Can functional traits account for phylogenetic signal in community composition?

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

[1]Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, WI 53706, USA; [2]Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, WI 53706, USA

Author for correspondence:
Daijiang Li
Tel: +1 608 265 2191
Email: daijianglee@gmail.com

Summary

- Phylogenetic and functional trait-based analyses inform our understanding of community composition, yet methods for quantifying the overlap in information derived from functional traits and phylogenies remain underdeveloped. Does adding traits to analyses of community composition reduce the phylogenetic signal in the residual variation? If not, then measured functional traits alone may be insufficient to explain community assembly.
- We propose a general statistical framework to quantify the proportion of phylogenetic pattern in community composition that remains after including measured functional traits. We then illustrate the framework with applications to two empirical data sets.
- Both data sets showed strong phylogenetic attraction, with related species likely to co-occur in the same communities. In one data set, including traits eliminated all phylogenetic signals in the residual variation of both abundance and presence/absence patterns. In the second data set, including traits reduced phylogenetic signal in residuals by 25% and 98% for abundance and presence/absence data, respectively.
- Our framework provides an explicit way to estimate how much phylogenetic community pattern remains in the residual variation after including measured functional traits. Knowing that functional traits account for most of the phylogenetic pattern should provide confidence that important traits for phylogenetic community structure have been identified. Conversely, knowing that there is unexplained residual phylogenetic information should spur the search for additional functional traits or other processes underlying community assembly.

Introduction

Functional traits, arising as innovations through evolution, 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 as a result of rapid evolution or ecological convergence (Losos, 2008, 2011), many functional traits are conserved enough to show strong phylogenetic signal (Freckleton *et al.*, 2002; Webb *et al.*, 2002; Moles *et al.*, 2005; Donoghue, 2008). Functional traits, with or without phylogenetic signal, affect how environmental conditions filter species into communities and how species compete, mechanistically linking fundamental ecological processes to 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 led to a proliferation of studies using functional traits to understand community composition. Functional trait-based approaches, however, are limited by the fact that it is impossible to measure all potentially important functional traits that affect the distribution of species.

Phylogenies play an important role in community ecology by providing information about evolutionary relationships among species (Graves & Gotelli, 1993; Losos, 1996; Cavender-Bares *et al.*, 2006; Baum & Smith, 2012). Because phylogenetically related species often have 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 the similar traits that phylogenetically related species have cause them to compete strongly, then closely related species may be less likely to co-occur. These and other processes relating functional traits to community composition often lead to phylogenetic signatures in how species are distributed among communities (Webb *et al.*, 2002).
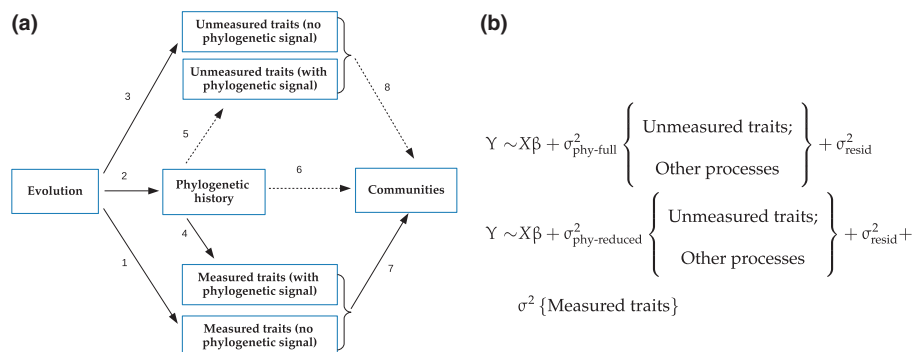
We expect the information provided by functional traits and the information provided by phylogenetic analyses to overlap (Vane-Wright *et al.*, 1991; Cavender-Bares *et al.*, 2004; Cadotte *et al.*, 2009; Ives & Helmus, 2011). We can test this expectation by incorporating functional trait information into analyses of community composition and asking whether there are residual

phylogenetic patterns that remain unaccounted for by the functional traits. If so, this would suggest one of two things. First, there may be additional functional traits that have not been measured and incorporated into the analysis. Second, there may be phylogenetically correlated factors aside from functional traits that affect community structure; for example, phylogenetic patterns could be generated if there is immigration from another region that contains phylogenetically related species, or if stochastic processes leave a phylogenetic signature (Ricklefs & Schluter, 1993; Hubbell, 2001; Leibold *et al.*, 2010). Thus, when residual phylogenetic patterns remain in analyses that incorporate functional trait information, we have evidence that there are additional factors affecting community assembly.

Here, we present a general three-step statistical framework to test for residual phylogenetic signal in community composition that remains unaccounted for after the inclusion of measured functional traits (Fig. 1). The first step is to test for phylogenetic pattern, either 'phylogenetic attraction' (phylogenetically related species more likely to occur in the same communities) or 'phylogenetic repulsion', using a generalized mixed model (Bolker *et al.*, 2009) that incorporates phylogenetic correlations in the distributions of species (Ives & Helmus, 2011). If there is phylogenetic pattern, then it could be produced by measured functional traits that themselves have phylogenetic signal (Fig. 1a, arrows 2, 4, and 7), unmeasured functional traits with phylogenetic signal (Fig. 1a, arrows 2, 5, and 8), or phylogenetic processes unrelated to functional traits (Fig. 1a, arrow 6). The second step first adds measured functional traits to the model, testing to see if they help to explain the distribution of species among communities. After incorporating these explanatory traits, we then ask whether phylogenetic signal remains in the residual variation in community composition (Fig. 1b). If there is residual phylogenetic pattern, we continue to add traits to see if these account for this signal. Thus, the second step investigates the extent to which we can account for the phylogenetic pattern in community composition using the measured functional traits. If phylogenetic patterns

remain even after incorporating all selected measured traits, we can then apply a third step. This uses environmental data to test whether phylogenetically related species respond similarly to environmental gradients across communities. Such parallel species' responses to environmental gradients allow us to indirectly identify possible unmeasured functional traits that could play a role in community assembly (Fig. 1a, arrow 8, unmeasured traits with phylogenetic signal). In cases where phylogenetically related species respond similarly to an environmental gradient, species presumably share traits that confer similar tolerances to, or preferences for, specific environmental conditions. Although these three steps employ existing statistical methodology, the specific models have not previously been incorporated into a general framework to investigate the overlap between information provided by functional traits and phylogenies. To illustrate this framework and explore its utility, we analyse two community data sets rich in information on traits, phylogenetic relationships, and environmental variables.

