

Phylogenetic Comparative Methods on Networks

Paul Bastide¹, Claudia Solís-Lemus², Ricardo Kriebel³,
K. William Sparks³, Cécile Ané^{3,4}

¹ Evolutionary and Computational Virology, Rega Institute, KU Leuven, Belgium

² Department of Human Genetics, Emory University, Atlanta, GA, 30322, USA

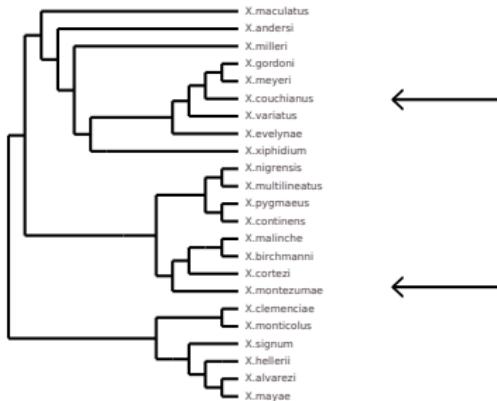
³ Department of Botany, University of Wisconsin–Madison, WI, 53706, USA

⁴ Department of Statistics, University of Wisconsin–Madison, WI, 53706, USA

28 September 2018



Phylogenetic Networks



X. Couchianus

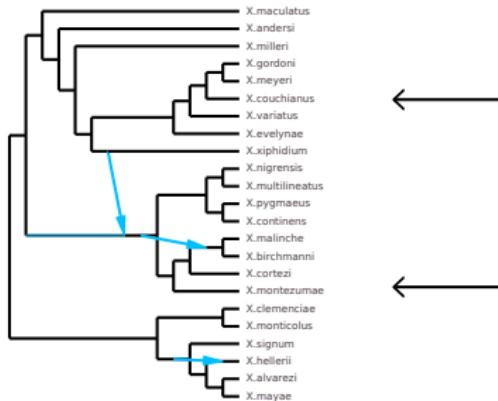


X. Montezumae

Measured Traits

- Sword index
- Female preference

Phylogenetic Networks



X. Couchianus

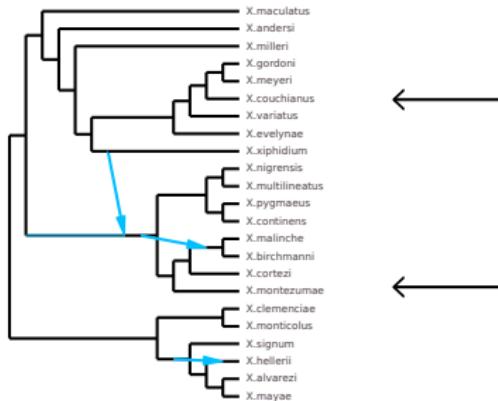


X. Montezumae

Measured Traits

- Sword index
- Female preference

Phylogenetic Networks



X. Couchianus



X. Montezumae

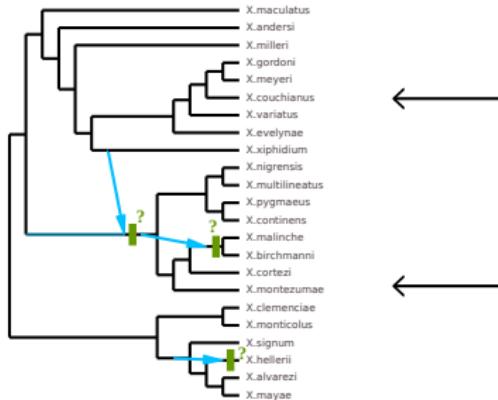
Measured Traits

- Sword index
- Female preference

Questions:

- PCM for a **network** structure ?

Phylogenetic Networks



X. Couchianus



X. Montezumae

Measured Traits

- Sword index
- Female preference

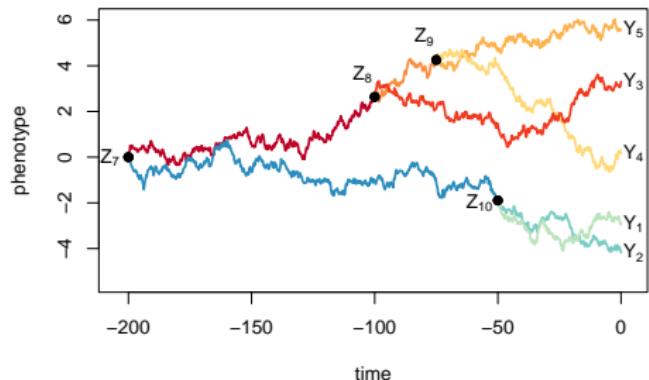
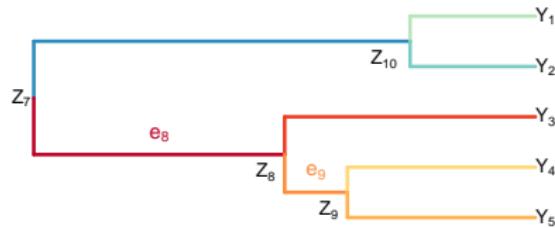
Questions:

- PCM for a **network** structure ?
- Can we detect ancestral **transgressive evolution** ?

