



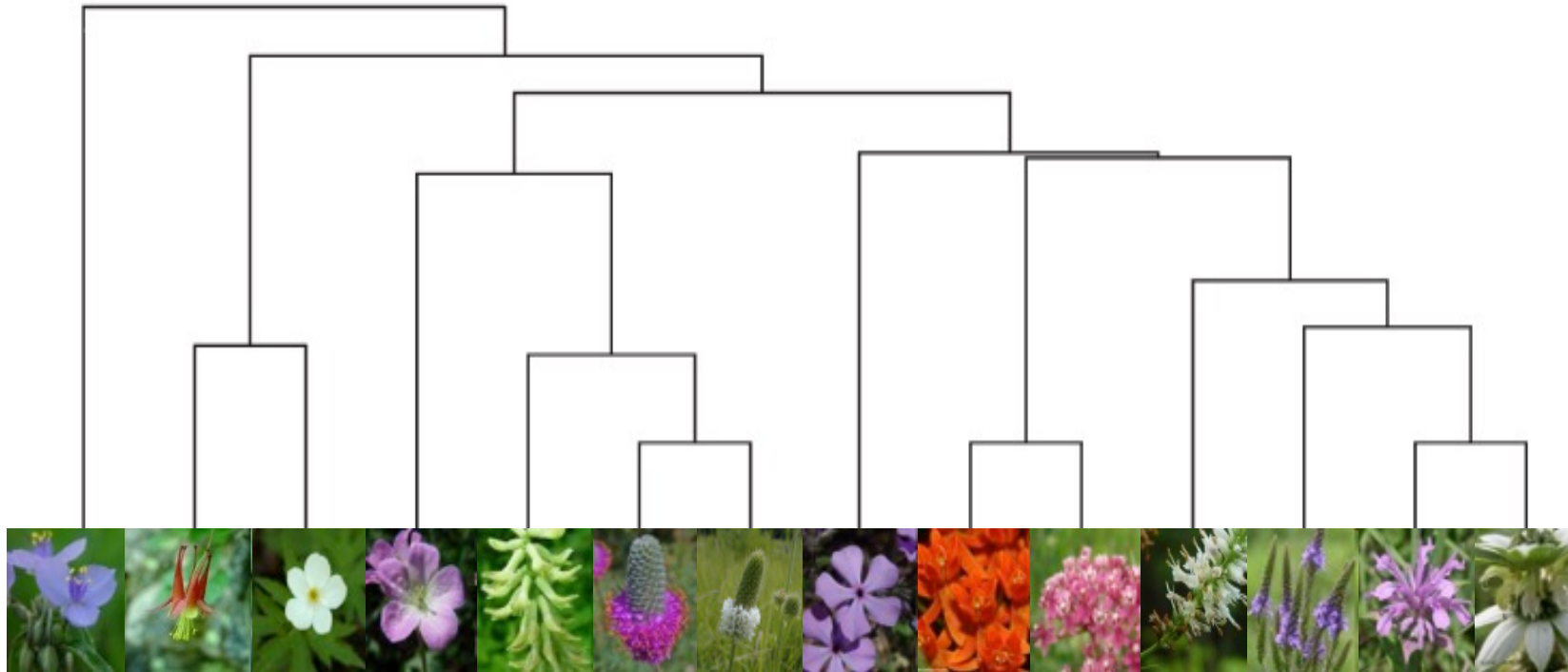
Biodiversity data wrangling: Linking large phylogenies with species traits and ecologies

Module: Phylogenetic Community Ecology

Daijiang Li

Botany 2024

Phylogeny: evolutionary relationships among species



1. Comparative methods among species
2. **Community ecology**

Community ecology

What species are present/absent in different communities?



Why use phylogenies in community ecology?

Phylogenies can be surrogates for missing trait data

Even if relevant traits are unknown, related species should have similar trait values

Phylogeny tells which species will likely have similar traits

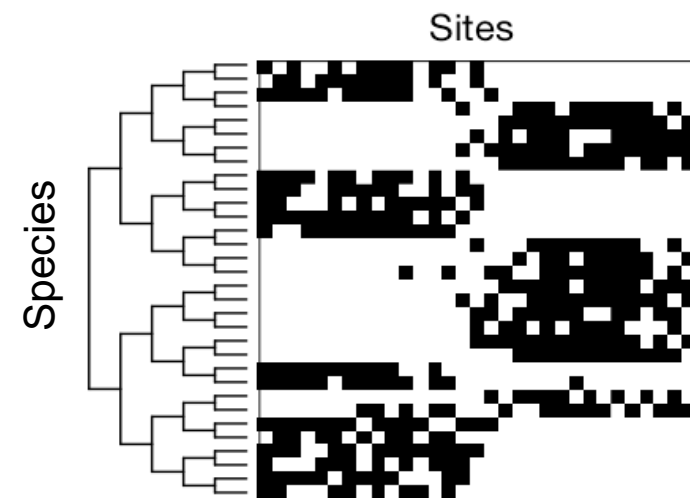
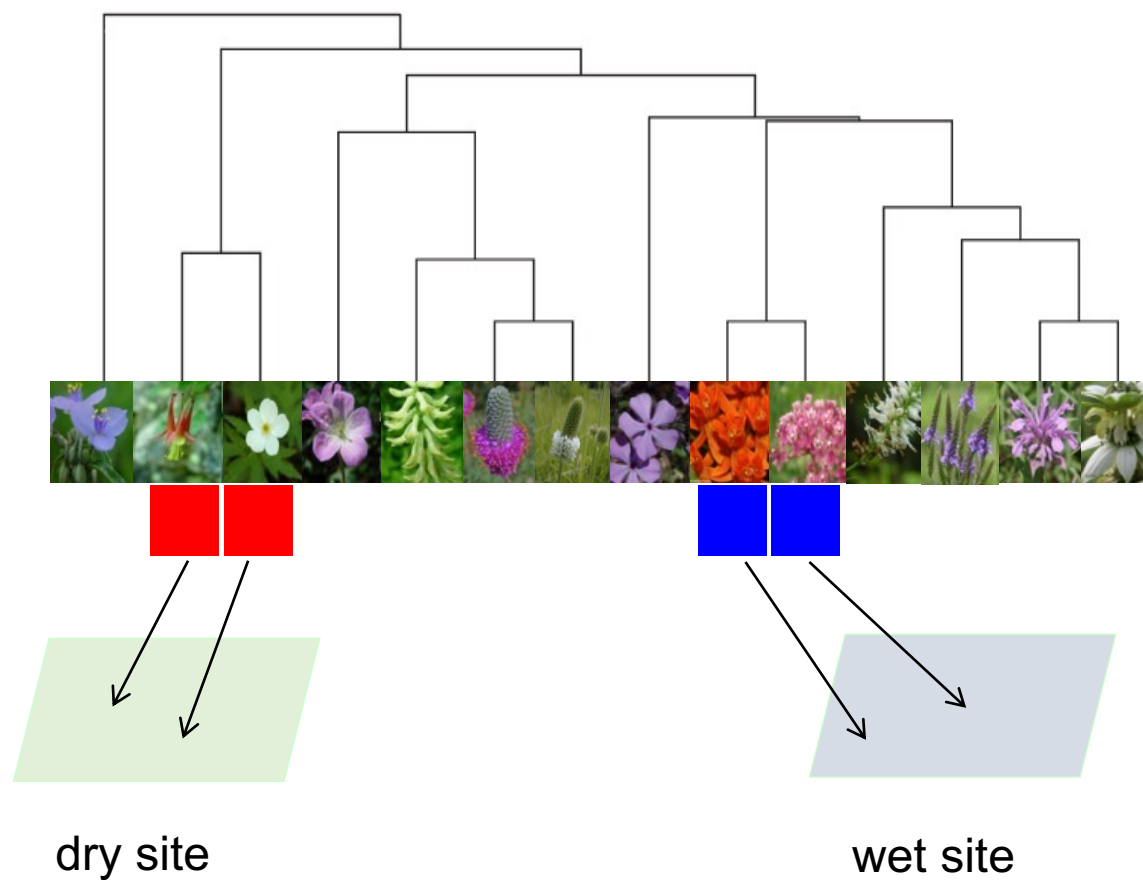