Other studies have simultaneously investigated both trait-based and phylogenetic patterns in community composition (e.g. Cavender-Bares *et al.*, 2006; Swenson & Enquist, 2009). As many functional traits can affect community composition, it is important to analyse several traits when testing whether residual phylogenetic patterns exist after incorporating traits. Previous studies investigating multiple functional traits generally use one metric to measure multidimensional differences between pairs of species in functional trait-space and a second metric to measure phylogenetic differences between pairs of species. They then compare the explanatory power of these two metrics. Recently, Cadotte *et al.* (2013) created a synthetic metric that combines trait and phylogenetic distances by varying a weighting parameter. This allowed them to analyse differences between communities using models that varied this metric to span the gradient from traits alone to phylogeny alone. Our approach differs from these metric-based approaches in several ways. First, rather than reducing information on traits to a one-dimensional gradient of

(a)

(b)

$$\Upsilon \sim X\beta + \sigma^2_{\text{phy-full}} \left\{ \begin{array}{l} \text{Unmeasured traits;} \\[1em] \text{Other processes} \end{array} \right\} + \sigma^2_{\text{resid}}$$

$$\Upsilon \sim X\beta + \sigma^2_{\text{phy-reduced}} \left\{ \begin{array}{l} \text{Unmeasured traits;} \\[1em] \text{Other processes} \end{array} \right\} + \sigma^2_{\text{resid}} +$$

$$\sigma^2 \{\text{Measured traits}\}$$

Fig. 1 Diagram of the conceptual framework of the study. (a) Some traits have phylogenetic signal that reflects phylogenetic history (arrows 2, 4 and 5) while other traits do not (arrows 1 and 3), possibly because these traits evolve rapidly or experience convergent evolution. Phylogenetic patterns in community composition can be generated by measured and unmeasured traits with phylogenetic signal (arrows 7 and 8), and by other phylogenetic processes such as biogeographical patterns in species distributions (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 there is residual phylogenetic signal that could have been generated by unmeasured traits or other phylogenetic processes. (b) Traits and phylogeny contain overlapping information about how communities are assembled. We estimate the proportion of this overlapping information by calculating the changes in the strength of community phylogenetic patterns without ($\sigma^2_{\text{phy-full}}$) and with ($\sigma^2_{\text{phy-reduced}}$) measured functional traits in the models.

differences among species, our modelling approach retains specific values for all traits for each species. This is consistent with typical phylogenetic analyses (Felsenstein, 1985; Harvey & Pagel, 1991; Garland *et al.*, 2005) that treat species' traits as independent variables (the 'mean' part of the model), while phylogenetic patterns are assessed as covariances in the residuals (the 'variance' part of the model). Second, because our model-based approach retains more information than metric-based approaches, it is expected to have greater statistical power (Ives & Helmus, 2011). Third, although metric-based approaches can effectively identify trait-based and phylogenetic patterns in specific communities, they do not provide any overall test for overlap between information derived from measured functional traits and phylogenies. Providing such a test is the focus of our analyses. If measured traits together reduce all phylogenetic signal in the residual variation in species distribution patterns, we can conclude that these traits (or other traits that are strongly correlated) account for phylogenetic patterns in community composition. Conversely, if measured traits do not reduce most of the residual phylogenetic patterns, then we can infer that either additional traits or additional biogeographical processes need to be investigated.

Materials and Methods

We apply our three-step statistical framework to both species abundance and presence/absence (incidence) data drawn from two illustrative empirical studies. In the main text we present results for abundance data because abundance data typically provide more information about community assembly (Freilich & Connolly, 2015). In the Supporting Information we present parallel results for the presence/absence data.

Phylogenetic community composition

Our first step tests for phylogenetic community structure in species abundances without including environmental or functional trait information by applying a phylogenetic linear mixed model (PLMM). 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:

$$\log(Y_i + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]} + e_i$$
$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$$
$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \Sigma_{\text{spp}})$$
$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \Sigma_{\text{nested}})) \qquad \text{Eqn 1}$$
$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{I}_m)$$
$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn})$$

Here we use the convention of multilevel models, with fixed and random effects given by Greek and Latin letters, respectively (Gelman & Hill, 2007). The function spp[i] maps the observation i in vector Y to the identity of the species, so i takes values from 1 to mn (Gelman & Hill, 2007, 251–252). 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, $a_{\text{spp}[i]}$ gives differences among species in mean log abundance across all sites, with differences among the n species assumed to be drawn independently from a Gaussian distribution with mean 0 and variance σ_a^2. $b_{\text{spp}[i]}$ also gives differences in mean log abundance among species across sites, but differences among species are assumed to be drawn from a multivariate Gaussian distribution with covariance matrix $\sigma_b^2 \Sigma_{\text{spp}}$, where the $n \times n$ matrix Σ_{spp} is derived from the phylogeny, and the scalar σ_b^2 dictates the overall strength of the phylogenetic signal (see next paragraph). 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, kron). Specifically, c_i assesses whether phylogenetically related plant species are more or less likely to co-occur at the same sites. Hence, c_i measures the overall strength of phylogenetic attraction or repulsion, and it is the key term we are interested in. Random effect $d_{\text{site}[i]}$ accounts for site-specific variation in the mean log abundances of all species among the m sites, with these m values assumed to be distributed by a Gaussian distribution with variance σ_d^2. Finally, e_i captures the residual variance σ_e^2.

We derive the phylogenetic covariance matrix Σ_{spp} from the assumption of Brownian motion evolution. If a continuous-valued trait evolves up a phylogenetic tree with a constant probability of slight increases or decreases, 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 & Hansen, 1997). This gives a direct way to convert the phylogeny into a hypothesis about the covariance matrix. For the mean abundances of species among sites given by $a_{\text{spp}[i]} + b_{\text{spp}[i]}$, the variance among species given by $\sigma_a^2 \mathbf{I} + \sigma_b^2 \Sigma_{\text{spp}}$ is equivalent to the model of evolution proposed by Pagel (1999). For the assessment of phylogenetic attraction within sites, c_i, we use $\Sigma_{\text{nested}} = \Sigma_{\text{spp}}$. For phylogenetic repulsion, we use the matrix inverse of Σ_{spp}, $\Sigma_{\text{nested}} = (\Sigma_{\text{spp}})^{-1}$. Theoretical justification for $\Sigma_{\text{nested}} = (\Sigma_{\text{spp}})^{-1}$ comes from a model of competition among community members (Appendix A of Ives & Helmus, 2011). Briefly, if the strength of competition between species is given by Σ_{spp}, as expected if closely related species exploit similar resources, then the relative abundances of species will have covariance matrix $(\Sigma_{\text{spp}})^{-1}$.

Eqn 1 is the same as model I in Ives & Helmus (2011) except that 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 Eqn 1 with Eqn 2 (later) 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 led to almost identical estimates of phylogenetic signal (estimates of σ_c^2), and so our treatment of $a_{\text{spp}[i]}$ and $b_{\text{spp}[i]}$ as random effects does not change our conclusions.

Statistical significance of the variance terms σ^2 can be determined using a likelihood ratio test. Because the null hypothesis $\sigma^2 = 0$ is on the boundary of the parameter space (σ^2 cannot be negative), we use the $0.5 \chi_0^2 + 0.5 \chi_1^2$ mixture distribution for significance tests (Self & Liang, 1987). χ_0^2 represents a distribution with a point mass at 0, so the P-values given by $0.5 \chi_0^2 + 0.5 \chi_1^2$ 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 (give higher P-values) than those derived from a parametric bootstrap (Supporting Information Notes S1).