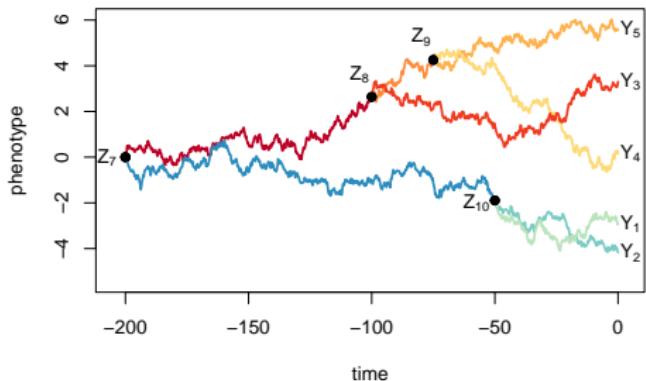
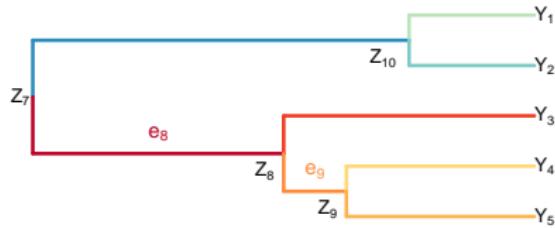
Outline

- ① Trait Evolution on Networks
- ② Transgressive Evolution
- ③ Remarks, Perspectives and Advertising

BM on a Tree



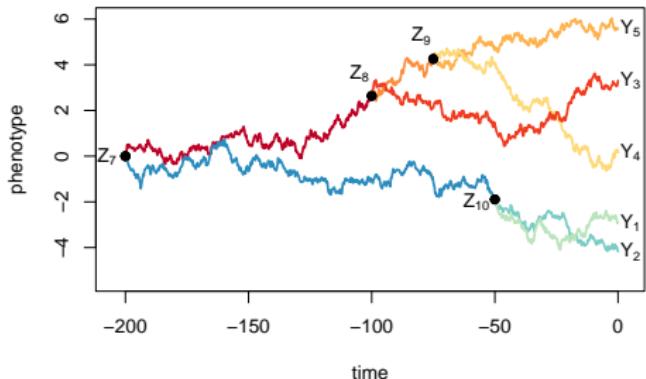
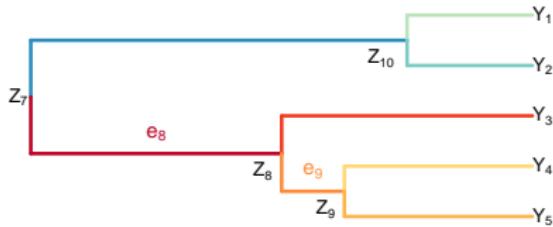
BM on a Tree



BM Variance: Shared Evolution Time

$$C_{45} = \ell_{e_8} + \ell_{e_9}$$

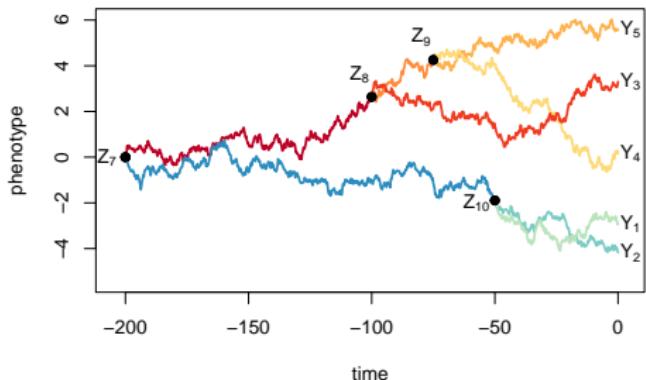
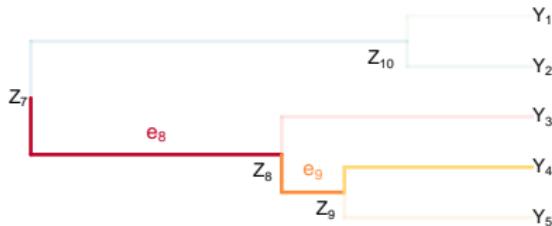
BM on a Tree