Plant Traits

 wet tolerant

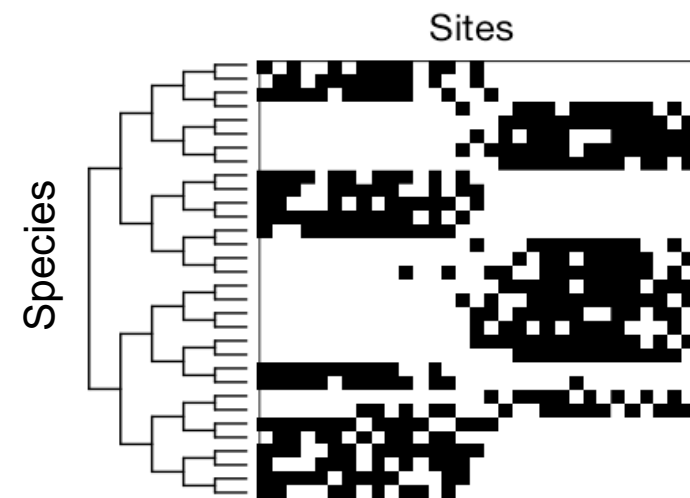
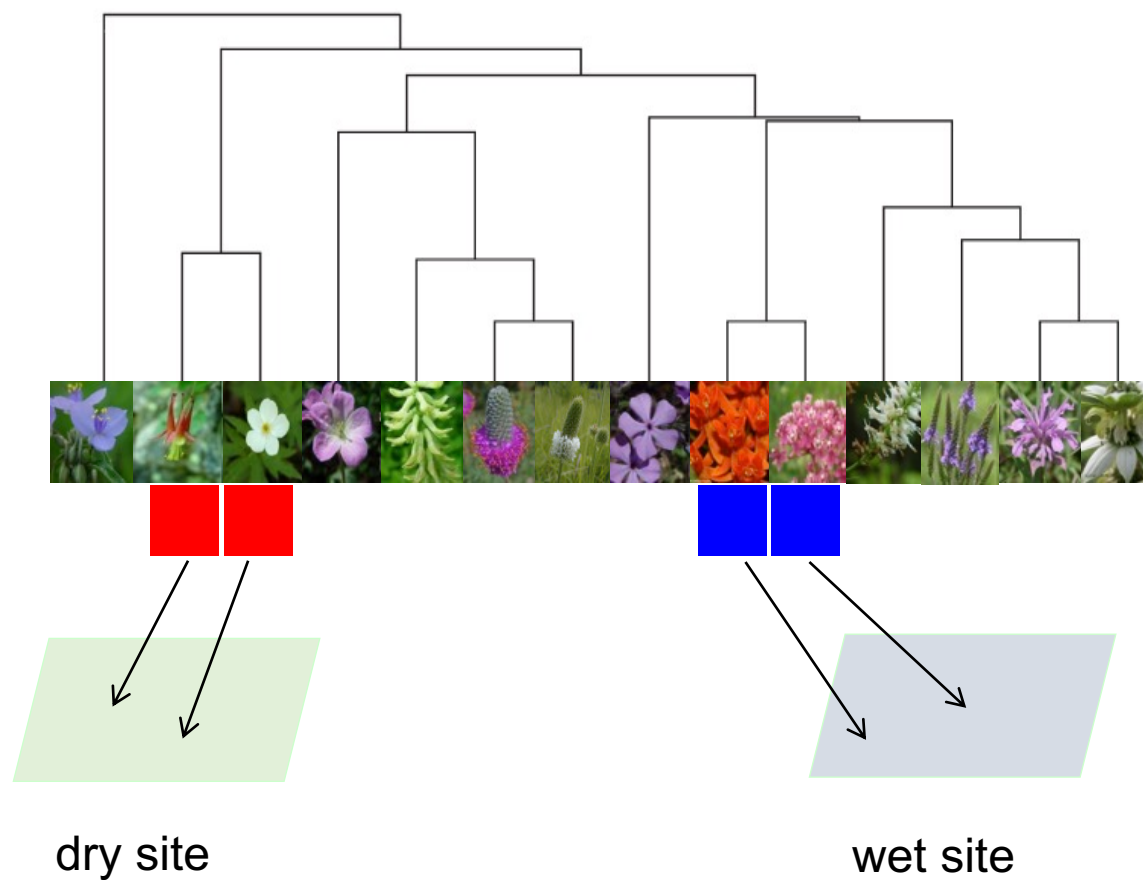
 dry tolerant

Phylogeny tells which species will likely have similar traits



Phylogenetically related species will occur in the same sites

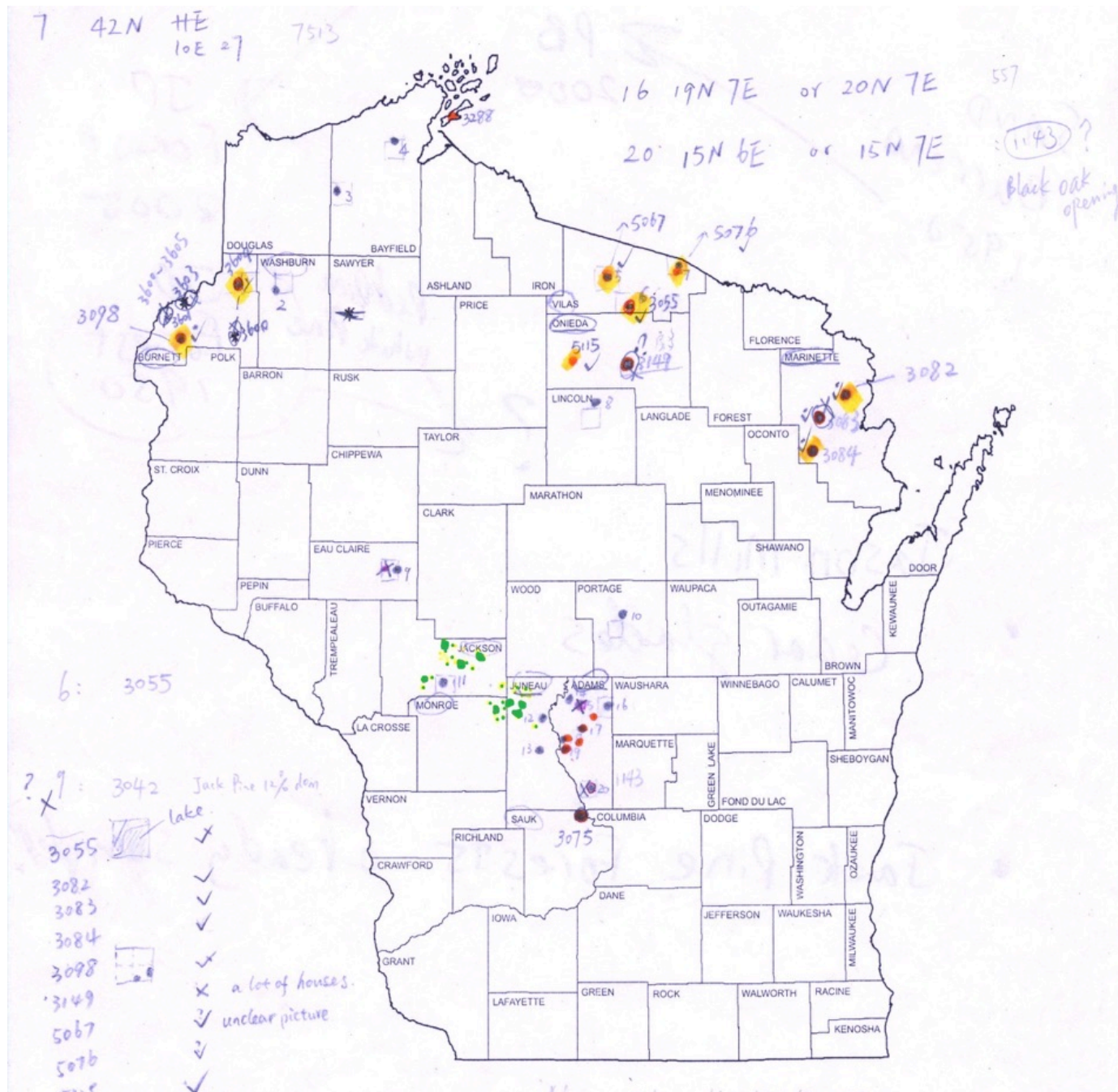
Phylogeny tells which species will likely have similar traits



