

Accounting for intraspecific variation in continuous trait evolution on a reticulate phylogeny

Benjamin Teo, Jeffrey P. Rose, Paul Bastide, Cécile Ané

June 13, 2022

Ignoring intraspecific variation adversely affects comparative methods.

Problem:

Comparative methods that account for intraspecific variation in trees exist **but network methods are still in their infancy**.

Our approach:

Extend the linear model in `PhyloNetworks.jl` for a **continuous response** to account for intraspecific variation assuming **equal intraspecific variances**.

We apply our extension to study how morphology and ecology relate in *Polemonium*.

bioRxiv doi: <https://doi.org/10.1101/2022.05.12.490814>

Comparative methods

Regression model:
$$\underbrace{y}_{\text{species response}} = \underbrace{x'\beta}_{\text{linear function of species predictors}} + \underbrace{\epsilon}_{\text{residual error}}$$

ϵ : Brownian motion

From Bastide et al. (2018)

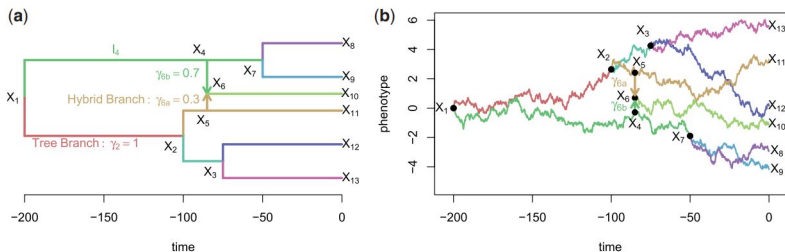


FIGURE 1. Realization of a BM (with $\mu=0$ and $\sigma^2=0.04$) on a calibrated network. Only tip values are observed (here at time $t=0$). For simplicity, the two hybrid branches were chosen to have a length of 0, but this need not be the case. Inheritance probabilities at the hybridization event are γ_{6a} and γ_{6b} , with $\gamma_{6a} + \gamma_{6b} = 1$. a) A time-calibrated phylogenetic network. b) BM on the branches of the network.

Brownian motion on a network

Tree node

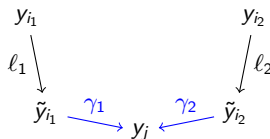


$$y_j \sim \mathcal{N}(y_i, \sigma_b^2 \ell)$$

σ_b^2 : variance-rate

ℓ : branch length

Hybrid node: weighted mean of parents



$$y_j = \gamma_1 \tilde{y}_{i1} + \gamma_2 \tilde{y}_{i2}$$

γ_1, γ_2 : inheritance weights/probabilities

Modeling intraspecific variation

Standard extension

$$\begin{aligned} y_i &= \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \\ &\downarrow \\ y_{ij} &= \underbrace{\mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i}_{\text{species response}} + \underbrace{\delta_{ij}}_{\text{intra}} \end{aligned}$$

$\mathbf{x}_i \approx \bar{\mathbf{x}}_i$: ignore intraspecific variation in \mathbf{x}

Want to estimate: $\boldsymbol{\beta}$, σ_b^2

“Plug-in” approach

Assume $\delta_{ij} \sim \mathcal{N}(0, \hat{\sigma}_{w,i}^2)$

Our approach

Assume $\delta_{ij} \sim \mathcal{N}(0, \sigma_w^2)$

Jointly estimate σ_w^2 , σ_b^2 , then $\boldsymbol{\beta}$ **OR**
all 3 parameters simultaneously

Simulation model

For individual j in species i

Predictor: $x_{ij} = \mathbf{x}_i + \delta_{x,ij}$

$$\mathbf{x}_i \sim \text{BM}$$

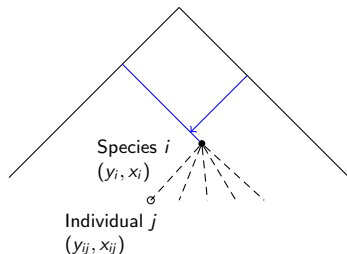
$$\delta_{x,ij} \sim \mathcal{N}(0, \sigma_{w,x}^2)$$

$$\text{Response: } y_{ij} = \underbrace{\beta_1 \mathbf{x}_i + \epsilon_i}_{y_i} + \underbrace{\beta_2 \delta_{x,ij} + \delta_{y,ij}}_{\substack{\text{pheno} \\ \text{correlated}}}$$

