

# 1 Can functional traits explain phylogenetic signal in the 2 composition of a plant community?

3 Daijiang Li<sup>1\*</sup>, Anthony R. Ives<sup>2</sup>, Donald M. Waller<sup>1</sup>

4  
5 <sup>1</sup>Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin,  
6 53706

7 <sup>2</sup>Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin,  
8 53706

9 \* Correspondence: [daijianglee@gmail.com](mailto:daijianglee@gmail.com); [arrives@wisc.edu](mailto:arrives@wisc.edu); [dmwaller@wisc.edu](mailto:dmwaller@wisc.edu)

10  
11 **Authorship:** DL and AI designed the study and performed the analyses. DL collected the  
12 vegetation and environmental data. DL and DW collected the functional traits data. DL wrote the  
13 first draft of the manuscript and all authors contributed substantially to revisions.

14 **Running title:** Functional traits and phylogeny

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

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

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

22 ***Abstract:***

23 Phylogeny-based and functional trait-based analyses are two principle ways to study community  
24 assembly and underlying ecological processes. In principle, knowing all information about  
25 species traits would make phylogenetic information redundant, at least that component of  
26 phylogenetic signal in the distribution of species among communities that is caused by  
27 phylogenetically related species sharing similar traits. In reality, phylogenies may contain more  
28 information than a set of singular, discretely measured traits because we cannot measure all  
29 species traits and may misjudge which are most important. The extent to which functional trait  
30 information makes phylogenetic information redundant, however, has not been explicitly studied  
31 with empirical data in community ecology. Here, we use phylogenetic linear mixed models to  
32 analyze community assembly of 55 understory plant species in 30 forest sites in central  
33 Wisconsin. These communities show strong phylogenetic attraction, yet variation among sites in  
34 20 environmental variables could not account for this pattern. Most of the 15 functional traits we  
35 measured had strong phylogenetic signal, but only three varied strongly among sites in ways that  
36 affected species' abundances. These three traits explained only 19% of variation in phylogenetic  
37 patterns of species co-occurrence. Thus, phylogenies appear to provide considerably more  
38 information about community assembly than the functional traits measured in this study,  
39 demonstrating the value of phylogeny in studying of community assembly processes even with  
40 abundant functional traits.

41

## 42 **Introduction**

43 Functional traits, arising as innovations through evolution, can capture essential aspects of  
44 species' morphology, ecophysiology, and life-history strategy (McGill *et al.* 2006; Violle *et al.*  
45 2007). Although closely related species can differ greatly in some functional traits due to rapid  
46 evolution or convergence (Losos, 2008, 2011), most functional traits show strong phylogenetic  
47 signal (Moles *et al.* 2005, Donoghue 2008, Freckleton *et al.* 2002; Webb *et al.* 2002). Functional  
48 traits, with or without phylogenetic signal, are known to influence the species composition of  
49 communities, thereby providing mechanistic links between fundamental ecological processes and  
50 community structure (McGill *et al.* 2006; Violle *et al.* 2007; Adler *et al.* 2013). Functional traits  
51 also provide a common currency that facilitates comparisons among species and across regions,  
52 allowing us to assess the generality of patterns and predictions in community ecology (McGill *et*  
53 *al.* 2006). This has led to a proliferation of studies using functional traits to understand  
54 community structure and composition. Functional trait-based approaches, however, are limited  
55 by the fact that it is impossible to measure all potentially important functional traits affecting the  
56 distribution of species across communities.

57 Even in the absence of functional trait information, it is still possible to infer the effects of  
58 (unmeasured) functional traits on community composition by investigating phylogenetic patterns  
59 in community composition. Phylogenies play an important role in community ecology by giving  
60 information about evolutionary relationships among species (Graves & Gotelli, 1993; Losos  
61 1996; Baum & Smith, 2012). Because phylogenetically related species often share similar  
62 functional trait values, we expect phylogenetically related species to co-occur more often in the  
63 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

87 “phylogenetic attraction” (phylogenetically related species more likely to occur in the same  
88 communities) and “phylogenetic repulsion.” If there is phylogenetic pattern, then it could be  
89 produced by measured functional traits that themselves have phylogenetic signal (Fig. 1, arrows  
90 2, 4, and 7), unmeasured functional traits with phylogenetic signal (Fig. 1, arrows 2, 5, and 8), or  
91 phylogenetic processes unrelated to functional traits (Fig. 1, arrow 6). We then develop a  
92 phylogenetic community mixed model incorporating the functional traits we measured to ask  
93 whether there is phylogenetic signal in the residual variation in community composition after the  
94 effects of these functional traits are removed. This analysis tests the hypothesis that we can  
95 explain all the phylogenetic pattern in community composition using measured functional traits.  
96 Finally, we use a phylogenetic community mixed model to investigate whether phylogenetically  
97 related species respond similarly to environmental gradients across the communities. The  
98 motivation for this analysis is to indirectly identify possible unmeasured functional traits that  
99 might play a role in community composition. Cases where phylogenetically related species  
100 respond similarly to an environmental gradient suggest that they share traits that confer similar  
101 tolerances to, or preferences for, specific environmental conditions; this analysis therefore may  
102 suggest what additional functional traits to measure in order to explain patterns in community  
103 composition.