Phylogeny is a surrogate for traits

Outline

- Phylogenetic diversity
- Phylogenetic Generalized Linear Mixed Models (PGLMMs)

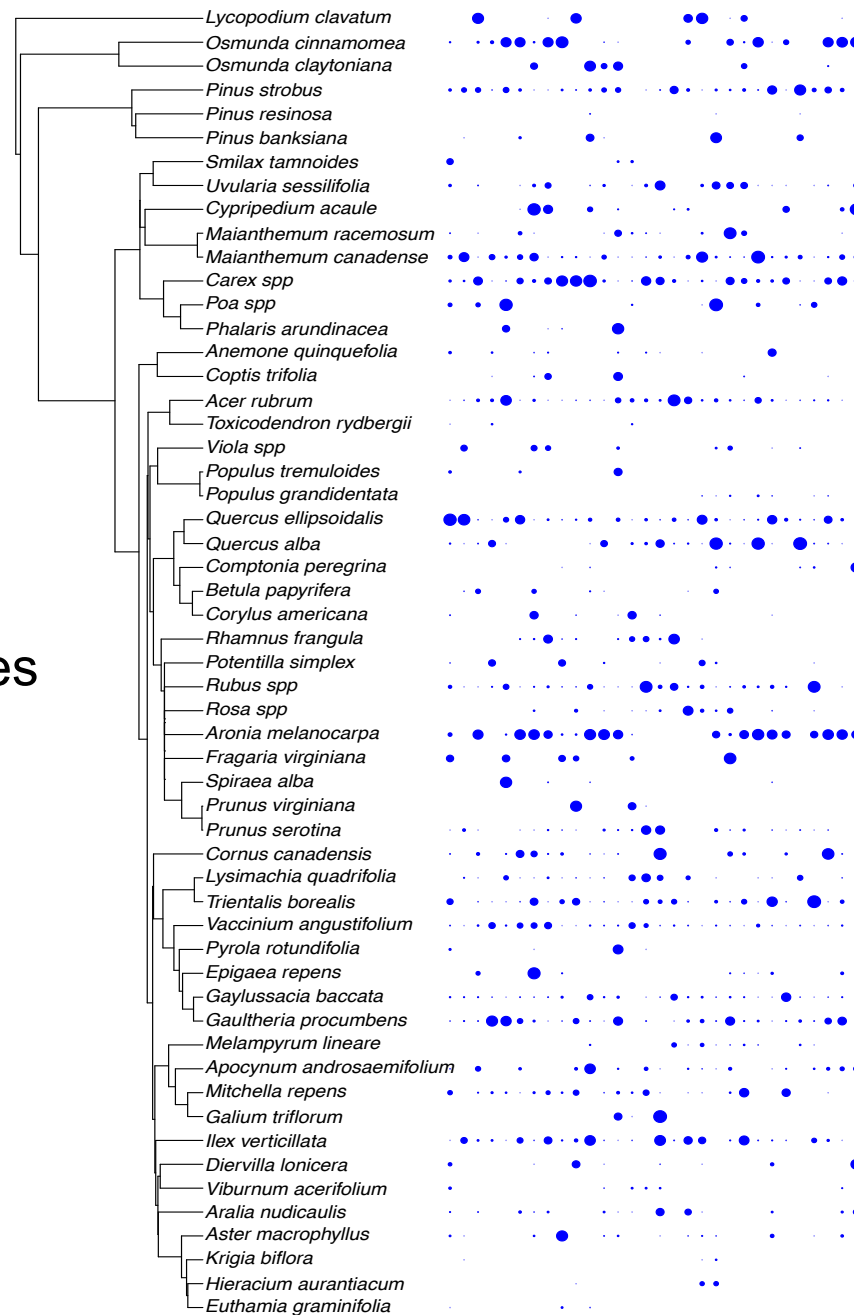


Dataset





species



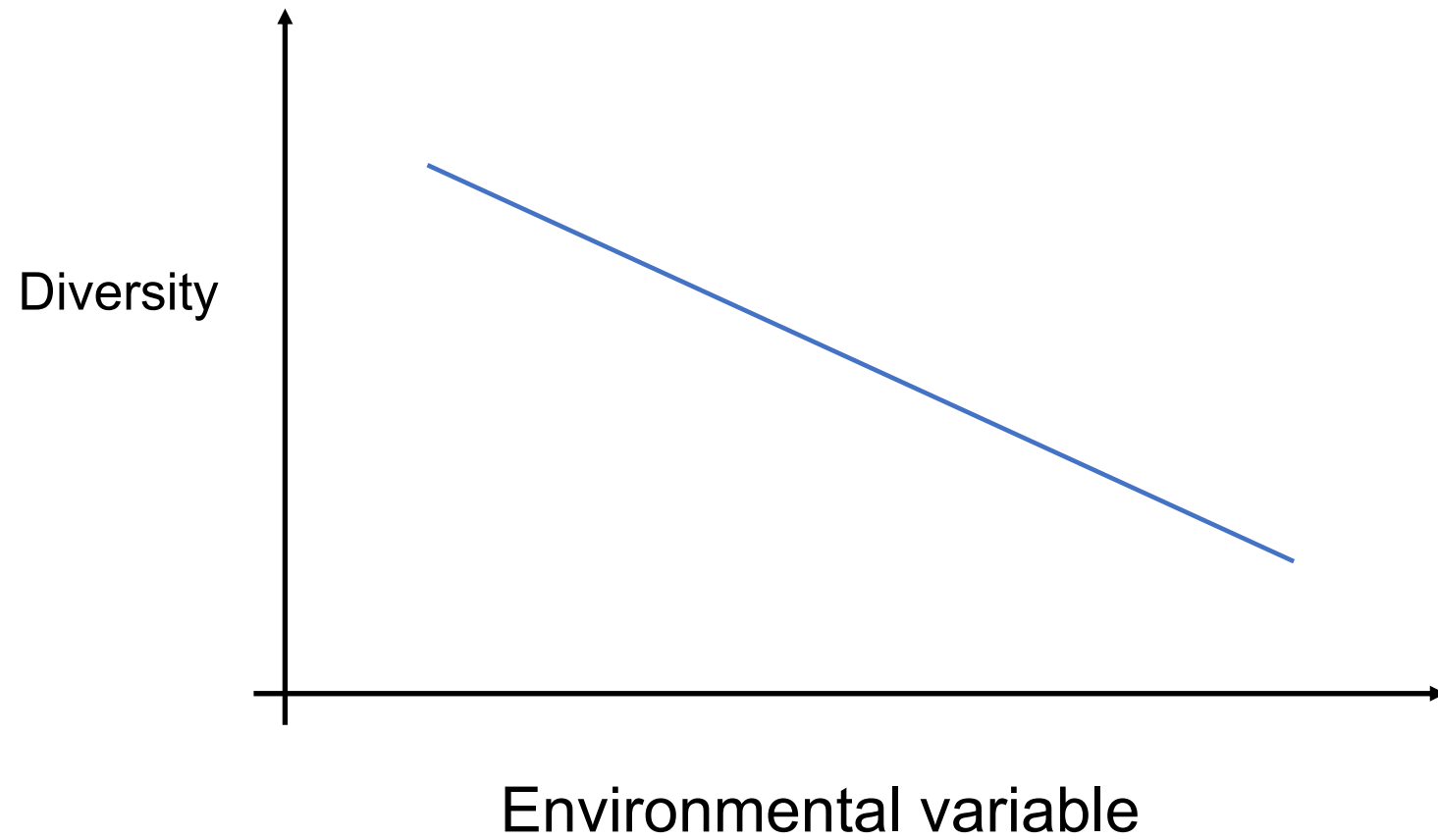
sites

Functional traits

- Woody
- Biotic Pollination
- Shade Tolerance
- Plant Height
- Seed Mass
- Specific Leaf Area (SLA, leaf area/dry mass)
- Leaf Carbon Content (LCC)
- Leaf Nitrogen Content (LNC)
- Stem Dry Mass Content (SDMC)