Are there residual phylogenetic patterns after incorporating functional traits?

After the detection of phylogenetic patterns, our second step asks how much of the phylogenetic patterns identified in step 1 remains after including functional traits in the model. We selected measured functional traits to be incorporated into the model with two aims. First, we wanted to include functional traits that are important in explaining species composition of communities, fulfilling one of the most common goals of community ecology. Second, we wanted to include additional functional traits that can further reduce phylogenetic patterns in the residual variation, fulfilling the goal of this study.

We incorporated functional traits into a statistical model of community composition using the model:

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

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$$
$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \Sigma_{\text{spp}})$$
$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \Sigma_{\text{nested}})) \qquad \text{Eqn 2}$$
$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{I}_m)$$
$$f \sim \text{Gaussian}(\mathbf{0}, \sigma_f^2 \mathbf{I}_m)$$
$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn})$$

This model is the same as Eqn 1, except that it includes the values of a single functional trait, $t_{\text{spp}[i]}$. The fixed and random effects terms $(\beta + f_{\text{site}[i]}) t_{\text{spp}[i]}$ allow the effect of trait $t_{\text{spp}[i]}$ to vary among sites, with the fixed term β giving the mean slope of the relationship between log abundance and $t_{\text{spp}[i]}$ across all sites, and the random term $f_{\text{site}[i]}$ giving variation in this slope among sites. The proportion of phylogenetic signal in species composition (estimated by σ_c^2) reduced by the incorporation of trait $t_{\text{spp}[i]}$ is assessed by comparing σ_c^2 between models with and without this trait in the product $f_{\text{site}[i]} t_{\text{spp}[i]}$. To assess more than one functional trait, we built a multivariate version of Eqn 2; each trait is incorporated in the same way by replicating the terms $(\beta + f_{\text{site}[i]}) t_{\text{spp}[i]}$.

To identify functional traits that are important in explaining species composition of communities, we performed model selection using the Akaike information criterion (AIC). Because of the large number of possible models, we used forward selection as proposed by Jamil *et al.* (2012). We did not use backward

selection, because backward selection is often unfeasible because of mathematical convergence problems if the full model has many variables. To select traits that explain patterns of community composition, we started with the model containing only species and sites as random terms and excluding phylogenetic variances (Eqn 1 without terms $b_{\text{spp}[i]}$ and c_i). We then selected traits as fixed terms $\beta t_{\text{spp}[i]}$ (Eqn 2 without terms $b_{\text{spp}[i]}$ and c_i). These fixed terms represent the average responses across all sites; thus, when a fixed term is included, it implies that, on average across all sites, species may have high or low abundance depending on their trait value. We then selected traits as random terms including also their fixed terms $(\beta + f_{\text{site}[i]}) t_{\text{spp}[i]}$. The random terms represent differences among sites in how species traits affect species abundances. Thus, if $f_{\text{site}[i]}$ is included, then some sites might have high abundances of species with a given trait value, while other sites would have low abundances of species with the same trait value. We always included the fixed term for a trait when including it as a random term, because we did not want to force the random effects to have a mean of zero. After all traits that improved the fit of the model (either as fixed terms or as fixed and random terms) were selected, we selected and added the phylogenetic terms $b_{\text{spp}[i]}$ and c_i (Eqn 2). This process should give the lowest AIC phylogenetic model, which we checked by removing each of the random and fixed effects while retaining the phylogenetic terms $b_{\text{spp}[i]}$ and c_i in the model. To ask whether the addition of more traits could reduce the residual phylogenetic pattern (σ_c^2) any further, regardless of whether this decreased the fit of the model (increased AIC), we then added each remaining trait as fixed and random terms. We tried all traits regardless of whether they were correlated with each other, because our goal was to test the overall phylogenetic pattern in the residual variation after including functional traits rather than to separate the contribution of each functional trait. For this purpose, multicollinearity does not interfere with the assessment of phylogenetic signal in the residual variation.

The process outlined in the above paragraph will identify functional traits that are important both in explaining species composition and in reducing phylogenetic signal in the residual variation. To provide additional evidence that the traits identified reduce phylogenetic signal in the residual variation, we analysed each trait separately to test for two anticipated properties. The first property is that the functional trait should itself show phylogenetic signal among species (i.e. related species have similar functional trait values); otherwise, a trait could not produce (and hence reduce in the model) phylogenetic signal in species' abundances. Therefore, we tested each continuous trait for phylogenetic signal among species using Pagel's λ (Pagel, 1999), and for binary traits we used phylogenetic logistic regression (Ives & Garland, 2010). We also tested for phylogenetic signal in continuous and binary functional traits using Blomberg's K (Blomberg *et al.*, 2003). We used Blomberg's K in addition to model-based methods because K makes no specific assumption about the model form of phylogenetic signal (e.g. whether phylogenetic signal varies as a result of stabilizing selection vs accelerating rates of evolution). As shown in the Results section, K identified some traits as having phylogenetic signal that were not identified by Pagel's λ or phylogenetic logistic regression. The second property

is that there should be variation among sites in the relationship between species trait values and abundances. If a trait has phylogenetic signal but there is no variation in relationships between plant functional trait values and abundances among sites, then the trait will contribute to the overall phylogenetic signal of species abundance as captured by $b_{spp[i]}$ in Eqn 1, but it probably will not affect phylogenetic patterns nested within sites captured by c_i. Therefore, we tested for variation among sites in the relationship between trait values and log abundances using the linear mixed model (LMM):

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

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

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{I}_m) \qquad\qquad \text{Eqn 3}$$

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

where $t_{spp[i]}$ is the focal functional trait value of the species corresponding to observation i, and σ_b^2 gives the variation among sites in the relationship between trait values and log abundances. This formulation is closely related to the model used by Pollock *et al.* (2012), although we included a main effect for traits. If $\sigma_b^2 > 0$, we concluded that different sites select species differently based on the focal trait. We used a significance threshold of $P < 0.1$ here to lower the risk of excluding potentially important functional traits.

Which environmental variables drive phylogenetic pattern?

Sometimes phylogenetic patterns in community composition are observed, and there is still strong phylogenetic signal in the residual variation after including measured functional traits. In such cases, how can we identify additional functional traits that might further reduce residual phylogenetic pattern? Phylogenetically related species are usually assumed to be ecologically similar as a consequence of niche conservatism (Wiens *et al.*, 2010). Therefore, related species will tend to have similar responses to environmental variables. If these environmental variables are strong enough to drive phylogenetic patterns in community composition, then functional traits associated with tolerance or sensitivity to these environmental variables will probably affect community composition. We can therefore investigate phylogenetic patterns in how species respond to environmental variables in order to infer possible additional, unmeasured functional traits that might help to explain patterns in community composition.