104

## 105 **Methods**

### 106 **Data**

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

116 *Functional traits.* – For the 55 focal species, we measured 11 continuous and four categorical  
117 functional traits on at least 12 individuals from at least three populations using standard protocols  
118 (Pérez-Harguindeguy *et al.* 2013). Continuous traits include seed mass (g/seed), plant height  
119 (cm), specific leaf area (SLA,  $m^2/kg$ ), leaf dry matter content (LDMC, %), leaf circularity  
120 (dimensionless), leaf length (cm), leaf width (cm), leaf thickness (mm), leaf carbon concentration  
121 (%), leaf nitrogen concentration (%), and stem dry matter content (SDMC, %). We aggregated  
122 categories of each categorical trait into two levels: growth form (woody vs. non-woody), life  
123 cycle (annual vs. non-annual), and pollination mode (biotic vs. abiotic). We divided seed  
124 dispersal mode into three binary variables (wind dispersed vs. not, animal dispersed vs. not, and  
125 unassisted vs. assisted dispersal). Collectively, these functional traits, covering the leaf-height-  
126 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<sup>1</sup>). 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<sub>AXML</sub> (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.*

---

<sup>1</sup> 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.

147 2007) and mean phylogenetic distance (MPD, Webb, 2000). For a site with  $m$  individuals from  $n$   
148 species, if we let  $\mathbf{M}$  to be a  $n \times 1$  column vector containing abundance of all species and let  
149 variance-covariance matrix  $\mathbf{C}$  describing phylogenetic relationship of these species, we can  
150 calculate PSE as  $\frac{m \text{diag}(\mathbf{C})' \mathbf{M} - \mathbf{M}' \mathbf{C} \mathbf{M}}{m^2(1-1/n)}$  (the prime denotes transpose) and calculate MPD as  
151  $\frac{\sum_i^n \sum_j^n \mathbf{C}_{[i,j]} \mathbf{M}_i \mathbf{M}_j}{\sum_i^n \sum_j^n \mathbf{M}_i \mathbf{M}_j}$ ,  $i \neq j$ . PSE is scaled from zero to one, with one occurring when  $\mathbf{C}$  is the identity  
152 matrix and all species have equal abundance. We then calculated the mean of these 30 PSE  
153 ( $\overline{\text{PSE}}_{\text{obs}}$ ) and 30 MPD values ( $\overline{\text{MPD}}_{\text{obs}}$ ). To test for phylogenetic pattern using these metrics, we  
154 permuted species randomly among sites (SIM2 in Gotelli, 2000) 4999 times and then calculated  
155 metrics base on each permutation data set. This null permutation model retains the prevalence of  
156 each species across sites, but allows sites to change in species richness. If  $\overline{\text{PSE}}_{\text{obs}}$  or  $\overline{\text{MPD}}_{\text{obs}}$   
157 falls above (or below) 97.5% of the permutation values, then we infer a statistically significant  
158 phylogenetic repulsion (attraction). Using this null model where sites can vary in species  
159 richness is justified because under the null hypothesis of no phylogenetic signal, the values of  
160 PSE and MPD are independent of species richness at the sites.

161 In addition to testing the significance of the observed PSE and MPD values via null models, we  
162 fit a phylogenetic linear mixed model (PLMM) to test for phylogenetic community patterns in  
163 species abundances. A PLMM establishes a flexible statistical base to subsequently incorporate  
164 functional trait and environmental variables. Furthermore, PLMMs tend to have greater  
165 statistical power than other metrics examined using permutation tests (Ives & Helmus, 2011). To  
166 build the PLMM, let  $n$  be the number of species distributed among  $m$  sites. Letting  $Y$  be the  $mn \times$   
167 1 vector containing the abundance of species  $j$  ( $j = 1, \dots, n$ ) at site  $s$  ( $s = 1, \dots, m$ ), the PLMM is

168 
$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]} + e_i$$



$$\begin{aligned} 169 \quad a &\sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n) \\ 170 \quad b &\sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{S}_{\text{spp}}) \\ 171 \quad c &\sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \mathbf{S}_{\text{nested}})) \\ 172 \quad d &\sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{I}_m) \\ 173 \quad e &\sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \end{aligned} \tag{1}$$

174 We use the convention of multilevel models here (Gelman & Hill, 2007), with fixed and random  
175 effects given by Greek and Latin letters, respectively. The function  $\text{spp}[i]$  maps the observation  $i$   
176 in vector  $Y$  to the identity of the species (Gelman & Hill, 2007, p251-252), so  $i$  takes value from  
177 1 to  $mn$ . The intercept  $\alpha$  estimates the overall average log abundance of species across all sites.  
178 The following three random variables  $a_{\text{spp}[i]}$ ,  $b_{\text{spp}[i]}$  and  $c_i$  incorporate variation in abundance  
179 among plant species. Specifically, the  $n$  values of  $a_{\text{spp}[i]}$  give differences among species in mean  
180 log abundance across all sites and are assumed to be drawn independently from a Gaussian  
181 distribution with mean 0 and variance  $\sigma_a^2$ . The  $n$  values of  $b_{\text{spp}[i]}$  also give differences in mean log  
182 abundance across sites but are assumed to be drawn from a multivariate Gaussian distribution  
183 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  
184 next paragraph below), and the scalar  $\sigma_b^2$  dictates the overall strength of the phylogenetic signal.  
185 Thus,  $a_{\text{spp}[i]}$  and  $b_{\text{spp}[i]}$  together capture variation in mean species log abundances that is either  
186 unrelated to phylogeny or has phylogenetic signal. The random variable  $c_i$  accounts for  
187 covariance in the log abundances of plant species nested within sites (using the Kronecker  
188 product,  $\otimes$ ). Specifically,  $c_i$  assesses whether phylogenetically related plant species are more or  
189 less likely to co-occur at the same sites. Hence,  $c_i$  is used to measure either phylogenetic