Common Phylogenetic Diversity Metrics

DIMENSION =>		1. Richness: how much? <i>Sum of ...</i>			2. Divergence: how different? <i>Mean of ...</i>				3. Regularity: how regular? <i>Variance of ...</i>			
<=Level of comparison	Unit =>				2.2 <i>Pairwise distances / similarities</i>		2.3 <i>Phylogenetic isolation</i>	3.1 <i>Tree topology</i>	3.2 <i>Pairwise distances</i>		3.3 <i>Phylogenetic isolation</i>	
		1.1 <i>Branch lengths</i>	1.2 <i>Pairwise distances</i>	1.3 <i>Phylogenetic isolation</i>	2.1 <i>Branch lengths</i>	2.2.1 <i>All</i>			3.2.1 <i>All</i>	3.2.2 <i>Nearest</i>		
<div>α-diversity (within sets)</div>		1.1a Amount of evolutionary history: <i>Across species</i> PD <i>Across individuals</i> ΔnPD <i>Effective (q=0)</i> ⁰D(T)* , ⁰PD(T)* <i>Per species</i> PD_{Ab} <i>Endemic</i> PE	1.2a Sum of pairwise distances: PSR F	1.3a Sum of evolutionary distinctiveness: ED AED	2.1a Sum of branch lengths divided by species richness: avPD avPD_{Ab} Effective number of species given phylogenetic balance and abundance evenness: (q>0) ^qD(T)* , ^qPD(T)* and associated entropies: H_p* I_q*	2.2.1a Effective number of species given phylogenetic balance and abundance evenness: ^qD^Z(p)* Mean of all distances including zero intra-species distances: Rao's QE, MPD_{Ab} PSE I Mean inter-species distances: MPD, AvTD, PSV interMPD_{Ab}	2.2.2a Mean shortest distance between a species and all others: MNTD MNTD_{Ab}	2.3a Mean of species evolutionary distinctiveness: mean(ED)	3.1a Branching symmetry and distribution: I_C γ IAC	3.2.1a Variance of pairwise distances: VPD, Δ⁺, VarTD VPD_{Ab} interVPD_{Ab}	3.2.2a Variance of nearest neighbour distances: VNTD VNTD_{Ab} PE_{ve}	3.3a Variance of species isolation metrics: var(ED) E_{ED} H_{ED} H_{AED} ^qD(P)* ^qD(AP)*



Model-based methods

- Why?
- No aggregation to one value/site
- Integrate multiple source of information (and their interactions)
- Model validation/selection/prediction ...
- Deeper understanding of community dynamics



Warning

This section has a lot of equations in it.

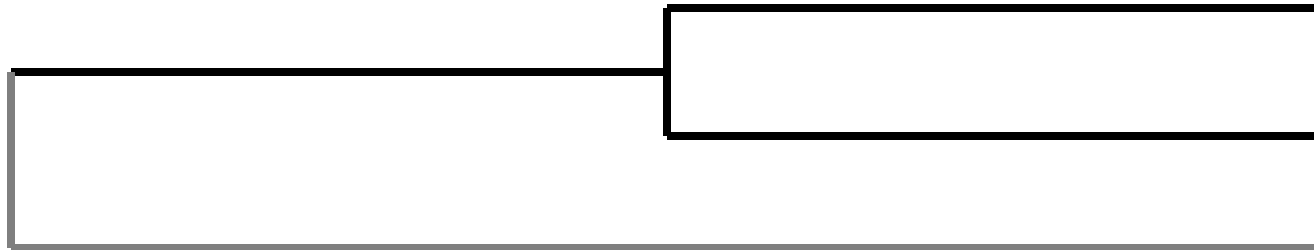
I want to show how useful it is to
analyze data with models.

The models look complicated only
because the questions are complicated.

Preliminaries

Incorporating phylogenies into
statistical models

Turning phylogenies into variances

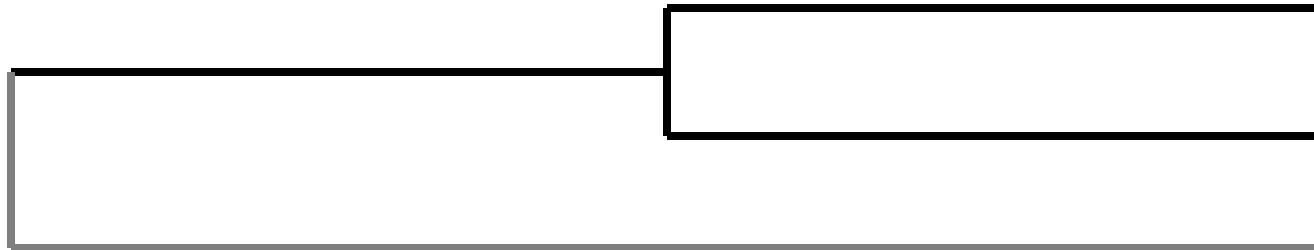
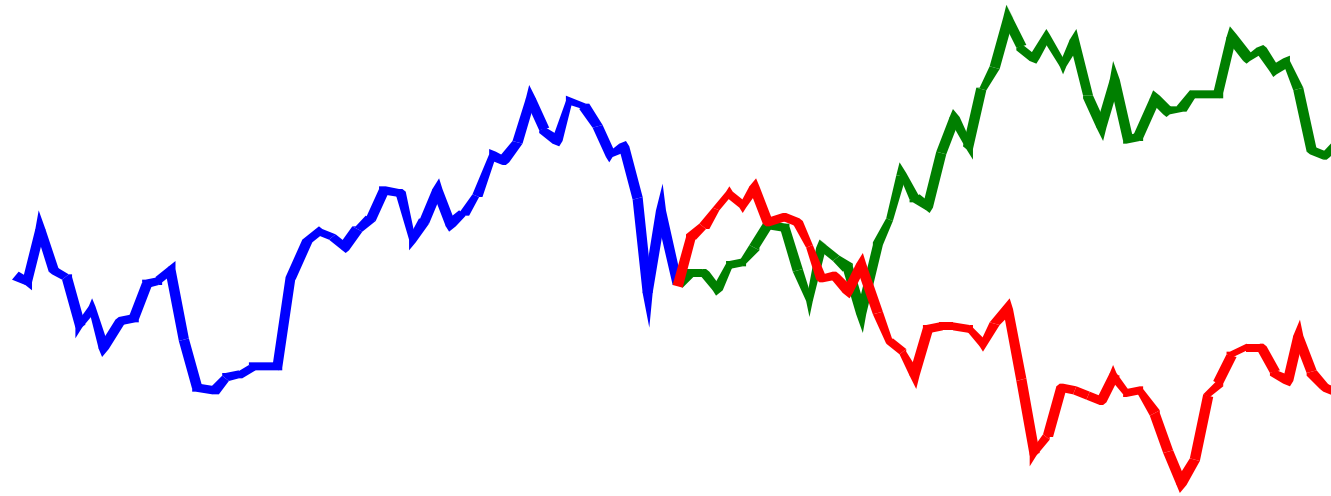