BM Variance: Shared Evolution Time

$$C_{45} = \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e$$

BM on a Tree



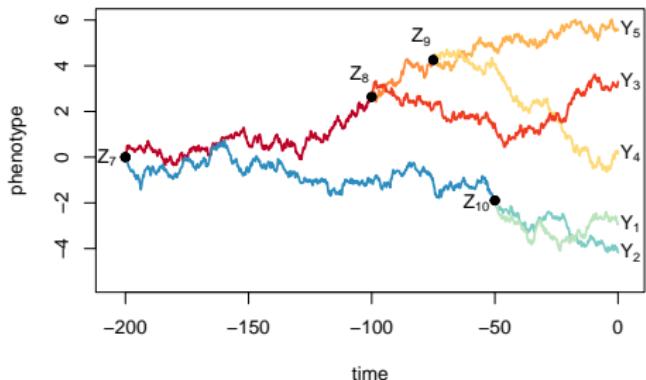
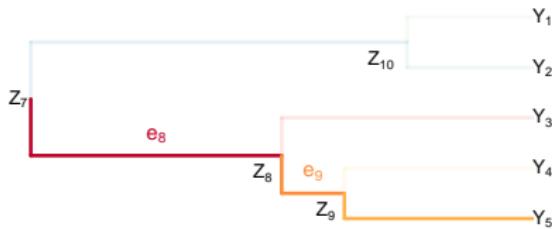
BM Variance: Shared Evolution Time

$$C_{45} = \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e$$

p_i : path from root to tip i :

$$p_4 = \{e_8, e_9, e_4\}$$

BM on a Tree



BM Variance: Shared Evolution Time

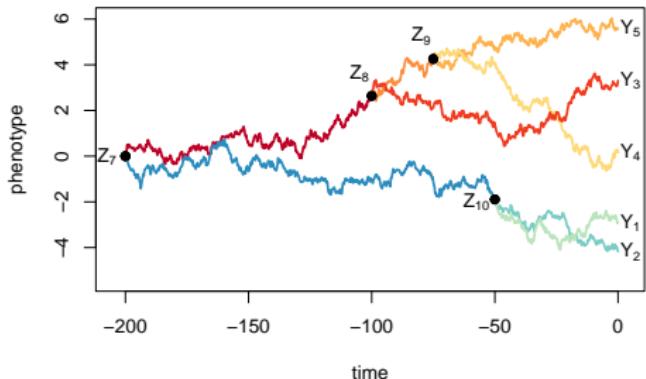
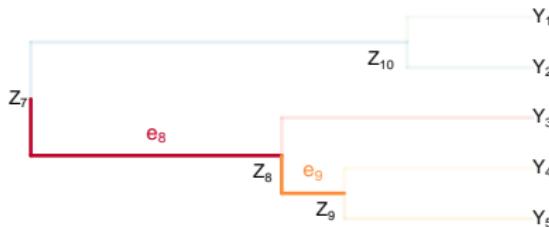
$$C_{45} = \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e$$

p_i : path from root to tip i :

$$p_4 = \{e_8, e_9, e_4\}$$

$$p_5 = \{e_8, e_9, e_5\}$$

BM on a Tree



BM Variance: Shared Evolution Time

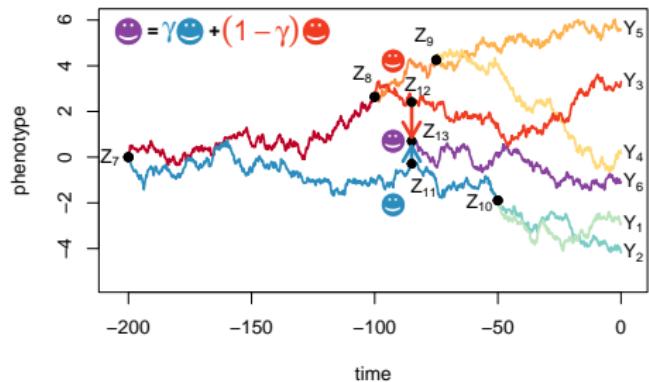
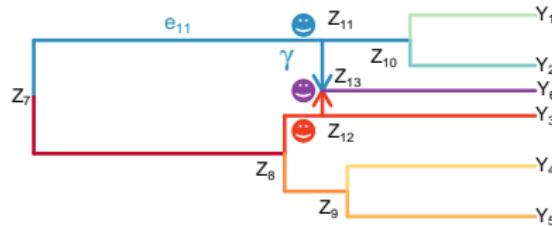
$$\begin{aligned} C_{45} &= \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e \\ &= \sum_{e \in p_4 \cap p_5} \ell_e \end{aligned}$$

p_i : path from root to tip i :