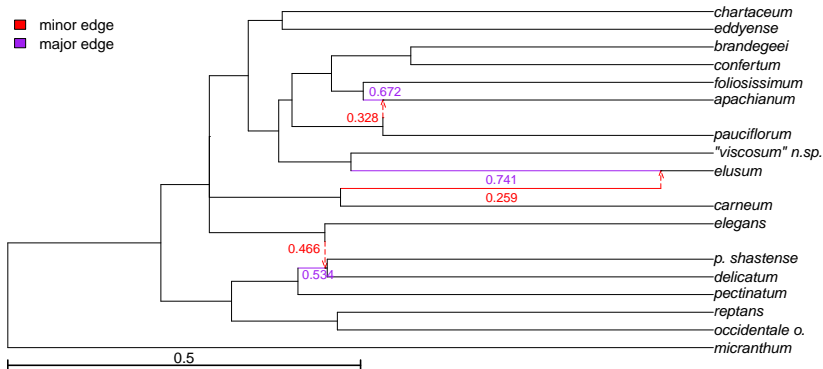
β_1 : phylo, β_2 : pheno

$$\epsilon \sim \text{BM}$$

$$\delta_{y,ij} \sim \mathcal{N}(0, \sigma_{w,y}^2)$$



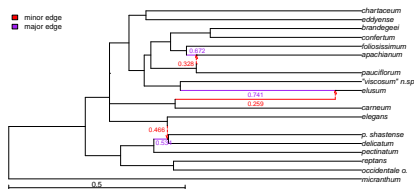
Simulation network



Polemonium



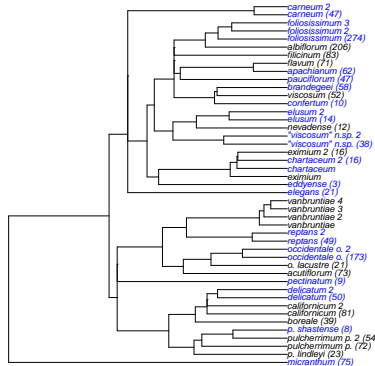
"*Polemonium reptans*" by Ryan Kaldari



17-taxon calibrated network

Phylogenies inferred from DNA in Rose et al. 2021 using SNaQ or ASTRAL.

Calibrated in Teo et al. 2022 following approach in Bastide et al. 2018.



45-taxon calibrated tree

Simulation results

① Unequal intraspecific variances for y , $\sigma_{w,y}^2$

Divide species into 2 groups $\begin{cases} \text{low : } \sigma_{w,y}^2 = \sigma_{lo}^2 \\ \text{high : } \sigma_{w,y}^2 = \sigma_{hi}^2 \end{cases}$

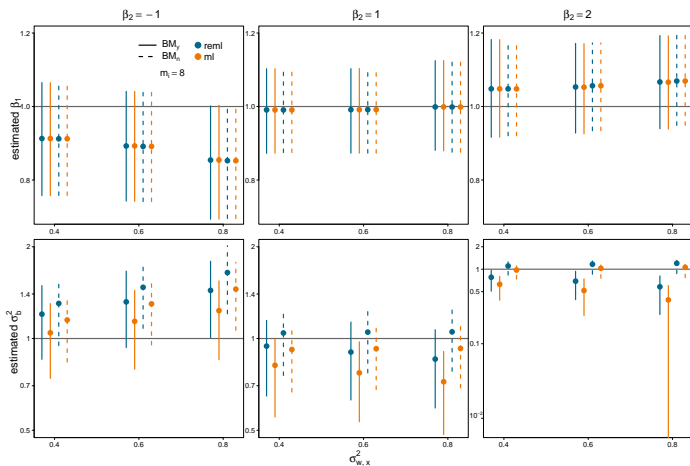
Within-species variation for y only: (y_{ij}, x_i)

$\hat{\beta}_1$ and $\hat{\sigma}_b^2$ are robust to increasing the ratio $\sigma_{hi}^2/\sigma_{lo}^2$.