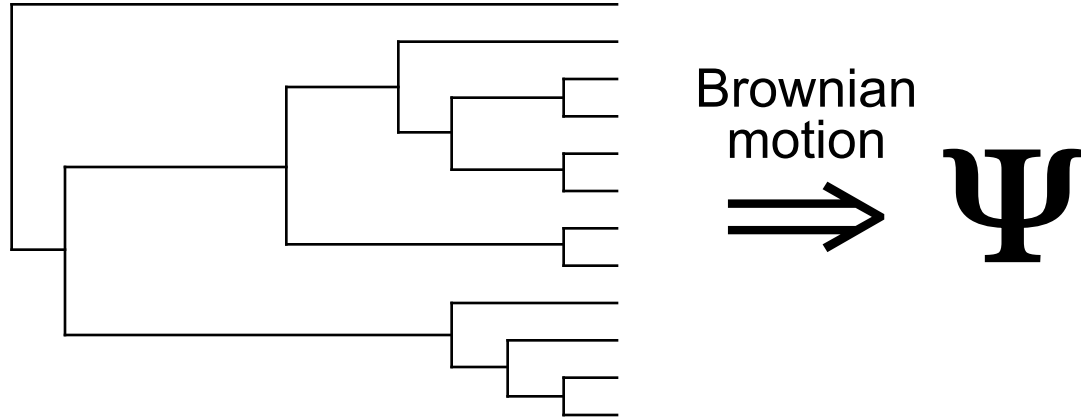
Total variance

Shared variance
= covariance

Variance since
speciation

↑
trait



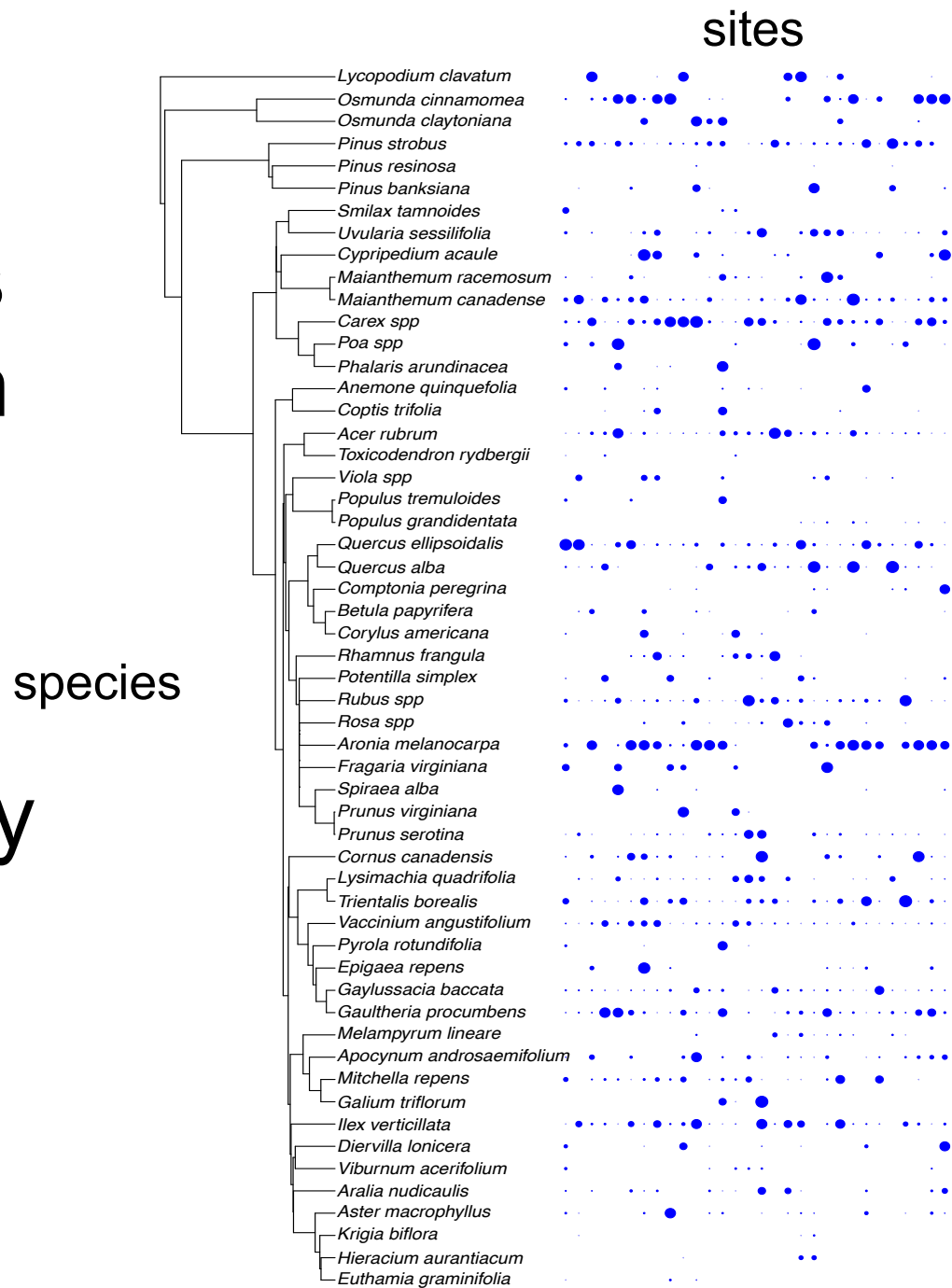


Ψ Matrix giving phylogenetic covariances among species

$$\Psi = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1. Are related species more likely to occur in the same sites?

2. Is this because they share functional traits?



PGLMM for community composition

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + e_i$$

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

$$\mathbf{a}^p \sim \text{Gaussian}(\mathbf{0}, \sigma_{\text{ap}}^2 \mathbf{\Psi})$$

$$\mathbf{c} \sim \text{Gaussian}(\mathbf{0}, \sigma_{\text{site}}^2 \mathbf{I}_m)$$

$$\mathbf{d} \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_{\text{attract}}^2 \mathbf{\Psi}))$$

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

PGLMM for community composition

$$Y_i = \boxed{a_{\text{spp}[i]}} + \boxed{a^p_{\text{spp}[i]}} + c_{\text{site}[i]} + d_i + e_i$$

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

differences in abundance

$$\mathbf{a}^p \sim \text{Gaussian}(\mathbf{0}, \sigma_{\text{ap}}^2 \mathbf{\Psi})$$

...that depend on phylogeny

PGLMM for community composition

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + \boxed{c_{\text{site}[i]}} + d_i + e_i$$