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

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

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

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

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

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

## 233 **Can functional traits explain phylogenetic community composition?**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

273  $c_i$ , and  $d_{\text{site}[i]}$  are defined as in equation 1. The proportion of phylogenetic signal in species  
274 composition (measured by  $\sigma_c^2$ ) that this trait can explain is then assessed by comparing  $\sigma_c^2$   
275 between models with and without this trait. Finally, to evaluate the overall contribution of  
276 functional traits to the observed phylogenetic patterns, we built a multivariate version of equation  
277 3 which included all traits that have both phylogenetic signal *and* strong variation among sites.

## 278 **Does any environmental variable drive phylogenetic pattern?**

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

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

$$292 \quad \log(Y + 1) = \alpha + a_{\text{spp}[i]} + (\beta + b_{\text{spp}[i]})x_{\text{site}[i]} + e_i$$

$$293 \quad a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$$

$$\begin{aligned} 294 \quad & b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{I}_n) \\ 295 \quad & e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \end{aligned} \tag{4}$$

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

$$\begin{aligned} 302 \quad & \log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + (\beta + c_{\text{spp}[i]} + d_{\text{spp}[i]})x_{\text{site}[i]} + e_i \\ 303 \quad & a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n) \\ 304 \quad & b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{S}_{\text{spp}}) \\ 305 \quad & c \sim \text{Gaussian}(\mathbf{0}, \sigma_c^2 \mathbf{I}_n) \\ 306 \quad & d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{S}_{\text{spp}}) \\ 307 \quad & e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \end{aligned} \tag{5}$$

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

## 314 **Results**

### 315 **Phylogenetic community composition**

316 Phylogenetically related species co-occur more often than expected by chance in pine barrens  
317 communities in central Wisconsin (Fig. 2). Permutation tests including all 152 species show that  
318 closely related species are likely to co-occur within communities, as judged by either  
319 phylogenetic species evenness ( $\overline{\text{PSE}}_{\text{obs}} = 0.32, p = 0.03$ ) or mean phylogenetic distance ( $\overline{\text{MPD}}_{\text{obs}}$   
320  $= 338, p = 0.01$ ). Confining the analyses to the 55 focal species that occurred in at least three  
321 communities, the permutation tests did not statistically significant phylogenetic patterns ( $\overline{\text{PSE}}_{\text{obs}}$   
322  $= 0.27, p = 0.29$ ;  $\overline{\text{MPD}}_{\text{obs}} = 286, p = 0.17$ ), although the PLMM gave statistically significant  
323 phylogenetic patterns ( $p = 0.008$ ; Table 1). Converting the data to presence/absence among  
324 communities, the PGLMM also confirmed phylogenetic attraction pattern in these communities  
325 (Appendix Table S1). The stronger statistical results for the PLMM and PGLMM are not  
326 unexpected, given that model-based approaches are generally statistically more powerful than  
327 metric-based approaches (Ives & Helmus 2011).

### 328 **Can functional traits explain phylogenetic community composition?**

329 Most functional traits show strong phylogenetic signal (Table 2). Three traits – SLA, leaf  
330 circularity, and leaf thickness – significantly affected plant species' abundances among sites ( $\sigma_b^2 >$   
331  $0$ , equation 2, Table 2). In other words, different sites selected different species based on these  
332 three functional traits. Individually, SLA, leaf circularity, and leaf thickness reduced the  
333 phylogenetic variance by 6%, 2%, and 6% measured by the reduction in  $\sigma_c^2$  with the inclusion of  
334 these functional traits (equation 3). Including all three traits in the final model, the phylogenetic



335 variation  $\sigma_c^2$  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,  $\sigma_c^2$  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*

378 We used our extensive database of functional traits to answer a key question in trait-based and  
379 phylogeny-based community ecology: Can information about functional traits explain  
380 phylogenetic patterns in community composition? Incorporating our measured functional traits  
381 into a PLMM for log species' abundance only reduced by 19% the phylogenetic component of  
382 residual variation in species composition (Table 3). For presence/absence data and a PGLMM,  
383 this increased to 53%, although there still remained considerable phylogenetic covariances. A  
384 possible explanation for this residual phylogenetic variation is some unknown historical process  
385 (Fig. 1B, IV). However, all of our sites are located within 100 km with each other (Li & Waller,  
386 2015), and therefore historical biogeographical processes are unlikely to affect plant community  
387 composition.