We tested for phylogenetic patterns in the responses of species to environmental variables using the PLMM:

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

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

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

$$g \sim \text{Gaussian}(\mathbf{0}, \sigma_g^2 \mathbf{I}_n) \qquad\qquad \text{Eqn 4}$$

$$h \sim \text{Gaussian}(\mathbf{0}, \sigma_h^2 \Sigma_{spp})$$

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

Here, $g_{spp[i]}$ and $h_{spp[i]}$ represent nonphylogenetic and phylogenetic variation among species in their response to environmental variable x, respectively (see Model II in Ives & Helmus, 2011). The key parameter of interest is σ_h^2 which we tested using a likelihood ratio test. If $\sigma_h^2 > 0$, phylogenetically related species respond to environmental variable x in similar ways, suggesting the existence of an unmeasured phylogenetically inherited trait associated with species tolerances or sensitivities to x. Similar to Eqn 2, multiple environmental variables can be included by replicating the term $(\beta + g_{spp[i]} + h_{spp[i]})x_{site[i]}$ for each additional variable x.

Empirical examples

We applied our analytical framework to two empirical data sets including trait, community, and environmental data. The first is the Dune Meadows data from Jongman *et al.* (1995) and the second is from the Wisconsin Pine Barrens (Li & Waller, 2015). The Dune Meadows data set consists of 30 plant species in 20 sites. Five environmental variables were measured for each site: thickness of the soil A1 horizon (A1), moisture content of the soil (moisture), agriculture land use (use), amount of manure applied (manure), and grassland management type (management). Five functional traits for higher plants were obtained by Jamil *et al.* (2013), including specific leaf area (SLA; $m^2 \, kg^{-1}$), plant height (cm), leaf dry matter content (LDMC; %), seed mass (g per seed), and life history (annual or perennial). We removed the two moss species as they lack functional trait data, resulting in 28 species across the 20 sites.

The Pine Barrens data consist of 152 species distributed among 30 Pine Barrens forest sites in the central Wisconsin sand plains (Li & Waller, 2015). We measured 20 environmental variables at each site, including soil properties (17 variables), climatic conditions (minimum temperature and precipitation), and canopy cover. The vegetation data set and canopy cover have been published in Li & Waller (2015). For the 55 focal species that occurred in three or more of the 30 communities, we measured 11 continuous and four categorical functional traits on at least 12 individuals (four from each of at least three populations) using standard protocols (Perez-Harguindeguy *et al.*, 2013). Continuous traits include seed mass (g per seed), plant height (cm), SLA ($m^2 \, kg^{-1}$), 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 each categorical trait into two levels: growth form (woody vs nonwoody), life cycle (annual vs nonannual), 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).

For both data sets, the available functional traits cover the leaf–height–seed (LHS) plant ecology strategy (Westoby, 1998) and represent multidimensional functions of plants associated with resource use, competitive ability, dispersal ability, and so on. For analyses, we log-transformed highly skewed traits and then Z-transformed all numerical trait values to have means of 0

and standard deviations of 1, allowing coefficients in the mixed models to be interpreted as effect sizes.

We obtained phylogenies for both sets of species from the super-phylogeny provided by Zanne *et al.* (2014) using the program PHYLOCOM (Webb *et al.*, 2008). This time-calibrated phylogeny was constructed from seven gene regions for 32 223 plant species using maximum-likelihood estimates in RAxML (Stamatakis, 2014). Use of the same super-phylogeny ensures that any differences between data sets in phylogenetic patterns are free from biases that might arise if we had instead used phylogenies constructed using different methods.

We fitted the PLMMs and LMMs (Eqns 1–4) with maximum likelihood using function COMMUNITYPGLMM in the PEZ (Pearse *et al.*, 2015) package of R (R Core Team, 2015). We have provided R code for the Dune Meadows as Notes S2. Phylogenetic signals of functional traits were tested using R packages PHYLOLM (Ho & Ané, 2014) and PICANTE (Kembel *et al.*, 2010).

Results

Phylogenetic community composition

Phylogenetically related species co-occurred more often than expected by chance in both Dune Meadows and Pine Barrens communities (Fig. 2). The PLMM (Eqn 1) revealed significant phylogenetic attraction for abundance data in both the Dune Meadows and the Pine Barrens communities ($P = 0.005$ and 0.013, respectively; Table 1). For presence/absence data, the Dune Meadows data had positive but nonsignificant phylogenetic attraction ($\sigma_c^2 = 0.032$, $P = 0.147$; Table S1), whereas phylogenetic attraction in the Pine Barrens was positive and significant ($\sigma_c^2 = 0.044$; $P = 0.018$; Table S1). We found no evidence for phylogenetic repulsion in either data set (Tables 1, S1).

Are there residual phylogenetic patterns after incorporating functional traits?

Because species abundances in both Dune Meadows and the Pine Barrens communities showed phylogenetic attraction, we investigated whether these phylogenetic patterns remained after incorporating information about functional traits. Based on Eqn 2, we first added traits as fixed and random effects that provided information (as measured by AIC) about the abundances of species among communities. We then added any additional traits to the model that reduced the remaining phylogenetic signal in the residual variation (if any) of species abundances among sites.

In the Dune Meadows communities, SLA and life history (annual or perennial) were selected as important traits in the model of species abundances as measured by the AIC (Eqn 2 without terms $b_{spp[i]}$ and c_i). Including these two traits as fixed terms and SLA as a random term reduced the phylogenetic variation (σ_c^2, Eqn 2) by 20% (Table 2a). Including the additional trait seed mass as both fixed and random terms further reduced σ_c^2 by 5%. SLA, life history, and seed mass thus reduced only a small fraction (25%) of the total phylogenetic signal in community composition (abundance patterns of species across sites), with strong

phylogenetic attraction remaining after including the traits (Table 2b). This finding is consistent with additional evidence provided by analysing all five traits separately (Tables 3, S3). Although one or two (based on Pagel's λ or Blomberg's *K*, respectively) of the five traits showed phylogenetic signal among species, no trait explained any of the variance among species (regardless of phylogeny) in the log abundance of species among sites (in Eqn 3 the estimate of σ_b^2 was zero). By contrast, the parallel analyses of species presence/absence in the Dune Meadows data show that including four of five traits reduced phylogenetic signal to almost zero (although the initial value of $\sigma_c^2 = 0.0398$ was not significant; Table S2).