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

differences among sites

PGLMM for community composition

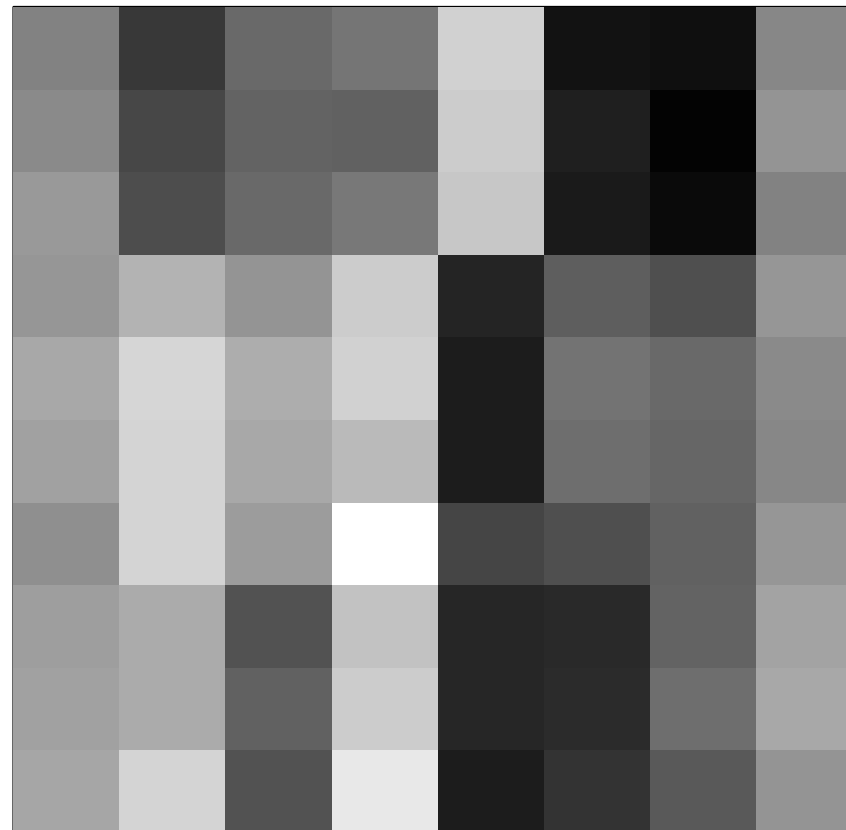
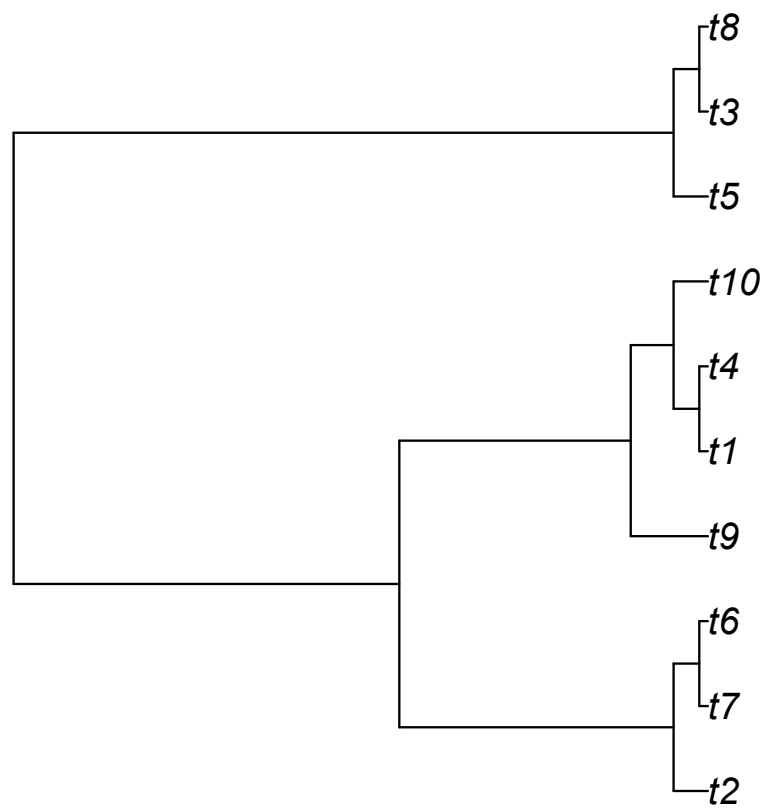
$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + \boxed{d_i} + e_i$$

$$\mathbf{d} \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_{\text{attract}} \mathbf{\Psi}))$$

phylogenetic attraction within sites

Sites contain phylogenetically related species

attraction within sites



sites

PGLMM for community composition

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + \boxed{e_i}$$

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

residual variation

PGLMM for community composition

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + \boxed{d_i} + e_i$$

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

$$\mathbf{a}^p \sim \text{Gaussian}(\mathbf{0}, \sigma_{\text{ap}}^2 \mathbf{\Psi})$$

$$\mathbf{c} \sim \text{Gaussian}(\mathbf{0}, \sigma_{\text{site}}^2 \mathbf{I}_m)$$

$$\mathbf{d} \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_{\text{attract}}^2 \mathbf{\Psi}))$$

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

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + e_i$$

$$\sigma_a = 0.997$$

$$\sigma_{\text{ap}} = 0.001$$

High variation among species in abundance but almost no phylogenetic signal

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + e_i$$

$$\sigma_a = 0.997$$

$$\sigma_{\text{ap}} = 0.001$$

$$\sigma_{\text{site}} = 0.004$$

Little site-to-site variation in mean abundance of species

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + e_i$$

$$\sigma_a = 0.997$$

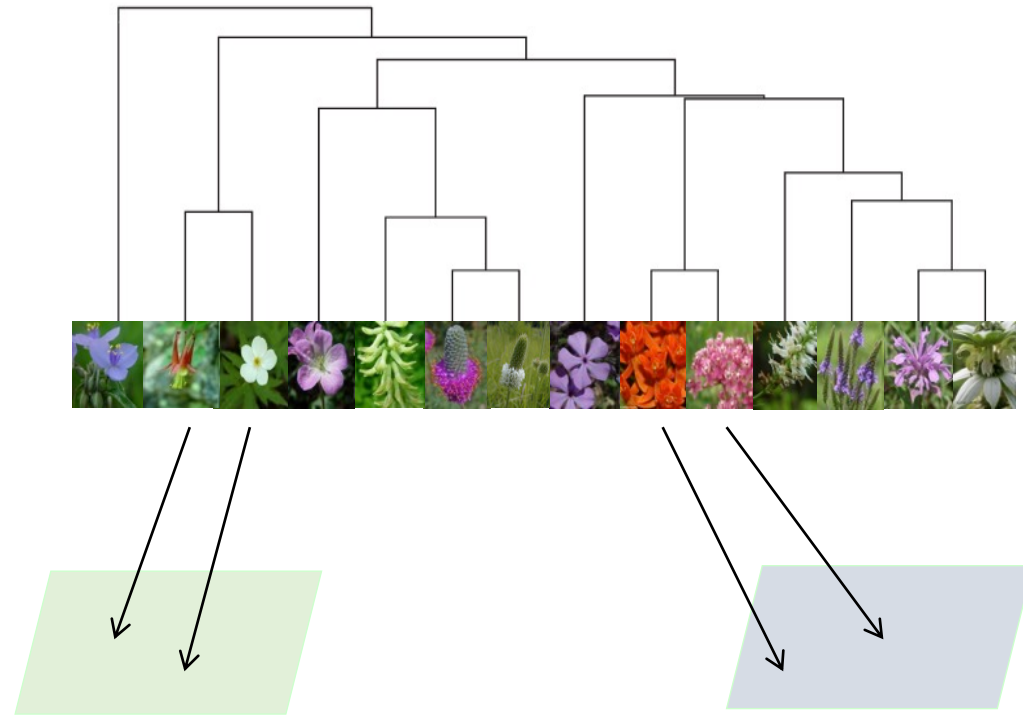
$$\sigma_{\text{ap}} = 0.001$$

$$\sigma_{\text{site}} = 0.004$$

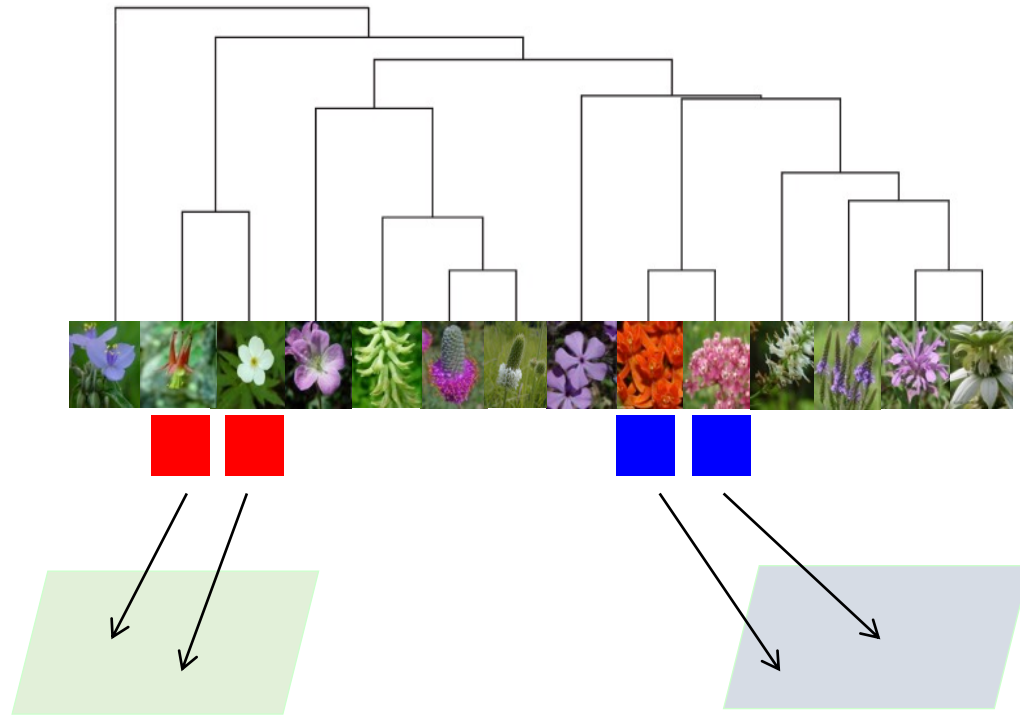
$$\sigma_{\text{attract}} = 0.096 \quad (P < 0.001)$$