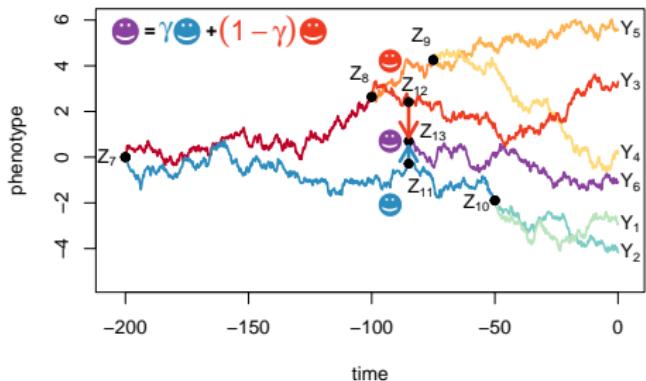
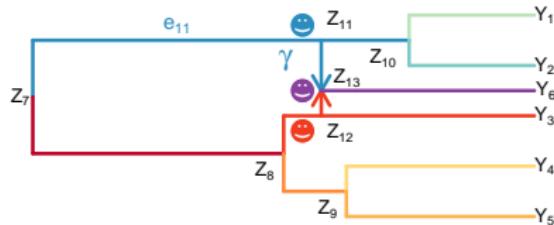
$$p_4 = \{e_8, e_9, e_4\}$$

$$p_5 = \{e_8, e_9, e_5\}$$

BM on a Network



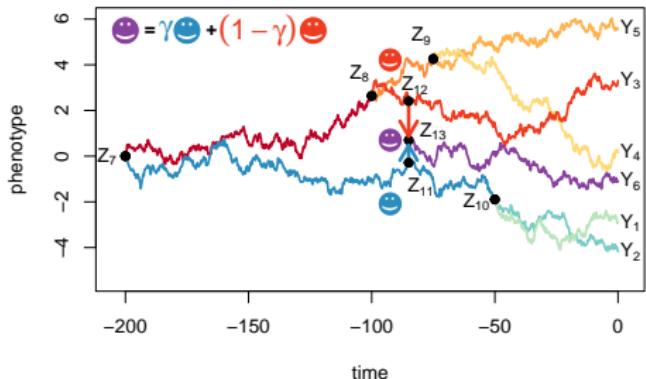
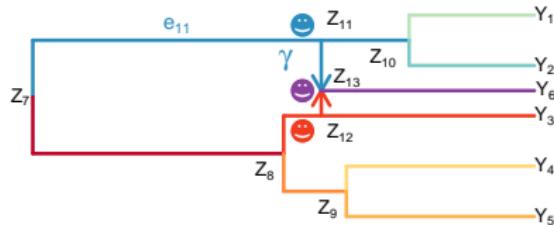
BM on a Network



BM Variance: “Shared Evolution Time”

$$C_{ij}^{\text{tree}} = \sum_{e \in p_i \cap p_j} \ell_e$$

BM on a Network

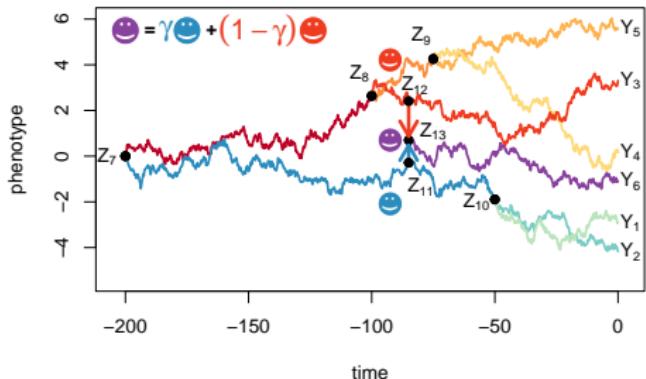
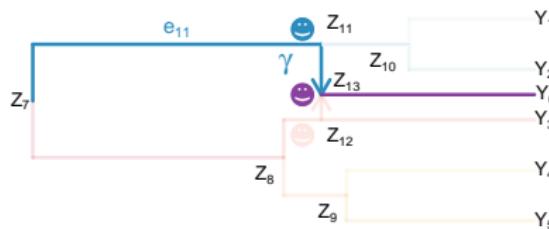


BM Variance: “Shared Evolution Time”

Problem Path to Y_6 not unique

$$C_{ij}^{\text{tree}} = \sum_{e \in p_i \cap p_j} \ell_e$$

BM on a Network

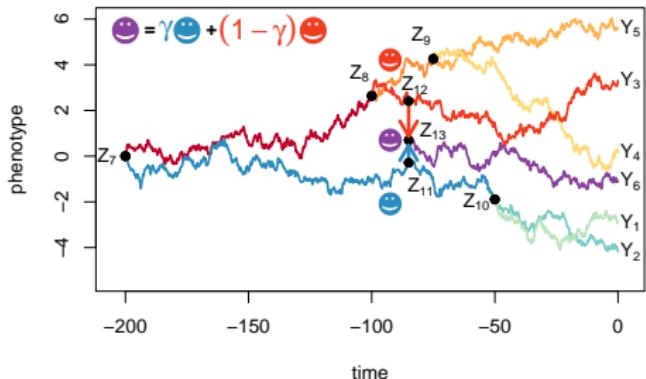
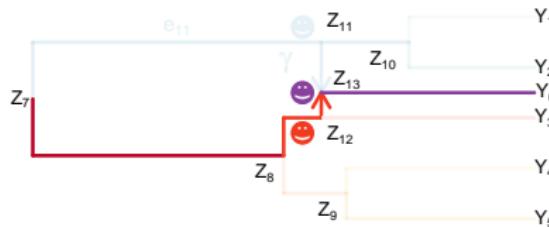


