trait	σ_c^2 with trait	σ_c^2 without trait	$100 \times \sigma^2_{c\ (with\ trait)}/\sigma^2_{c\ (without\ trait)}$
SLA	0.036213	0.042060	13.90
Leaf circularity	0.034731	0.041094	15.48
Leaf thickness	0.024457	0.042004	41.77
SLA + circularity + thickness	0.019557	0.044395	53.17

Table S3 There are strong variations in species' relationships between their abundance and most environmental variables (*p* value of each environmental variable was presented in the *P*-values for variation column). However, none of these variations show phylogenetic signal. For environmental variable that has no strong variation in species' responses, no further test for phylogenetic signal of variation was conducted (thus "-" in the third column). P-values that are less than 0.05 are in bold.

Environmental variables	P-values for variation	P-values for phylogenetic signal of variation
Minimum temperature	**<0.001**	**0.002**
Precipitation	**<0.001**	0.500
Canopy shade	**0.001**	0.500
Total exchange capacity	0.149	-
Organic matter	0.161	-
pH	**0.005**	**<0.001**
N	0.052	-
P	0.343	-
Mg	0.500	-
K	0.206	-
Na	**0.004**	0.500
Mn	**<0.001**	**<0.001**
Ca	**0.012**	**<0.001**
Clay	0.431	-
Silt	0.494	-
Sand	0.500	-
Fe	0.379	-
S	0.500	-
Zn	0.500	-
Al	0.500	-