Significant phylogenetic attraction:
related species are more likely to have
high abundance in the same site

Phylogenetic attraction: related species occur in the same sites

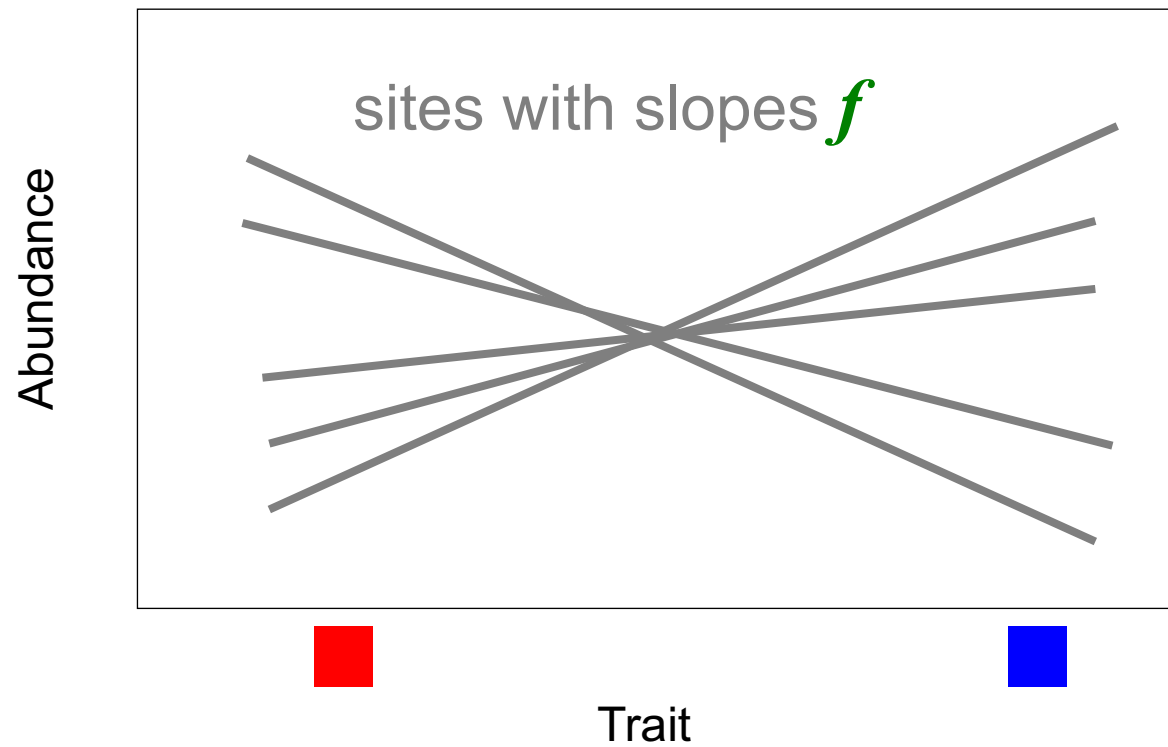


Are there traits that explain the phylogenetic attraction?



$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + f_{\text{site}[i]} \text{trait}_{\text{spp}[i]} + e_i$$

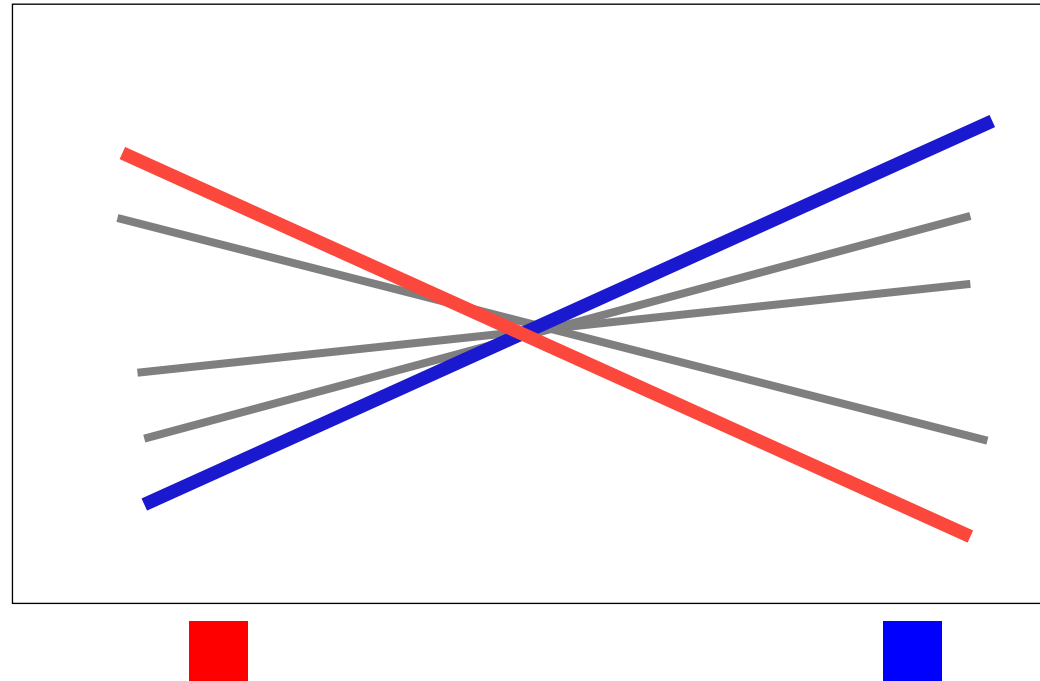
$$\mathbf{f} \sim \text{Gaussian}(\boldsymbol{\beta}, \sigma^2_{\text{trait}} \mathbf{I}_m)$$



$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + f_{\text{site}[i]} \text{trait}_{\text{spp}[i]} + e_i$$

$$\mathbf{f} \sim \text{Gaussian}(\boldsymbol{\beta}, \sigma^2_{\text{trait}} \mathbf{I}_m)$$