In the Pine Barrens communities, leaf circularity, leaf width, seed mass, and pollination mode were identified as important in explaining species abundances among sites as measured by the reduction in AIC of the nonphylogenetic model of species abundances (Eqn 2 without terms $b_{spp[i]}$ and c_i; Table 2a). Including these four traits in the phylogenetic model (all of them as fixed terms; leaf circularity and leaf width as random terms) eliminated phylogenetic variation in the residuals (σ_c^2 went to 0 in Eqn 2). Thus, these functional traits suffice to account for all the phylogenetic signal in the composition of the Pine Barrens communities. In analyses of each trait, most traits showed strong phylogenetic signal (Table 3). Five traits – leaf circularity, leaf width, leaf thickness, SLA, and animal dispersal (marginally) – also affected plant species' abundances among sites ($\sigma_b^2 > 0$, Eqn 3; Table 3), indicating that different sites favour particular species based on these traits. Among these five traits, each trait covaries with either leaf circularity or leaf width. For example, animal dispersal covaries with leaf circularity ($r = 0.484$), and SLA and leaf thickness covary with leaf width ($r = 0.458$ and -0.504, respectively). Therefore, even though only leaf circularity and leaf width were included via forward selection in the final model (Table 2a), these might have accounted for possible phylogenetic signal produced by the three other traits. This result provides supporting evidence that these traits are important in accounting for phylogenetic patterns in community structure. In parallel analyses of species presence/absence, σ_c^2 again declined to zero after including traits via forward selection (Table S2a).

The overall fits of the models give a detailed statistical description of the variation in abundance and presence/absence among communities. For abundance in the Pine Barrens data set (Table 2a), the model with variation in the effects of traits among sites (random effects) has fixed effects of leaf circularity, leaf width, seed mass, and pollination mode. Coefficients of seed mass and leaf circularity are positive, suggesting that species with higher values of each of the traits are more common across all communities. Leaf width has associated nonzero random effects, implying that sites differ in the relationships between leaf width and species abundance. There is large among-species variation in mean log abundance ($\sigma_a^2 = 0.878$). Interestingly, this variation lacks any phylogenetic component ($\sigma_b^2 = 0$), meaning that there is no phylogenetic signal in overall abundance. There is some site-to-site variation in mean log total abundance across all species ($\sigma_{site}^2 = 0.0058$). This reflects the fact that some sites have higher overall species abundance. As discussed in the preceding
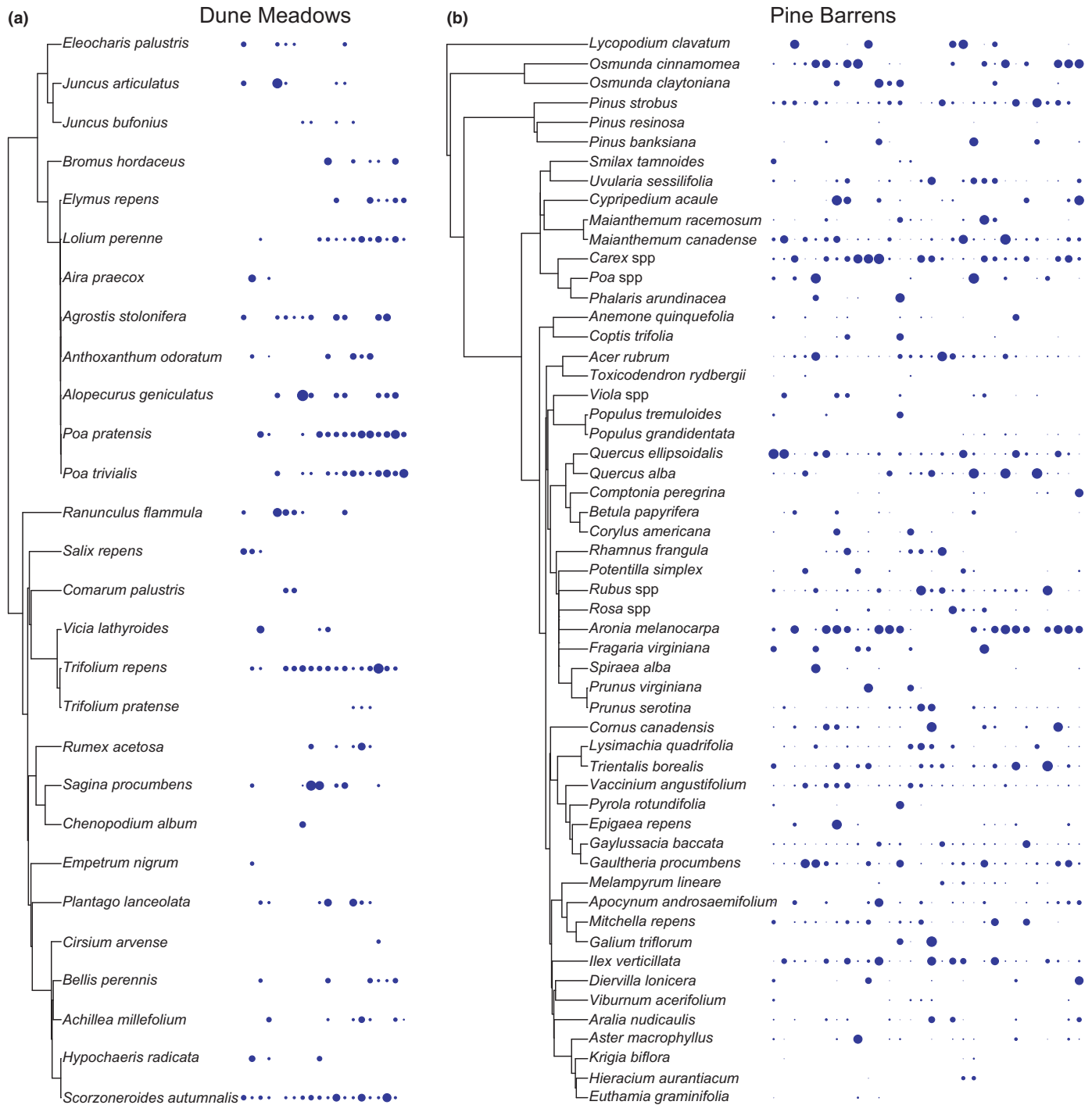
Fig. 2 Phylogeny and relative abundance of plant species found in the Dune Meadows and the Pine Barrens communities. The size of dots is proportional to abundances within each site.

paragraph, there is no site-to-site variation in log abundances in which phylogenetically related species are more likely to show high abundance in the same sites ($\sigma_c^2 = 0$). Finally, the residual variance $\sigma_{resid}^2 = 0.512$ implies a lot of variation in the log abundances of individual species among sites that is not explained by any other component of the model. This example shows that the fully fitted PLMM provides a detailed accounting of the effects of traits on variation in community composition, assessing patterns of both variation among species and the variation among sites.

Which environmental variables drive phylogenetic pattern?