BM Variance: “Shared Evolution Time”

Problem Path to Y_6 not unique

$$\mathcal{P}_6 = \left\{ \{e_{11}, e_{13}, e_6\}, \quad \right\}$$

BM on a Network

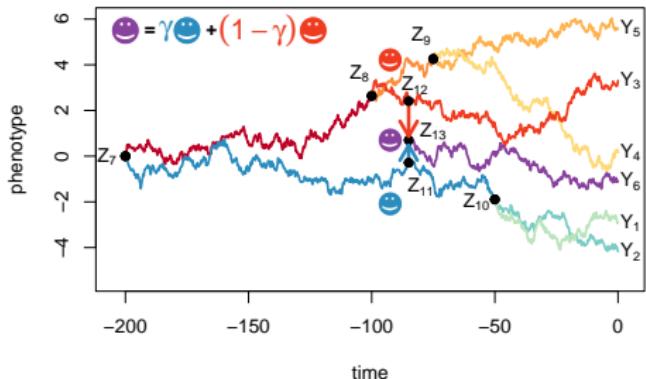
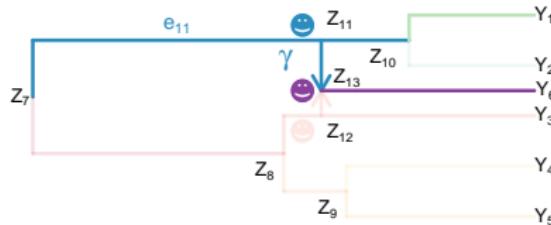


BM Variance: “Shared Evolution Time”

Problem Path to Y_6 not unique

$$\mathcal{P}_6 = \left\{ \begin{array}{l} \{e_{11}, e_{13}, e_6\}, \\ \{e_8, e_{12}, e_{12}, e_6\} \end{array} \right\}$$

BM on a Network

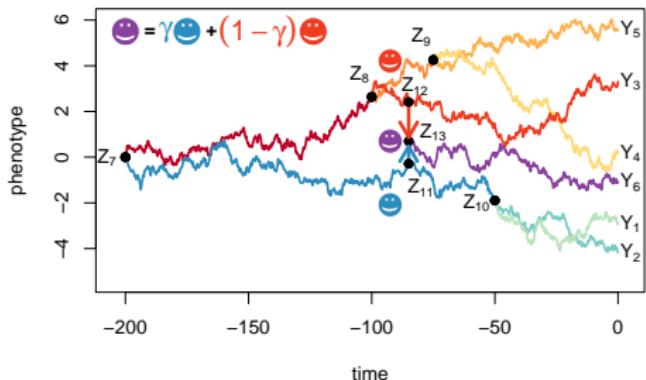
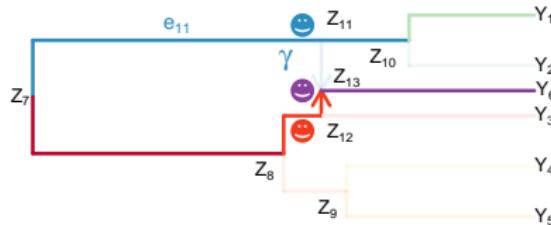


BM Variance: “Shared Evolution Time”

$$C_{16} = \ell_{e_{11}} \times \gamma$$

$$\mathcal{P}_6 = \left\{ \{e_{11}, e_{13}, e_6\}, \quad \right\}$$

BM on a Network

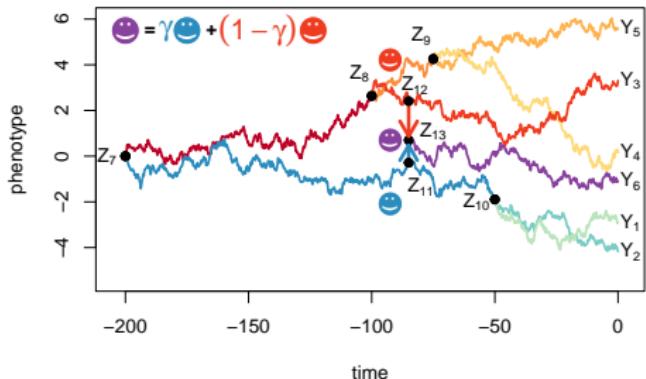
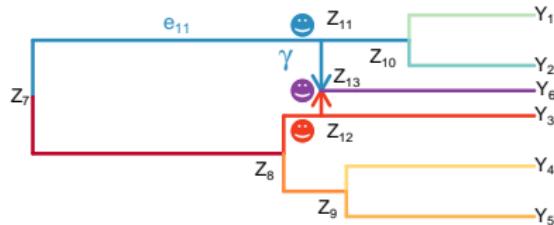