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

### 397 *Implications for other studies*

398 Our study has several implications for community ecology. First, it is clear that studying  
399 community composition should incorporate analyses of both phylogenetic structure and variation  
400 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

## 441 **References:**

- 442 Adler, P.B., Fajardo, A., Kleinhesselink, A.R. & Kraft, N.J.B. (2013). Trait-based tests of  
443 coexistence mechanisms. *Ecology Letters*, 16, 1294–1306.
- 444 Baum, D.A. & Smith, S.D. (2012). *Tree Thinking: An Introduction to Phylogenetic Biology*. 1st  
445 Edition edition. Roberts; Company Publishers, Greenwood Village, Colo.
- 446 Cadotte, M.W., Cavender-Bares, J., Tilman, D. & Oakley, T.H. (2009). Using Phylogenetic,  
447 Functional and Trait Diversity to Understand Patterns of Plant Community Productivity. *PLoS*  
448 *ONE*, 4, e5695.
- 449 Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of  
450 community ecology and phylogenetic biology. *Ecology Letters*, 12, 693–715.
- 451 Donoghue, M.J. (2008). A phylogenetic perspective on the distribution of plant diversity.  
452 *Proceedings of the National Academy of Sciences*, 105, 11549–11555.
- 453 Emerson, B.C. & Gillespie, R.G. (2008). Phylogenetic analysis of community assembly and  
454 structure over space and time. *Trends in Ecology & Evolution*, 23, 619–630.
- 455 Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic Analysis and Comparative  
456 Data: A Test and Review of Evidence. *The American Naturalist*, 160, 712–726.
- 457 Freilich, M.A. & Connolly, S.R. (2015). Phylogenetic community structure when competition  
458 and environmental filtering determine abundances. *Global Ecology and Biogeography*, n/a–n/a.
- 459 Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*.  
460 Cambridge University Press.
- 461 Gotelli, N.J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606–  
462 2621.
- 463 Graham, C.H., Parra, J.L., Tinoco, B.A., Stiles, F.G. & McGuire, J.A. (2012). Untangling the  
464 influence of ecological and evolutionary factors on trait variation across hummingbird  
465 assemblages. *Ecology*, 93, S99–S111.

- 466 Graves, G.R. & Gotelli, N.J. (1993). Assembly of avian mixed-species flocks in Amazonia.  
467 *Proceedings of the National Academy of Sciences*, 90, 1388–1391.
- 468 Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007). Phylogenetic Measures of  
469 Biodiversity. *The American Naturalist*, 169, E68–E83.
- 470 Ho, L.S.T. & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait  
471 evolution models. *Systematic Biology*, syu005.
- 472 Ives, A.R. & Garland, T. (2010). Phylogenetic Logistic Regression for Binary Dependent  
473 Variables. *Systematic Biology*, 59, 9–26.
- 474 Ives, A.R. & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses  
475 of community structure. *Ecological Monographs*, 81, 511–525.
- 476 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H. & Ackerly, D.D. *et al.*  
477 (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.
- 478 Kraft, N.J.B. & Ackerly, D.D. (2010). Functional trait and phylogenetic tests of community  
479 assembly across spatial scales in an Amazonian forest. *Ecological Monographs*, 80, 401–422.
- 480 Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait Evolution,  
481 Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The*  
482 *American Naturalist*, 170, 271–283.
- 483 Kucharik, C.J., Serbin, S.P., Vavrus, S., Hopkins, E.J. & Motew, M.M. (2010). Patterns of  
484 Climate Change Across Wisconsin From 1950 to 2006. *Physical Geography*, 31, 1–28.
- 485 Li, D. & Waller, D. (2015). Drivers of observed biotic homogenization in pine barrens of central  
486 Wisconsin. *Ecology*, 96, 1030–1041.
- 487 Losos, J.B. (1996). Phylogenetic Perspectives on Community Ecology. *Ecology*, 77, 1344–1354.
- 488 Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship  
489 between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11,  
490 995–1003.

- 491 Losos, J.B. (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in  
492 Comparative Biology. *The American Naturalist*, 177, 709–727.
- 493 Mayfield, M.M. & Levine, J.M. (2010). Opposing effects of competitive exclusion on the  
494 phylogenetic structure of communities: Phylogeny and coexistence. *Ecology Letters*, 13, 1085–  
495 1093.
- 496 McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology  
497 from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.
- 498 Moen, D.S., Smith, S.A. & Wiens, J.J. (2009). Community Assembly Through Evolutionary  
499 Diversification and Dispersal in Middle American Treefrogs. *Evolution*, 63, 3228–3247.
- 500 Moles, A.T., Ackerly, D.D., Webb, C.O., Tweddle, J.C., Dickie, J.B. & Westoby, M. (2005). A  
501 Brief History of Seed Size. *Science*, 307, 576–580.
- 502 Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.
- 503 Pakeman, R.J. & Quested, H.M. (2007). Sampling plant functional traits: What proportion of the  
504 species need to be measured? *Applied Vegetation Science*, 10, 91–96.
- 505 Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H. & Jaureguiberry, P. *et al.*  
506 (2013). New handbook for standardised measurement of plant functional traits worldwide.  
507 *Australian Journal of Botany*, 61, 167–234.
- 508 Ricklefs, R.E., Schluter, D. & others. (1993). Species diversity in ecological communities:  
509 Historical and geographical perspectives. *Species diversity in ecological communities: historical*  
510 *and geographical perspectives*.
- 511 Self, S.G. & Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators  
512 and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical*  
513 *Association*, 82, 605–610.
- 514 Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis  
515 of Large Phylogenies. *Bioinformatics*, btu033.