If including trait information in the PLMM of community composition (Eqn 2) leaves residual phylogenetic signal in the unexplained variation, this would suggest that there are other unmeasured traits that affect community composition. To search for any such additional traits, we investigated whether there was phylogenetic signal in the response of species to different environmental variables. Such signals could help us narrow the search

Table 1 Estimated components of variance from the phylogenetic linear mixed model (PLMM) of species abundances (log-transformed) in Dune Meadows and Pine Barrens data sets

Data set	PLMM	σ_a^2	σ_b^2	σ_c^2	σ_d^2	σ_e^2	$P(\sigma_c^2 = 0)$
Dune Meadows	Phylogenetic attraction: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \Sigma_{spp}))$	0.80	0.00958	0.00781	0.00265	0.36	**0.005**
	Phylogenetic repulsion: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 (\Sigma_{spp})^{-1})$	0.90	0.00126	0	0.00291	0.39	0.500
	Nonnested model: c removed	0.90	0.00291	–	0.00126	0.39	–
Pine Barrens	Phylogenetic attraction: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \Sigma_{spp}))$	0.98	0	0.00940	0.00266	0.51	**0.013**
	Phylogenetic repulsion: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 (\Sigma_{spp})^{-1})$	0.98	0	0	0.0228	0.53	0.500
	Nonnested model: c removed	0.98	0.0229	–	0	0.53	–

Phylogenetic attraction and repulsion are estimated in separate models by σ_c^2 (Eqn 1). σ_a^2 and σ_b^2 are the estimated variances of the overall abundance of species partitioned into nonphylogenetic (σ_a^2) and phylogenetic (σ_b^2) components; σ_d^2 is the estimated variance of overall abundance for all sites; and σ_e^2 is the residual variance. Significant results ($P < 0.05$) are given in bold.

Table 2 Proportion of phylogenetic signal of species abundances in the Dune Meadows and Pine Barrens communities reduced after including measured functional traits (Eqn 2); (a) models for both data sets with measured functional traits that are important in explaining species composition; (b) models with additional measured functional traits that can further reduce phylogenetic variation in the residual variation; because there is no residual phylogenetic variance ($\sigma_c^2 = 0$) for Pine Barrens after the first selection, additional functional traits were not added

	Dune Meadows				Pine Barrens			
(a)	Terms	With traits	Without traits	Decrease in signal	Terms	With traits	Without traits	Decrease in signal
Random terms	σ_a^2	0.0344	0.0316		σ_a^2	0.8780	0.8771	
	σ_b^2	0.0004	0.0005		σ_b^2	0	0	
	σ_c^2	0.0066*	0.0082*	20%	σ_c^2	0.0000	0.0094*	100%
	σ_{SLA}^2	0.0030			σ_{circ}^2	0.0000		
					$\sigma_{l.width}^2$	0.0984		
	σ_{site}^2	0.0144	0.0026		σ_{site}^2	0.0058	0.0027	
	σ_{resid}^2	0.3467	0.3554		σ_{resid}^2	0.5120	0.5137	
Fixed terms	Intercept	0.510***	0.510***		Intercept	1.612**	1.612**	
	SLA	0.172**	0.170***		Seed.mass	0.082*	0.083*	
	Annual	−0.387**	−0.387**		Polli.mode	−0.844	−0.844	
					Circ	0.249	0.249	
					L.width	−0.029	−0.029	

	Dune Meadows			
(b)	Terms	With traits	Without traits	Decrease in signal
Random terms	σ_a^2	0.0364	0.0312	
	σ_b^2	0.0004	0.0004	
	σ_c^2	0.0062*	0.0083*	25%
	σ_{SLA}^2	0.0155		
	$\sigma_{Seed.mass}^2$	0.0134		
	σ_{site}^2	0.0036	0.0027	
	σ_{resid}^2	0.3332	0.3551	
Fixed terms	Intercept	0.456***	0.455***	
	SLA	0.172**	0.169***	
	Annual	−0.138**	−0.140**	
	Seed.mass	−0.029	−0.028	

σ_a^2 and σ_b^2 are the estimated variances of the overall abundance of species partitioned into nonphylogenetic (σ_a^2) and phylogenetic (σ_b^2) components. SLA, specific leaf area; Seed.mass, seed mass; Polli.mode, pollination mode; Circ, leaf circularity; L.width, leaf width. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

for additional traits to those that affect the response of species to the environmental variables.

In both data sets, species varied greatly in how their abundances responded to the measured environmental variables

(Table 4). However, in the Dune Meadows data set we only found strong phylogenetic signal in the variation of species responses to manure (last column, Table 4). Thus, related species tend to respond similarly to similar applications of manure. This

Table 3 Phylogenetic signal present in the measured functional traits of Dune Meadows and Pine Barrens

Data set	Trait	Pagel's λ	K	P ($\sigma_b^2 = 0$)
Dune Meadows	Leaf specific area (SLA; m^2 kg^{-1})	0.63	0.09	0.500
	Leaf dry mass content (LDMC; %)	0.76*	0.11*	0.500
	Plant height (cm)	0.00	0.12*	0.500
	Seed mass (g per seed)	0.00	0.07•	0.428
	Life cycle (annual or nonannual)	0.00	0.06	0.500
Pine Barrens	**Leaf specific area (SLA; m^2 kg^{-1})**	0.57**	0.32**	**0.002**
	Leaf circularity (dimensionless)	1.00***	0.95***	**0.001**
	Leaf thickness (mm)	0.75***	0.78***	**0.001**
	Leaf width (cm)	1.00***	0.61***	**0.008**
	Animal dispersal (yes or no)	0.83***	0.52***	**0.054**
	Life cycle (annual or nonannual)	0.00	0.28	0.479
	Growth habit (woody or nonwoody)	1.37***	0.48**	0.500
	Pollination mode (biotic or abiotic)	0.08	0.20	0.500
	Seed mass (g per seed)	0.66	0.36*	0.373
	Leaf dry mass content (LDMC; %)	0.54*	0.29**	0.500
	Stem dry mass content (SDMC; %)	0.48	0.25*	0.500
	Plant height (cm)	0.76**	0.31**	0.500
	Leaf length (cm)	0.77***	0.39**	0.500
	Leaf carbon content (%)	0.68**	0.16	0.500
	Leaf nitrogen content (%)	0.00	0.15	0.334
	Wind dispersal (yes or no)	1.17***	0.61***	0.265
	Unassisted dispersal (yes or no)	0.00	0.15	0.500

We expected that functional traits playing roles in phylogenetic patterns will have phylogenetic signal and differences among sites in how they affect the abundance of species ($\sigma_b^2 = 0$; Eqn 3; P-values in last column). •, $P < 0.1$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Significant results ($P < 0.10$) are given in bold.

Table 4 Variation in species abundances in relation to environmental variables (Eqn 4)

Data set	Environmental variables	P-values σ_g^2 (no phylogenetic signal)	P-values for σ_h^2 (phylogenetic signal)
Dune Meadows	Manure	**< 0.001**	**0.025**
	Soil moisture	**< 0.001**	0.500
	Soil A1 depth	**< 0.001**	0.500
	Management	0.221	0.186
	Land use	0.099	0.350
Pine Barrens	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
	Nitrogen (N)	**< 0.001**	0.500
	Phosphorus (P)	**0.039**	0.500
	Magnesium (Mg)	**0.030**	0.500
	Potassium (K)	**0.007**	0.500
	Sodium (Na)	**< 0.001**	0.500
	Manganese (Mn)	**< 0.001**	0.500
	Calcium (Ca)	**< 0.001**	0.208
	Clay	0.110	0.500
	Silt	0.070	0.500
	Sand	0.117	0.500
	Iron (Fe)	0.500	0.500
	Sulphur (S)	0.458	0.500
	Zinc (Zn)	0.500	0.500
	Aluminium (Al)	0.500	0.500

In the Dune Meadows data set, only species' responses to manure levels showed phylogenetic signal; in the Pine Barrens data set, although 13 of the 20 environmental variables co-varied with variation in species abundances among sites, none showed phylogenetic signal in these responses. Significant results ($P < 0.05$) are given in bold.

suggests that functional traits related to responses of plants to soil nutrients may help to further explain patterns in community composition.