BM Variance: “Shared Evolution Time”

$$\begin{aligned} C_{16} = & \ell_{e_{11}} \times \gamma \\ & + 0 \times (1 - \gamma) \end{aligned}$$

$$\mathcal{P}_6 = \left\{ \begin{array}{l} \{e_{11}, e_{13}, e_6\}, \\ \{e_8, e_{12}, e_{12}, e_6\} \end{array} \right\}$$

BM on a Network



BM Variance: “Shared Evolution Time”

$$C_{16} = \ell_{e_{11}} \times \gamma \\ + 0 \times (1 - \gamma)$$

$$C_{ij}^{\text{net}} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e \right) \left(\prod_{e \in p_j} \gamma_e \right) \sum_{e \in p_i \cap p_j} \ell_e$$

Note: Variance reduction

Path formula:

$$C_{ij}^{\text{net}} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e \right) \left(\prod_{e \in p_j} \gamma_e \right) \sum_{e \in p_i \cap p_j} \ell_e$$

Variance reduction:

$$\begin{aligned} C_{ii}^{\text{net}} &\leq \sum_{p_i, q_i \in \mathcal{P}_i} \pi_{p_i} \pi_{q_i} \sum_{e \in p_i} \ell_e \\ &\leq \sum_{p_i, q_i \in \mathcal{P}_i} \pi_{p_i} \pi_{q_i} t_i \\ &\leq t_i = C_{ii}^{\text{tree}} \end{aligned}$$

Phylogenetic Linear Regression

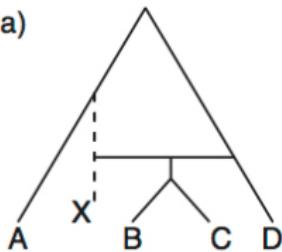
$$\mathbf{Y} = \mu \mathbf{1}_n + \mathbf{R}\boldsymbol{\theta} + \sigma \mathbf{E} \quad \text{with} \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{C}^{\text{net}})$$

- Core of PCM methods
- Account for network correlations in a regression
- Cf function `phylolm` in R
 - `phylolNetworklm` in Julia !

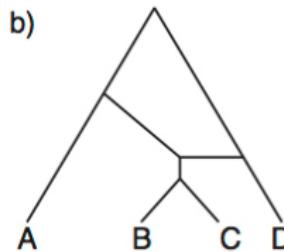
Extinctions

(Degnan, 2018)

a)

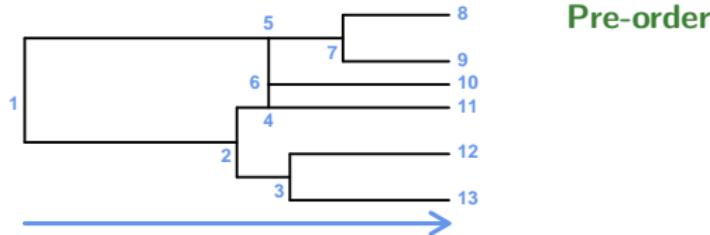


b)

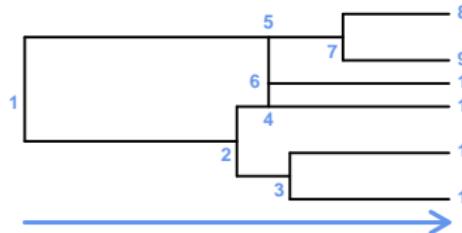


- Hybridization is instantaneous.
- Bad sampling gives it a branch length.