site favors
species
with ■

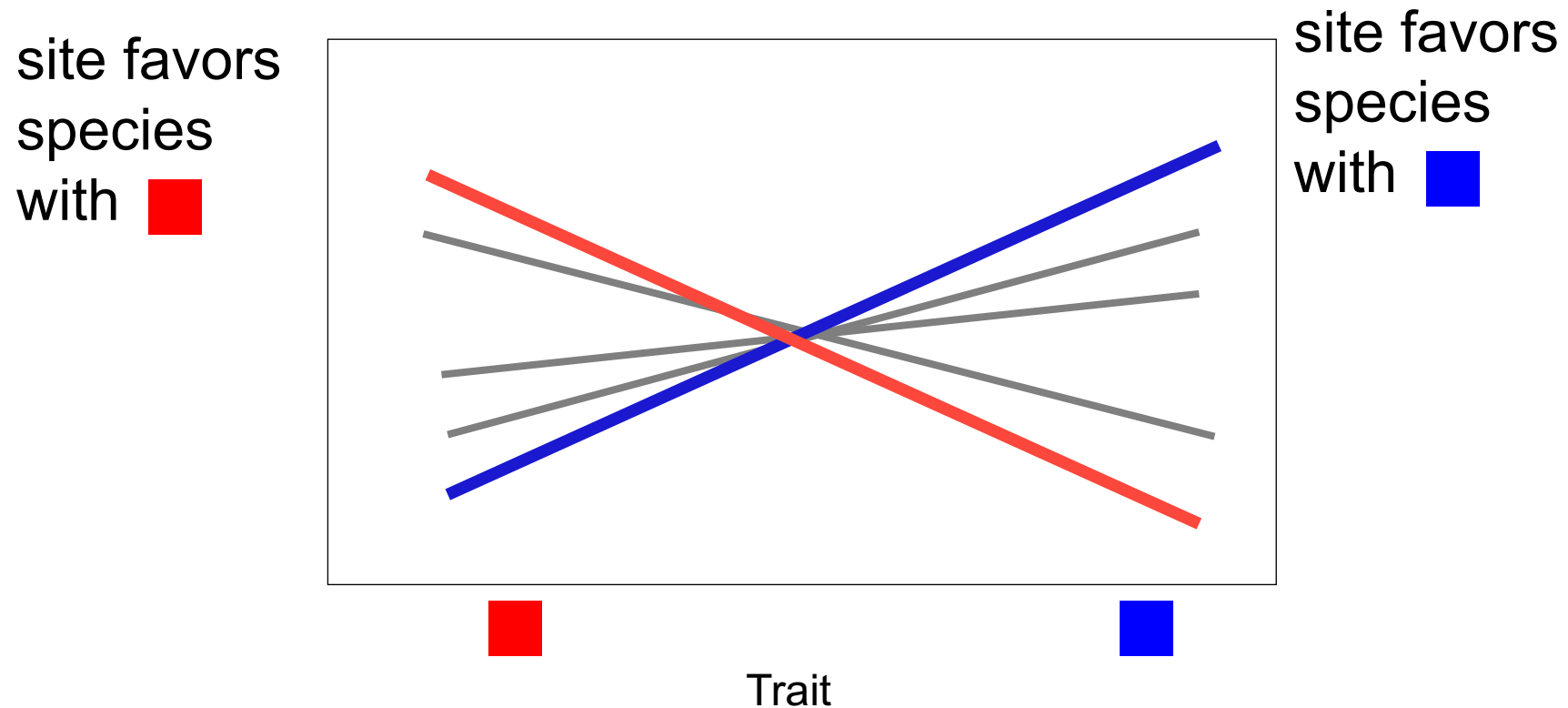


site favors
species
with ■

Trait

$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + f_{\text{site}[i]} \text{trait}_{\text{spp}[i]} + e_i$$

Do traits explain (remove) the effect of phylogeny?



$$Y_i = a_{\text{spp}[i]} + a^p_{\text{spp}[i]} + c_{\text{site}[i]} + d_i + f_{\text{site}[i]} \text{trait}_{\text{spp}[i]} + e_i$$

$$\sigma_{\text{leaf width}} = 0.076$$

$$\sigma_{\text{leaf circularity}} = 0.29$$

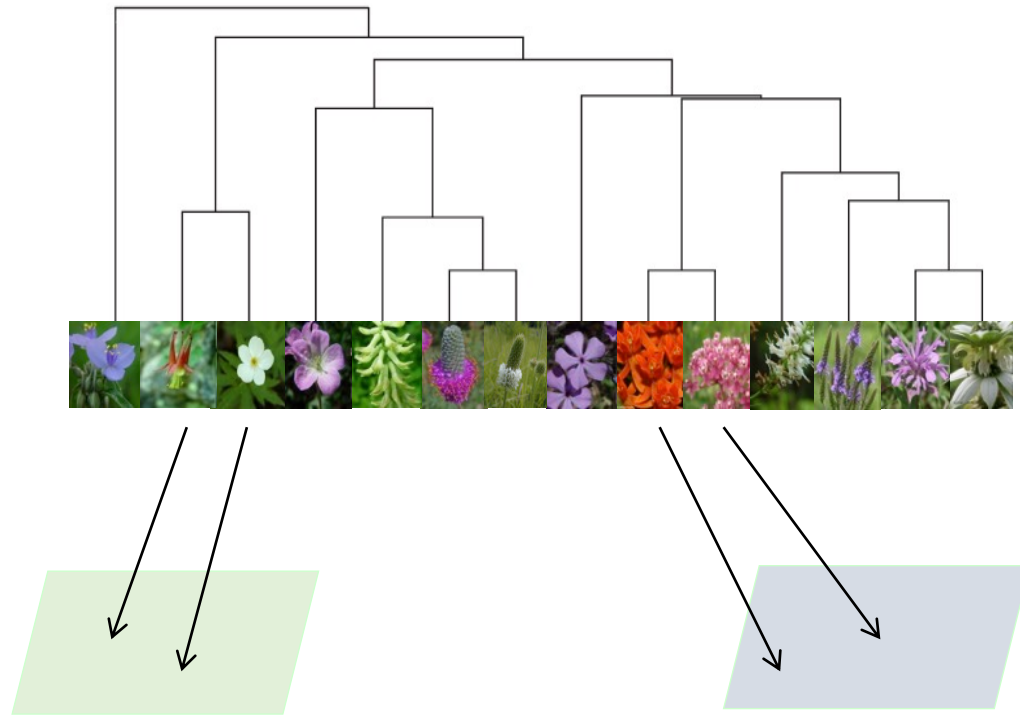
$$\sigma_{\text{leaf thickness}} = 0.0006$$

$$\sigma_{\text{dispersal}} = 0.087$$

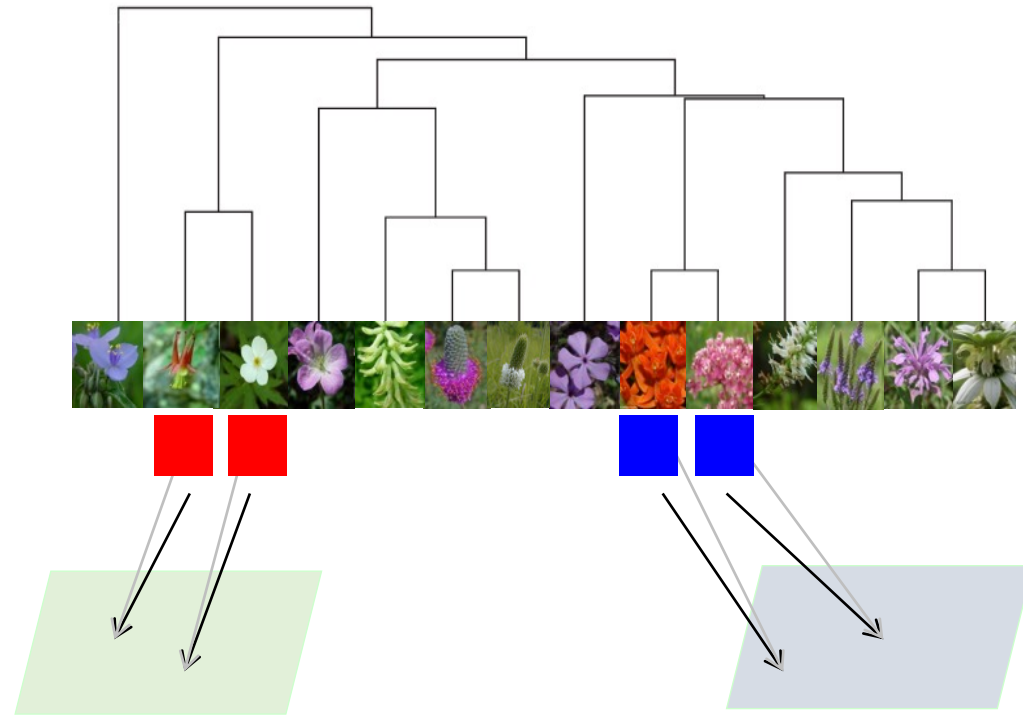
$$\sigma_{\text{attract}} = 0.00016$$

There is almost no remaining phylogenetic attraction

Are there traits that explain the phylogenetic attraction?



Are there traits that explain the phylogenetic attraction?



Yes

Going through R code