In the Pine Barrens abundance data, we found no phylogenetic signal in the variation among species in their responses to any measured environmental variable. By contrast, in the presence/absence data we found strong phylogenetic signal in species' responses to minimum temperature, soil pH, and calcium (Ca) and manganese (Mn) concentrations (PGLMM, Table S4). Related species thus tend to occupy similar Pine Barrens sites as measured by these environmental variables. This pattern, again reflecting mostly soil conditions, suggests that functional traits related to nutrient acquisition and physiology are affecting species incidence across sites. This is true even though four traits – pollination mode, seed mass, leaf width, and wind dispersal – together eliminated phylogenetic signal in the residual variation (Table S2).

Discussion

Our analyses address the question: when models of community structure incorporate functional trait data, how much phylogenetic pattern remains in the unexplained variation? If no residual phylogenetic patterns in species abundance or presence/absence remain after including measured functional traits, then phylogenetic information itself may provide little additional insight into community assembly. In contrast, if significant residual phylogenetic patterns remain after including measured functional traits, then phylogenetic relationships include important information beyond what is provided by these traits. Therefore, answers to this question provide an important starting point for the search for additional traits or other factors that could underlie community assembly.

Our statistical framework addresses this question using phylogenetic mixed models (PLMMs and PGLMMs). We found that phylogenetically related species are indeed more likely to occupy the same sites and reach similar abundances within two quite distinct communities: Dune Meadows in the Netherlands and Pine

Barrens forests in Wisconsin, USA. In the Pine Barrens communities with more species, sites, and functional traits, we found no phylogenetic signal in the residual variation in species abundance and presence/absence among sites after including functional traits in the phylogenetic mixed models (Eqn 2). In the simpler Dune Meadows communities, two of five traits had phylogenetic signal, and these traits removed all phylogenetic pattern in the residual variation in species presence/absence among sites. By contrast, these traits only reduced phylogenetic patterns in species abundance by 25% in the PLMM.

Incorporating functional traits reduced the phylogenetic component of residual variation in species composition for both data sets, but what could explain the remaining phylogenetic component of the variation in abundance of species among sites in the Dune Meadow communities (Fig. 1b, $\sigma^2_{phy-reduced}$)? Some unknown historical biogeographical process or meta-community dynamics could account for this. However, sites in the Dune Meadow data sets all occur within the same 86-km^2 island, making it unlikely that historical biogeographical processes played a key role in the assembly of these communities. It thus seems more likely that unmeasured functional traits played a role. Species abundances among Dune Meadows sites varied in apparent response to several measures of soil conditions, with phylogenetically related species responding similarly to nutrient concentrations (manure in Table 4). This response suggests that unmeasured traits (with a phylogenetic component) related to nutrient uptake and physiological processes affect the abundance of species across these Dune Meadows communities. The scarcity of data on traits related to belowground nutrient acquisition processes (root structure, mycorrhizal associations, etc.) may thus limit our ability to account for variation in species' abundances.

Despite the fact that measured traits appear unrelated to species' responses to the soil and climate gradients used in Table 4, measured traits still reduced much of the residual phylogenetic pattern in community composition, especially in the Pine Barrens communities. A possible explanation for this is that traits affecting how species respond to soil and climate variables covary with some of the measured traits. If traits covary, phylogenetic patterns driven by one trait could be statistically absorbed by another trait (similar to the effects of collinearity of independent variables). Conversely, soil and climate variables could covary with other environmental variables that affect plant species through the measured traits. It is difficult to infer just what mechanisms may be acting here without additional (ideally experimental) data. Although we have shown that the measured functional traits reduce most of the residual phylogenetic structure in the Pine Barrens communities, these associations could still mask important unmeasured traits.

We found that functional traits reduce a greater portion of the phylogenetic signal in species presence/absence than in species abundances in the Dune Meadows communities. This suggests that, although these functional traits strongly affect the suitability of sites for species, they have less effect on the ability of these species to attain large population sizes. Abundance may instead reflect local colonization dynamics or interactions with herbivores and pathogens influenced by traits that we did not measure. Thus, abundance data may provide more information about community assembly than just presence/absence data when analysing the phylogenetic components of community assembly (Freilich & Connolly, 2015).

Implications

It is often assumed that phylogenetic relationships among species contain additional, and possibly much more, ecological information relevant for predicting community assembly than what we find in commonly measured functional traits. This has led some community ecologists to argue that studies analysing community composition should incorporate information from both phylogenies and functional traits (e.g. Cadotte *et al.*, 2013). With the methods we have presented, this is a testable hypothesis, and the Dune Meadows and Pine Barrens data sets give mixed results. On the one hand, the Dune Meadows example showed that phylogenies can indeed provide ecological information in addition to that contained in a small set of functional traits for predicting abundances of species among sites (Vane-Wright *et al.*, 1991; Cadotte *et al.*, 2009). Although functional traits are necessary to accurately infer the processes driving phylogenetic patterns (Kraft *et al.*, 2007; Cavender-Bares *et al.*, 2009), measured functional traits alone appear to provide an incomplete picture of abundances in the Dune Meadows. A recent study suggested that measured functional traits may contain less information about responses to environmental factors than the identity of species (Clark, 2016), implying limitations of using functional traits alone in explaining community assembly. On the other hand, our results from the Pine Barrens suggest that, in some communities, measured functional traits can reduce most of the phylogenetic pattern of community composition.

Our analyses are based on statistical models rather than metrics applied to data and tested using randomization methods. The current popularity of model-based methods in ecology reflects the fact that they are more interpretable, flexible, and statistically powerful than either null models or conventional algorithmic multivariate analyses (Warton *et al.*, 2014). Here, we showed the power of phylogenetic mixed models (PLMMs and PGLMMs) for detecting and investigating phylogenetic patterns in community composition. This ability to combine phylogenies and functional traits into the same statistical model provides an integrated, quantitative framework for analysing ecological communities and predicting the incidence and abundance of one taxon from others. Using our methods to examine the overlaps of information about community composition derived from functional traits and phylogeny can provide useful insights and directions for subsequent analyses.

Statistical models also show the distinction between functional traits as explanations of community structure and phylogenies as patterns in the residual variation not accounted for by traits. Statistical models of community structure treat phylogenies as hypotheses for the pattern of covariances in the residual, unexplained variation. This implies that phylogenetic information

does not explain anything in the same way that functional traits explain patterns as independent variables. Although researchers commonly talk about community patterns being 'explained by phylogenies', it is more accurate to say that covariances in the unexplained variation in community structure are consistent with the patterns anticipated to emerge from phylogenetic relationships.