Efficient Computation



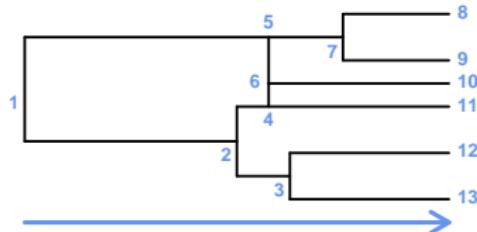
Efficient Computation



Pre-order

Root: $C_{11} = 0$

Efficient Computation



Pre-order

Root: $C_{11} = 0$

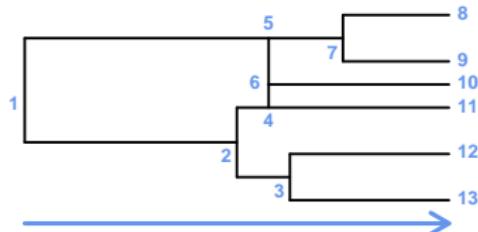
Tree node i with parent a

$$\begin{cases} C_{ij} = C_{aj} & j < i \\ C_{ii} = C_{aa} + \ell_a & \end{cases}$$



$$X_i = X_a + \epsilon_a \quad \epsilon_a \sim \mathcal{N}(0, \ell_a)$$

Efficient Computation

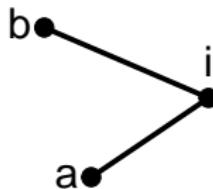


Pre-order

$$\text{Root: } C_{11} = 0$$

Tree node i with parent a

$$\begin{cases} C_{ij} = C_{aj} & j < i \\ C_{ii} = C_{aa} + \ell_a & \end{cases}$$

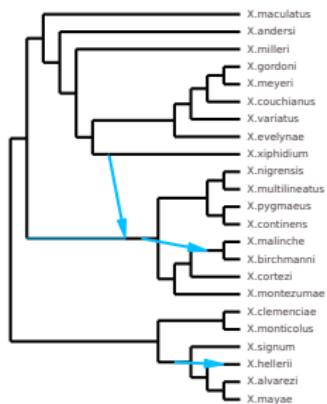


Hybrid node i with parents a and b

$$\begin{cases} C_{ij} = \gamma_a C_{aj} + \gamma_b C_{bj} & j < i \\ C_{ii} = \gamma_a^2(C_{aa} + \ell_a) + \gamma_b^2(C_{bb} + \ell_b) \\ \quad + 2\gamma_a\gamma_b C_{ab} & \end{cases}$$

$$\begin{aligned} X_i &= \gamma_a(X_a + \epsilon_a) & \epsilon_a &\sim \mathcal{N}(0, \ell_a) \\ &+ \gamma_b(X_b + \epsilon_b) & \epsilon_b &\sim \mathcal{N}(0, \ell_b) \end{aligned}$$

Xiphophorus Fish Dataset

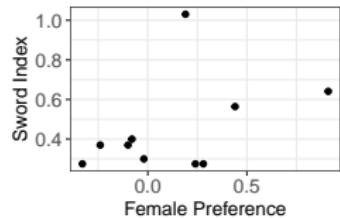


X. Couchianus



X. Montezumae

Two traits:



using PhyloNetworks

```
net = readTopology("xiphophorus_network.newick");
dat = CSV.read("xiphophorus_data.csv");
```

Xiphophorus Fish Dataset

```
fit = phyloNetworklm(@formula(sword_index ~ preference), dat, net)
```

Formula: sword_index ~ 1 + preference

Model: BM

Parameter(s) Estimates:

Sigma2: 0.0041735

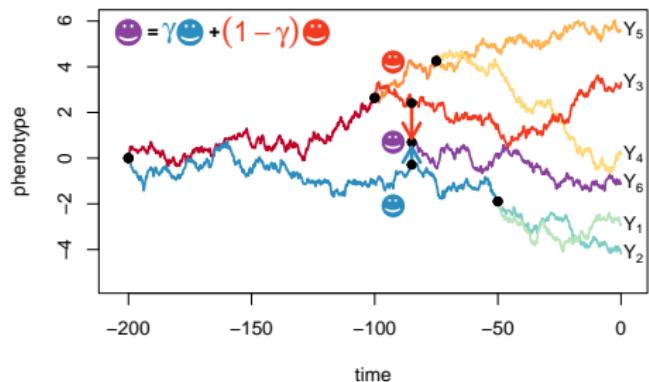
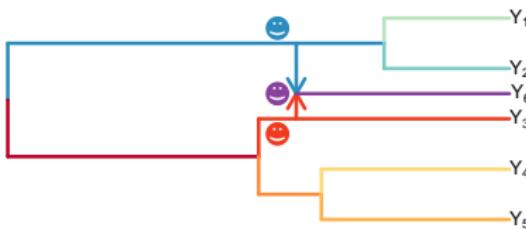
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.298191	0.17724	1.68241	0.1310
preference	0.55627	0.305627	1.8201	0.1062