Text S1 Codes to compare p-values of null hypothesis $\sigma^2 = 0$ calculated from the $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution and paramatric bootstrap. The p-values based on the mixture Chi-squre districution are conservative.

```
# packages used
library(ape)  # for phylogeny reading
library(plyr)
library(MASS)
library(dplyr, quietly = TRUE)
library(pez)  # for commuityPGLMM function
library(parallel)  # for multiple cores parallel computation, not available
# for Windows operation system

# data: vegetation data, phylogeny
load("d_li_data.RData")
# select 20 sites and 20 species
test = veg.aggr.wide.1958[1:20, 1:20]
test1 = filter(veg.aggr.long.1958, sp %in% names(test), site %in% rownames(te
st))

# this function calculates log likelihood of the fitted model on observed
# data, then simulates data based on the fitted model, and fits model on
# simulated data and calculates the log likelihood of the fitted model; then
# calculates the p-value of the log likelihood of the fitted model on
# observed data based all simulated ones (i.e. parametric bootstrap); so we
# can compare the p-value get in this way (parametric bootstrap) with the
# one from the mixture Chi-square distribution.
q1_obs_sim = function(veg.long, phylo = pb.phylo, date = 1958, trans = NULL,
    nsim = 100, ncores = 5) {
    # transformation of freq
    if (!is.null(trans)) {
        if (trans == "log") {
            veg.long$Y <- log(veg.long$freq + 1)
        }

        if (trans == "asin") {
            veg.long <- group_by(veg.long, site) %>% mutate(Y = asin(sqrt((fr
eq + 1)/ifelse(date == 1958, 20 + 2, 50 + 2)))) %>% ungroup() %>%
                as.data.frame()
        }
    }

    veg.long$sp = as.factor(veg.long$sp)
    veg.long$site = as.factor(veg.long$site)
    nspp <- nlevels(veg.long$sp)
    nsite <- nlevels(veg.long$site)

    # Var-cov matrix for phylogeny
    phy <- drop.tip(phylo, tip = phylo$tip.label[!phylo$tip.label %in% levels
(veg.long$sp)])
```

```r
    Vphy <- vcv(phy)
    Vphy <- Vphy[order(phy$tip.label), order(phy$tip.label)]
    Vphy <- Vphy/max(Vphy)
    Vphy <- Vphy/det(Vphy)^(1/nspp)
    Vphy.inv = solve(Vphy)

    show(c(nlevels(veg.long$sp), Ntip(phy)))  # should be equal

    # random effect for site
    re.site <- list(1, site = veg.long$site, covar = diag(nsite))
    re.sp <- list(1, sp = veg.long$sp, covar = diag(nspp))
    re.sp.phy <- list(1, sp = veg.long$sp, covar = Vphy)
    # sp is nested within site, to test phylo attraction or repulsion
    re.nested.phy <- list(1, sp = veg.long$sp, covar = Vphy, site = veg.long$site)
    re.nested.rep <- list(1, sp = veg.long$sp, covar = Vphy.inv, site = veg.long$site)

    z <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = veg.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.site, re.nested.phy), REML = F, verbose = F, s2.init = 0.1)
    show(z$ss)
    z0 <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = veg.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.site),
        REML = F, verbose = F, s2.init = 0.1)
    z.rep <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = veg.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.site, re.nested.rep), REML = F, verbose = F, s2.init = 0.1)
    show(z.rep$ss)

    # observed ouput, p-values are get from Chisq approx.
    output_obs = data.frame(LRT_attract = (z$logLik - z0$logLik), p_attract = pchisq(2 * (z$logLik - z0$logLik), df = 1, lower.tail = F)/2, LRT_repulse = (z.rep$logLik - z0$logLik), p_repulse = pchisq(2 * (z.rep$logLik - z0$logLik), df = 1, lower.tail = F)/2, obs_sim = "obs")


    # the fitting model z0: log(y_i + 1) = alpha + a_spp[i] +
    # b_spp.phy[i] + c_site[i] + err[i]

    alpha = z0$B  # intercept, overall mean of all sp
    alpha.se = z0$B.se  # SE
    LRT_sim = mclapply(1:nsim, function(x) {
        # multi-cores
        set.seed(x)
        # z0$ss: random effects' SD for the cov matrix \sigma^2 * V, in order: [1]
        # sp with no phylo; [2] sp with Vphy; [3] site random effect
        a_spp = rnorm(nspp, 0, z0$ss[1])  # simulate a_spp
```

```
        # simulate b_spp.phy
        b_spp.phy = MASS::mvrnorm(1, mu = rep(0, nspp), Sigma = z0$ss[2] * Vphy)
        mu_spp = alpha + a_spp + b_spp.phy  # mean freq of sp
        c_site = rnorm(nsite, 0, z0$ss[3])  # site random
        mu_spp_site = rep(mu_spp, nsite) + rep(c_site, each = nspp)  # each sp at each site
        y_i = rnorm(nspp * nsite, mean = mu_spp_site, sd = alpha.se)  # include SE of intercept
        y_i_count = ceiling(exp(y_i) - 1)  # exp transf and round to positive interge
        test1_sim = data.frame(sp = names(mu_spp_site), site = rep(1:nsite,
            each = nspp), Y = y_i, freq = y_i_count)

        test1_sim$sp = as.factor(test1_sim$sp)
        test1_sim$site = as.factor(test1_sim$site)

        # refit models on simulated data random effect for site
        re.site.sim <- list(1, site = test1_sim$site, covar = diag(nsite))
        re.sp.sim <- list(1, sp = test1_sim$sp, covar = diag(nspp))
        re.sp.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy)
        # sp is nested within site
        re.nested.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy, site = test1_sim$site)
        re.nested.rep.sim <- list(1, sp = test1_sim$sp, covar = Vphy.inv, site = test1_sim$site)

        z_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian",
            sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim, re.sp.phy.sim, re.site.sim, re.nested.phy.sim), REML = F, verbose = F, s2.init = 0.1)
        # show(z_sim$ss)
        z0_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian", sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim,
                re.sp.phy.sim, re.site.sim), REML = F, verbose = F, s2.init = 0.1)
        # show(z0_sim$ss)
        z.rep_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian", sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim, re.sp.phy.sim, re.site.sim, re.nested.rep.sim), REML = F, verbose = F, s2.init = 0.1)
        # show(z.rep_sim$ss)

        # log lik of refitted models on simulated data
        data.frame(LRT_attract = (z_sim$logLik - z0_sim$logLik), LRT_repulse = (z.rep_sim$logLik - z0_sim$logLik))
    }, mc.cores = ncores)
```

```r
        # output results
    list(output_obs, ldply(LRT_sim))
}

qqq = q1_obs_sim(test1, trans = "log", nsim = 1000, ncores = 6)
saveRDS(qqq, "qqq.rds")

qqq = readRDS("qqq.rds")
qqq[[1]]

##   LRT_attract p_attract  LRT_repulse p_repulse obs_sim
## 1   0.3006013 0.2190598 -1.82719e-05       0.5     obs

head(qqq[[2]])

##   LRT_attract  LRT_repulse
## 1  -0.9611412 -15.68606765
## 2   0.1303866  -0.14712624
## 3  -2.9661583   0.06437319
## 4  -0.2152182  -1.91538503
## 5  -0.4204626   0.04073069
## 6  -1.3125998  -0.40523844

qqq[[2]]$obs_sim = "sim"
q1_sim = rbind(select(qqq[[1]], -p_attract, -p_repulse), qqq[[2]])
1 - (rank(q1_sim$LRT_attract)[1] + 1)/1001  # 0.12088 vs 0.219 from Chisq

## [1] 0.1208791

1 - (rank(q1_sim$LRT_repulse)[1] + 1)/1001  # 0.40959 vs 0.5 from Chisq

## [1] 0.4095904
```