- 516 Swenson, N.G. & Enquist, B.J. (2009). Opposing assembly mechanisms in a Neotropical dry  
517 forest: Implications for phylogenetic and functional community ecology. *Ecology*, 90, 2161–  
518 2170.
- 519 Vamosi, S.M., Heard, S.B., Vamosi, J.C. & Webb, C.O. (2009). Emerging patterns in the  
520 comparative analysis of phylogenetic community structure. *Molecular Ecology*, 18, 572–592.
- 521 Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991). What to protect? Systematics and  
522 the agony of choice. *Biological Conservation*, 55, 235–254.
- 523 Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C. & Hummel, I. *et al.* (2007). Let the  
524 concept of trait be functional! *Oikos*, 116, 882–892.
- 525 Warton, D.I., Foster, S.D., Death, G., Stoklosa, J. & Dunstan, P.K. (2014). Model-based thinking  
526 for community ecology. *Plant Ecology*, 1–14.
- 527 Webb, C.O. (2000). Exploring the phylogenetic structure of ecological communities: An  
528 example for rain forest trees. *The American Naturalist*, 156, 145–155.
- 529 Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002). Phylogenies and  
530 community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505.
- 531 Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*,  
532 199, 213–227.
- 533 Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B. & Cornell, H.V. *et al.*  
534 (2010). Niche conservatism as an emerging principle in ecology and conservation biology.  
535 *Ecology Letters*, 13, 1310–1324.
- 536
- 537
- 538
- 539
- 540

541 **Tables:**

542

543 Table 1 Estimated variance of random effects for the PLMM (equation 1) used to detect phylogenetic

544 patterns in community composition.

PLMM	$\sigma_a^2$	$\sigma_b^2$	$\sigma_c^2$	$\sigma_d^2$	$\sigma_e^2$	$p(\sigma_c^2 = 0)$	AIC
Phylogenetic attraction:							
$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \mathbf{S}_{\text{spp}}))$	0.9789	$1.265 \times 10^{-7}$	$6.503 \times 10^{-3}$	$7.712 \times 10^{-10}$	0.5154	0.008	3900
Phylogenetic repulsion:							
$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 (\mathbf{S}_{\text{spp}})^{-1}))$	0.9785	$1.715 \times 10^{-7}$	$1.227 \times 10^{-9}$	$2.282 \times 10^{-2}$	0.5308	0.496	3906
Non-nested model:							
$c$ removed	0.9838	$2.287 \times 10^{-2}$	-	$8.828 \times 10^{-7}$	0.5306	-	3904

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561 Table 2 Phylogenetic signal and site variation for each functional trait. P-values for the null hypothesis

562  $\sigma_b^2 = 0$  (equation 2) implying no difference among sites in the effects of trait values on log abundance are

563 given in the column labeled  $p(\sigma_b^2 = 0)$ . Functional traits with strong phylogenetic signal and  $p(\sigma_b^2 = 0) <$

564 0.05 are considered to be important in explaining phylogenetic patterns.

Trait	Pagel's $\lambda$	$K$	$p(\sigma_b^2 = 0)$
<b>Leaf specific area (SLA, <math>m^2/kg</math>)</b>	0.70**	0.26**	<b>0.002</b>
<b>Leaf circularity (Dimensionless)</b>	1.00***	0.71***	<b>0.001</b>
<b>Leaf thickness (<math>mm</math>)</b>	0.96***	1.80***	<b>0.032</b>
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

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

566

567

568

569

570

571 Table 3 Reduction of the phylogenetic variance in community composition caused by the inclusion of  
572 functional traits (equation 3).

Trait	$\sigma_c^2$ with traits	$\sigma_c^2$ without traits	$100 \times \sigma_c^2(\text{with traits})/\sigma_c^2(\text{without traits})$
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	<i>P</i> -values $\sigma_c^2$ (no phylogenetic signal)	<i>P</i> -values for $\sigma_d^2$ (phylogenetic signal)
Minimum temperature	< <b>0.001</b>	0.500
Precipitation	< <b>0.001</b>	0.500
Canopy shade	<b>0.002</b>	0.500
Total exchange capacity	<b>0.002</b>	0.500
Organic matter	<b>0.001</b>	0.500
pH	< <b>0.001</b>	0.500
N	< <b>0.001</b>	0.500
P	<b>0.039</b>	0.500
Mg	<b>0.030</b>	0.500
K	<b>0.007</b>	0.500
Na	< <b>0.001</b>	0.500
Mn	< <b>0.001</b>	0.354
Ca	< <b>0.001</b>	0.122
Clay	0.110	-
Silt	0.070	-
Sand	0.117	-
Fe	0.500	-
S	0.458	-
Zn	0.500	-
Al	0.500	-