REML is more accurate than ML for $\hat{\sigma}_b^2$.

Still applicable even if unequal intraspecific variances suspected.

2 Correlated intraspecific variation between x and y ($\beta_2 \neq 0$)

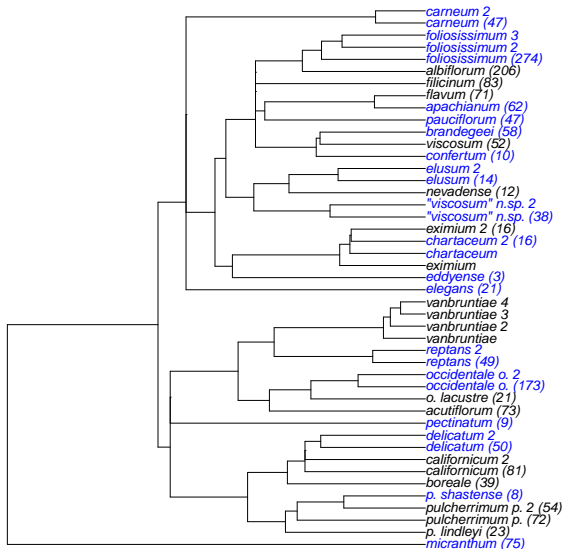


Both $\hat{\beta}_1$ and $\hat{\sigma}_b^2$ can be upwards or downwards biased.

We did theory to understand how this bias is induced by β_1 and β_2 .

Possible post-hoc correction?

Polemonium leaflets



"*Polemonium acutiflorum*" by
Walter Siegmund CC BY-SA 3.0

30 "morphs"

Several tips per morph

y = leaflet size

Sample size: 3 – 274

x = elevation, latitude

Sample size: 3 – 1300+

Findings

Accounting for intraspecific variation

$$\log(\text{length}_{ij}) = \beta_0 + \beta_{\text{elev}} \cdot \text{elevation}_i + \beta_{\text{lat}} \cdot \text{latitude}_i + \epsilon_i + \delta_{ij}$$

① Small leaflets correlate with high elevation and high latitude

Stronger evidence for elevation than latitude

Consistent with ecological literature

② Gene flow explains residual variation

AIC(network) marginally better than AIC(major/minor tree): $|\Delta\text{AIC}| < 1$

Morphological signal is (tenuously) consistent with genetic signal

Single continuous response has low information about inheritance weights

Ignoring intraspecific variation and fitting Pagel's- λ

$$\log(\text{length}_i) = \beta_0 + \beta_{\text{elev}} \cdot \text{elevation}_i + \beta_{\text{lat}} \cdot \text{latitude}_i + \epsilon_i$$

3 ML vs REML can strongly influence conclusions on phylogenetic signal λ

Network (17 taxa): $\hat{\lambda}_{\text{ML}} = 0$ (no signal) and $\hat{\lambda}_{\text{REML}} \approx 1$ (full signal)

Tree (30 taxa): $\hat{\lambda}_{\text{ML}} = \hat{\lambda}_{\text{REML}} = 1$

More taxa with a tree is more powerful than less taxa with a network

Acknowledgements/Funding

This work was supported in part by

- ◇ National Science Foundation (DMS-1902892 and DMS-2023239)
- ◇ NSF doctoral dissertation improvement grant (DEB 1501867) to JPR

We also thank Cathy Cao for technical help in setting up the simulations.

Bibliography



Bastide, Paul et al. (2018). “Phylogenetic comparative methods on phylogenetic networks with reticulations”. In: *Systematic Biology* 67.5, pp. 800–820. DOI: [10.1093/sysbio/syy033](https://doi.org/10.1093/sysbio/syy033).



Rose, Jeffrey P et al. (2021). “Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal”. In: *Systematic Biology* 70.1, pp. 162–180. DOI: [10.1093/sysbio/syaa049](https://doi.org/10.1093/sysbio/syaa049).



Teo, Benjamin Sean et al. (2022). “Accounting for intraspecific variation in continuous trait evolution on a reticulate phylogeny”. In: *bioRxiv*.