We can use phylogenetic analyses to infer other unmeasured functional traits that may underlie patterns in community composition. Species often differ in how they respond to gradients in environmental conditions, but related species often respond similarly. In such cases, we expect some functional trait or traits to underlie these responses. The phylogenetic patterns we found here show that plant species commonly respond to edaphic conditions such as soil nutrient and chemistry. This highlights our frequent lack of data on functional traits related to roots and water/nutrient uptake. Thus, the integrated PLMM models provide valuable tools in cases where measured traits cannot fully account for phylogenetic patterns in ecological communities by suggesting which additional traits might be most informative for improving our ability to account for ecological patterns.

The assembly of plant and animal communities is clearly a complex phenomenon involving many processes. Some of these reflect differences among species assessed using the traits that can be measured. Some of these trait differences, in turn, reflect shared ancestry and conservative patterns of trait change among evolving lineages. Therefore, measured functional traits are expected to overlap with phylogenies in their information about species composition. The proportion of this overlap, however, varies from community to community, as found here between the two communities we studied. The tools presented here allowed us to explore these differences. We might also envision developing analogous methods to partition and assess the effects of other processes affecting community assembly (e.g. meta-community or neutral dynamics). This would require us to explicitly model the effects of these processes on variation in species incidence and abundance, and to be able to distinguish the patterns that emerge from those of alternative models.

Author contributions

D.L. and A.R.I. designed the study and performed the analyses. D.M.W. and D.L. designed the Pine Barrens field study and collected the functional trait data. D.L. collected the Pine Barrens vegetation and environmental data. D.L. wrote the initial manuscript and all authors collaborated in revising the paper.

References

Adler PB, Fajardo A, Kleinhesselink AR, Kraft NJB. 2013. Trait-based tests of coexistence mechanisms. *Ecology Letters* 16: 1294–1306.

Baum DA, Smith SD. 2012. *Tree thinking: an introduction to phylogenetic biology.* Greenwood Village, CO, USA: Roberts & Company Publishers.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717–745.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24: 127–135.

Cadotte M, Albert CH, Walker SC. 2013. The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecology Letters* 16: 1234–1244.

Cadotte MW, Cavender-Bares J, Tilman D, Oakley TH. 2009. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS ONE* 4: e5695.

Cavender-Bares J, Ackerly D, Baum D, Bazzaz F. 2004. Phylogenetic overdispersion in Floridian oak communities. *American Naturalist* 163: 823–843.

Cavender-Bares J, Keen A, Miles B. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87: S109–S122.

Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12: 693–715.

Clark JS. 2016. Why species tell more about traits than traits about species: predictive analysis. *Ecology* 97: 1979–1993.

Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549–11555.

Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* 125: 1–15.

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* 160: 712–726.

Freilich MA, Connolly SR. 2015. Phylogenetic community structure when competition and environmental filtering determine abundances. *Global Ecology and Biogeography* 24: 1390–1400.

Garland T, Bennett AF, Rezende EL. 2005. Phylogenetic approaches in comparative physiology. *Journal of Experimental Biology* 208: 3015–3035.

Gelman A, Hill J. 2007. *Data analysis using regression and multilevel/hierarchical models.* New York, NY, USA: Cambridge University Press.

Graves GR, Gotelli NJ. 1993. Assembly of avian mixed-species flocks in Amazonia. *Proceedings of the National Academy of Sciences, USA* 90: 1388–1391.

Harvey PH, Pagel MD. 1991. *The comparative method in evolutionary biology.* Oxford, UK: Oxford University Press.

Ho LST, Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* 63: 397–408.

Hubbell SP. 2001. *The unified neutral theory of biodiversity.* Princeton, NJ, USA: Princeton University Press.

Ives AR, Garland T. 2010. Phylogenetic logistic regression for binary dependent variables. *Systematic Biology* 59: 9–26.

Ives AR, Helmus MR. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs* 81: 511–525.

Jamil T, Opdekamp W, van Diggelen R, ter Braak CJF. 2012. Trait–environment relationships and tiered forward model selection in linear mixed models. *International Journal of Ecology* 2012: ID 947103.

Jamil T, Ozinga WA, Kleyer M, ter Braak CJ. 2013. Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. *Journal of Vegetation Science* 24: 988–1000.

Jongman RH, Ter Braak CJ, Van Tongeren OF, eds. 1995. *Data analysis in community and landscape ecology.* Cambridge, UK: Cambridge University Press.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Kraft NJB, Cornwell WK, Webb CO, Ackerly DD. 2007. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist* 170: 271–283.

Leibold MA, Economo EP, Peres-Neto P. 2010. Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology Letters* 13: 1290–1299.

Li D, Waller DM. 2015. Drivers of observed biotic homogenization in Pine Barrens of central Wisconsin. *Ecology* 96: 1030–1041.

Losos JB. 1996. Phylogenetic perspectives on community ecology. *Ecology* 77: 1344–1354.

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

Losos JB. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology. *American Naturalist* 177: 709–727.

Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149: 646–667.

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

Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, 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.

Pearse WD, Cadotte MW, Cavender-Bares J, Ives AR, Tucker CM, Walker SC, Helmus MR. 2015. Pez: phylogenetics for the environmental sciences. *Bioinformatics* 31: 2888–2890.

Perez-Harguindeguy N, Diaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte MS, Cornwell WK, Craine JM, Gurvich DE *et al.* 2013. New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 61: 167–234.

Pollock LJ, Morris WK, Vesk PA. 2012. The role of functional traits in species distributions revealed through a hierarchical model. *Ecography* 35: 716–725.

R Core Team. 2015. *R: a language and environment for statistical computing, version 3.2.3.* Vienna, Austria: R Foundation for Statistical Computing.

Ricklefs RE, Schluter D. 1993. *Species diversity in ecological communities: historical and geographical perspectives.* Chicago, IL, USA: University of Chicago Press.

Self SG, 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* 30: 1312–1313.

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

Vane-Wright RI, Humphries CJ, Williams PH. 1991. What to protect? Systematics and the agony of choice. *Biological Conservation* 55: 235–254.

Violle C, Navas ML, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E. 2007. Let the concept of trait be functional!. *Oikos* 116: 882–892.

Warton DI, Foster SD, Stoklosa GDJ, Dunstan PK. 2014. Model-based thinking for community ecology. *Plant Ecology* 216: 669–682.

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

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. 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 JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Jonathan Davies T, Grytnes J-A, Harrison SP *et al.* 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13: 1310–1324.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Table S1 Phylogenetic linear mixed models (PGLMM) of species presence/absence

Table S2 Proportion of phylogenetic signal of species presence/absence reduced after including measured functional traits

Table S3 Phylogenetic signal of measured functional traits

Table S4 Variation in species presence/absence in relation to environmental variables

Notes S1 R code to compare *P*-values from likelihood ratio test and parametric bootstrap.

Notes S2 R code and data for analyses of the Dune Meadows data set.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.