596

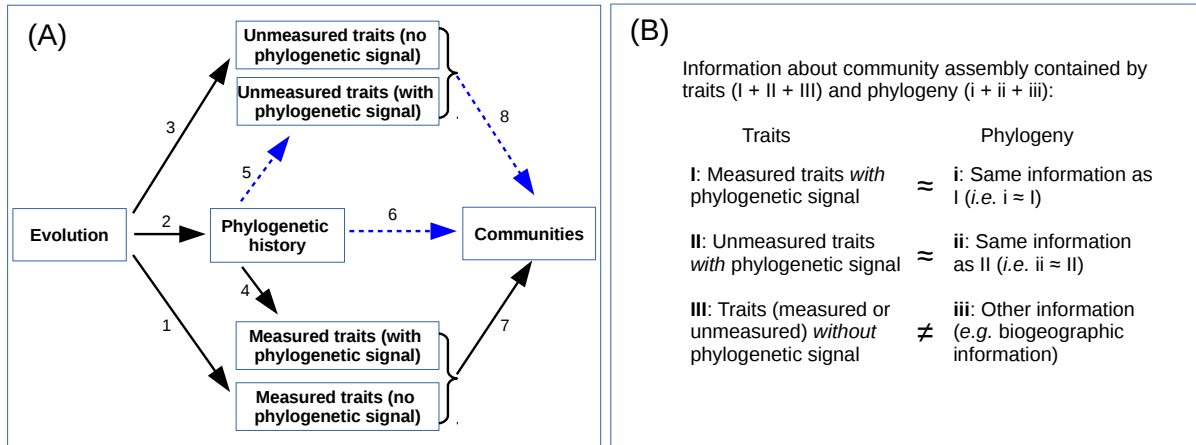
597 **Figures:**

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

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

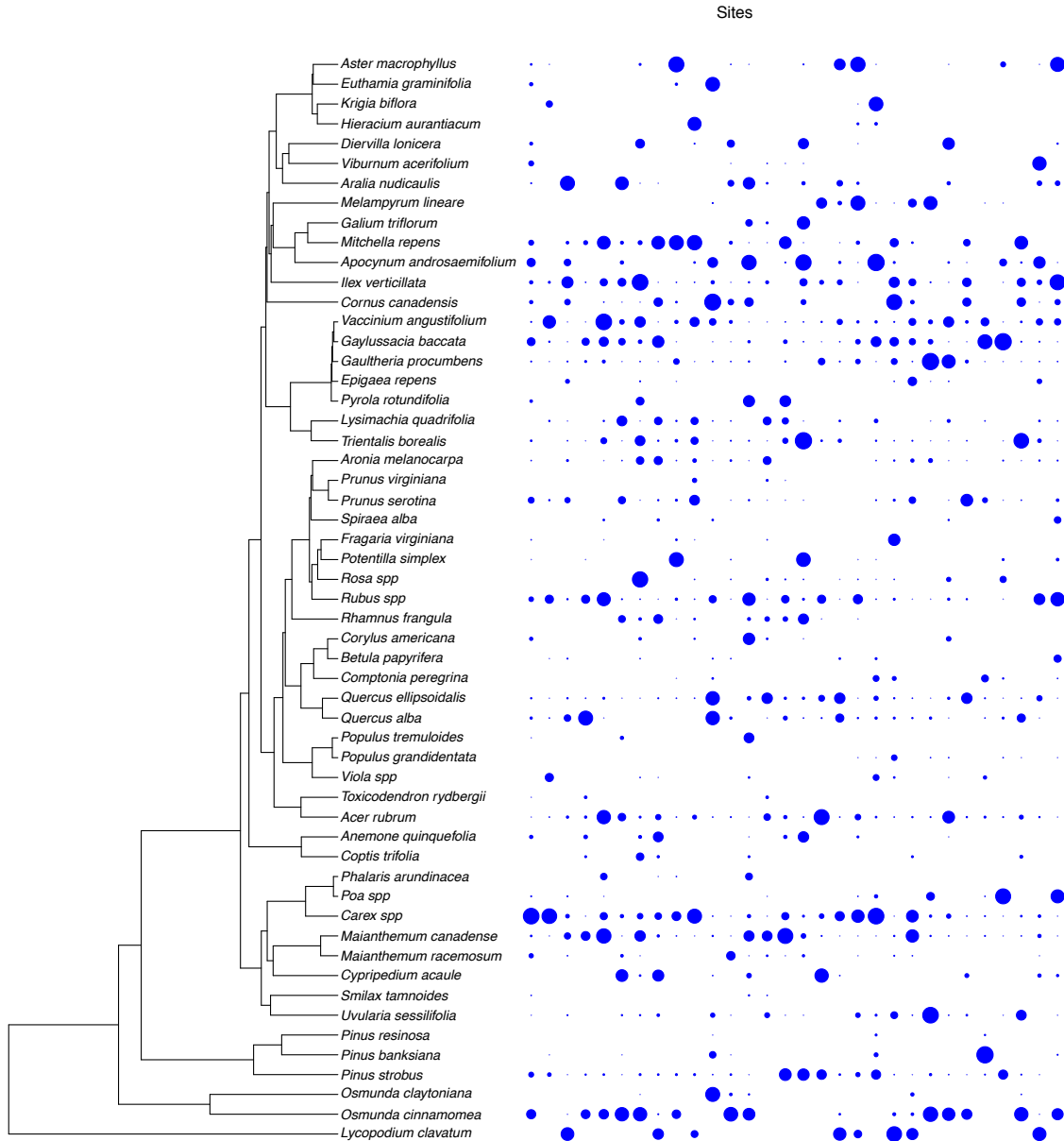
618

619



620

621 Figure 1



622

623 Figure 2

624

625

626

627



## 628 Appendix

629

630 Table S1 Estimated variance of random effects within phylogenetic generalized linear mixed models

631 used to detect phylogenetic patterns (phylogenetic attraction and phylogenetic repulsion, estimated by  $\sigma_c^2$ ).

632

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

633

634

635

636

637

638

639

640

641

642

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

648

Trait	$\sigma_c^2$ with trait	$\sigma_c^2$ without trait	$100 \times \sigma_c^2(\text{with trait})/\sigma_c^2(\text{without trait})$
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

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663 Table S3 There are strong variations in species' relationships between their abundance and most  
 664 environmental variables (*p* value of each environmental variable was presented in the *P*-values for  
 665 variation column). However, none of these variations show phylogenetic signal. For environmental  
 666 variable that has no strong variation in species' responses, no further test for phylogenetic signal of  
 667 variation was conducted (thus "-" in the third column). P-values that are less than 0.05 are in bold.  
 668

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

669

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

```
673 # packages used
674 library(ape) # for phylogeny reading
675 library(plyr)
676 library(MASS)
677 library(dplyr, quietly = TRUE)
678 library(pez) # for communityPGLMM function
679 library(parallel) # for multiple cores parallel computation, not available
680 # for Windows operation system

681 # data: vegetation data, phylogeny
682 load("d_li_data.RData")
683 # select 20 sites and 20 species
684 test = veg.aggr.wide.1958[1:20, 1:20]
685 test1 = filter(veg.aggr.long.1958, sp %in% names(test), site %in% rownames(test))
686
687
688 # this function calculates log likelihood of the fitted model on observed
689 # data, then simulates data based on the fitted model, and fits model on
690 # simulated data and calculates the log likelihood of the fitted model; then
691 # calculates the p-value of the log likelihood of the fitted model on
692 # observed data based all simulated ones (i.e. parametric bootstrap); so we
693 # can compare the p-value get in this way (parametric bootstrap) with the
694 # one from the mixture Chi-square distribution.
695 q1_obs_sim = function(veg.long, phylo = pb.phylo, date = 1958, trans = NULL,
696   nsim = 100, ncores = 5) {
697   # transformation of freq
698   if (!is.null(trans)) {
699     if (trans == "log") {
700       veg.long$Y <- log(veg.long$freq + 1)
701     }
702
703     if (trans == "asin") {
704       veg.long <- group_by(veg.long, site) %>% mutate(Y = asin(sqrt((freq + 1)/ifelse(date == 1958, 20 + 2, 50 + 2)))) %>% ungroup() %>%
705         as.data.frame()
706     }
707   }
708 }
709
710 veg.long$sp = as.factor(veg.long$sp)
711 veg.long$site = as.factor(veg.long$site)
712 nspp <- nlevels(veg.long$sp)
713 nsite <- nlevels(veg.long$site)
714
715 # Var-cov matrix for phylogeny
716 phy <- drop.tip(phylo, tip = phylo$tip.label[!phylo$tip.label %in% levels
717 (veg.long$sp)])
```

```
718 Vphy <- vcv(phy)
719 Vphy <- Vphy[order(phy$tip.label), order(phy$tip.label)]
720 Vphy <- Vphy/max(Vphy)
721 Vphy <- Vphy/det(Vphy)^(1/nspp)
722 Vphy.inv = solve(Vphy)
723
724 show(c(nlevels(veg.long$sp), Ntip(phy))) # should be equal
725
726 # random effect for site
727 re.site <- list(1, site = veg.long$site, covar = diag(nsites))
728 re.sp <- list(1, sp = veg.long$sp, covar = diag(nspp))
729 re.sp.phy <- list(1, sp = veg.long$sp, covar = Vphy)
730 # sp is nested within site, to test phylo attraction or repulsion
731 re.nested.phy <- list(1, sp = veg.long$sp, covar = Vphy, site = veg.long$
732 site)
733 re.nested.rep <- list(1, sp = veg.long$sp, covar = Vphy.inv, site = veg.l
734 ong$site)
735
736 z <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = veg
737 .long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.si
738 te, re.nested.phy), REML = F, verbose = F, s2.init = 0.1)
739 show(z$ss)
740 z0 <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = ve
741 g.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.s
742 ite),
743 REML = F, verbose = F, s2.init = 0.1)
744 z.rep <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp =
745 veg.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re
746 .site, re.nested.rep), REML = F, verbose = F, s2.init = 0.1)
747 show(z.rep$ss)
748
749 # observed output, p-values are get from Chisq approx.
750 output_obs = data.frame(LRT_attract = (z$logLik - z0$logLik), p_attract =
751 pchisq(2 * (z$logLik - z0$logLik), df = 1, lower.tail = F)/2, LRT_repulse = (
752 z.rep$logLik - z0$logLik), p_repulse = pchisq(2 * (z.rep$logLik - z0$logLik),
753 df = 1, lower.tail = F)/2, obs_sim = "obs")
754
755 # the fitting model z0:  $\log(y_i + 1) = \alpha + a_{spp}[i] +$ 
756 #  $b_{spp.phy}[i] + c_{site}[i] + err[i]$ 
757
758 alpha = z0$B # intercept, overall mean of all sp
759 alpha.se = z0$B.se # SE
760 LRT_sim = mclapply(1:nsim, function(x) {
761 # multi-cores
762 set.seed(x)
763 # z0$ss: random effects' SD for the cov matrix  $\sigma^2 * V$ , in order
764 : [1]
765 # sp with no phylo; [2] sp with Vphy; [3] site random effect
766 a_spp = rnorm(nspp, 0, z0$ss[1]) # simulate a_spp
```

```
766     # simulate b_spp.phy
767     b_spp.phy = MASS::mvrnorm(1, mu = rep(0, nspp), Sigma = z0$ss[2] * Vp
768 hy)
769     mu_spp = alpha + a_spp + b_spp.phy # mean freq of sp
770     c_site = rnorm(nsites, 0, z0$ss[3]) # site random
771     mu_spp_site = rep(mu_spp, nsites) + rep(c_site, each = nspp) # each s
772 p at each site
773     y_i = rnorm(nspp * nsites, mean = mu_spp_site, sd = alpha.se) # inclu
774 de SE of intercept
775     y_i_count = ceiling(exp(y_i) - 1) # exp transf and round to positive
776 interge
777     test1_sim = data.frame(sp = names(mu_spp_site), site = rep(1:nsites,
778         each = nspp), Y = y_i, freq = y_i_count)
779
780     test1_sim$sp = as.factor(test1_sim$sp)
781     test1_sim$site = as.factor(test1_sim$site)
782
783     # refit models on simulated data random effect for site
784     re.site.sim <- list(1, site = test1_sim$site, covar = diag(nsites))
785     re.sp.sim <- list(1, sp = test1_sim$sp, covar = diag(nspp))
786     re.sp.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy)
787     # sp is nested within site
788     re.nested.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy, site =
789 test1_sim$site)
790     re.nested.rep.sim <- list(1, sp = test1_sim$sp, covar = Vphy.inv, sit
791 e = test1_sim$site)
792
793     z_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian",
794         sp = test1_sim$sp, site = test1_sim$site, random.effects = list(r
795 e.sp.sim, re.sp.phy.sim, re.site.sim, re.nested.phy.sim), REML = F, verbose =
796 F, s2.init = 0.1)
797     # show(z_sim$ss)
798     z0_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian"
799 , sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim,
800     re.sp.phy.sim, re.site.sim), REML = F, verbose = F, s2.init =
801 0.1)
802
803     # show(z0_sim$ss)
804     z.rep_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussi
805 an", sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.si
806 m, re.sp.phy.sim, re.site.sim, re.nested.rep.sim), REML = F, verbose = F, s2.
807 init = 0.1)
808     # show(z.rep_sim$ss)
809
810     # Log Lik of refitted models on simulated data
811     data.frame(LRT_attract = (z_sim$logLik - z0_sim$logLik), LRT_repulse
812 = (z.rep_sim$logLik - z0_sim$logLik))
813     }, mc.cores = ncores)
814
```

```
815     # output results
816     list(output_obs, ldply(LRT_sim))
817 }
818
819 qqq = q1_obs_sim(test1, trans = "log", nsim = 1000, ncores = 6)
820 saverDS(qqq, "qqq.rds")
821
822 qqq = readRDS("qqq.rds")
823 qqq[[1]]
824
825 ##   LRT_attract p_attract  LRT_repulse p_repulse obs_sim
826 ## 1  0.3006013 0.2190598 -1.82719e-05      0.5      obs
827
828 head(qqq[[2]])
829
830 ##   LRT_attract  LRT_repulse
831 ## 1  -0.9611412 -15.68606765
832 ## 2   0.1303866  -0.14712624
833 ## 3  -2.9661583   0.06437319
834 ## 4  -0.2152182  -1.91538503
835 ## 5  -0.4204626   0.04073069
836 ## 6  -1.3125998  -0.40523844
837
838 qqq[[2]]$obs_sim = "sim"
839 q1_sim = rbind(select(qqq[[1]], -p_attract, -p_repulse), qqq[[2]])
840 1 - (rank(q1_sim$LRT_attract)[1] + 1)/1001 # 0.12088 vs 0.219 from Chisq
841
842 ## [1] 0.1208791
843
844 1 - (rank(q1_sim$LRT_repulse)[1] + 1)/1001 # 0.40959 vs 0.5 from Chisq
845
846 ## [1] 0.4095904
847
848
849
850
```