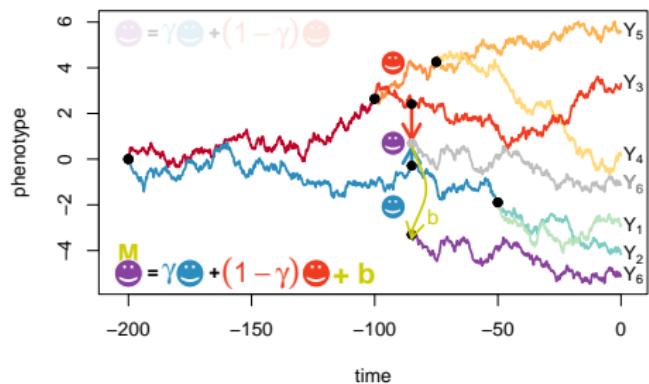
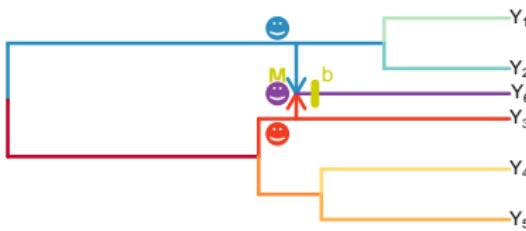
Log Likelihood: 1.1366987194

AIC: 3.7266025612

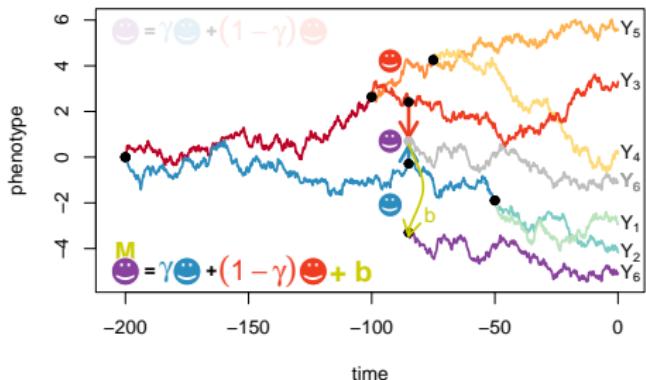
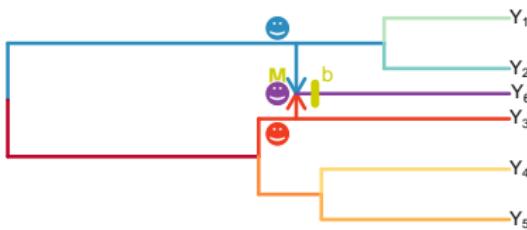
Transgressive Evolution



Transgressive Evolution



Transgressive Evolution



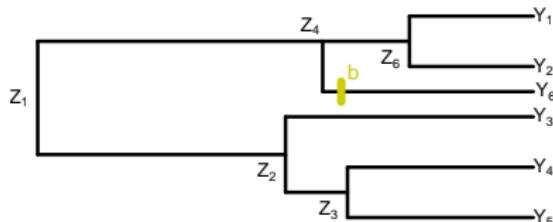
Idea: Write as a linear model

$$\mathbf{Y} = \mu \mathbf{1} + \mathbf{Nb} + \sigma \mathbf{E}$$

$$\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

Variance unchanged

Linear Regression Model



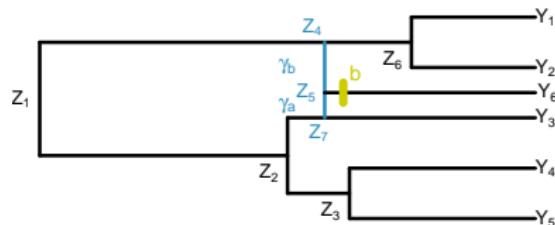
$$\Delta = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_6 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} \quad T\Delta = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}$$

μ μ
 μ μ
 μ μ
 μ μ
 μ μ
 $\mu + b$

$$T = \begin{pmatrix} Z_1 & Z_2 & Z_3 & Z_4 & Z_6 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \\ Y_1 & 1 & . & . & 1 & 1 & 1 & . & . & . & . \\ Y_2 & 1 & . & . & 1 & 1 & . & 1 & . & . & . \\ Y_3 & 1 & 1 & . & . & . & . & . & 1 & . & . \\ Y_4 & 1 & 1 & 1 & . & . & . & . & . & 1 & . \\ Y_5 & 1 & 1 & 1 & . & . & . & . & . & . & 1 \\ Y_6 & 1 & . & . & 1 & . & . & . & . & . & 1 \end{pmatrix}$$

$$\mathbb{E}[\mathbf{Y}] = T\Delta$$

Linear Regression Model



$$\Delta = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} \quad T\Delta = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}$$

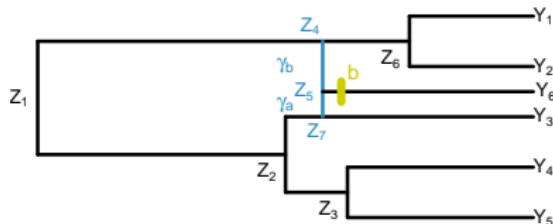
$\begin{pmatrix} \mu \\ \vdots \\ 0 \\ \vdots \\ \cdot \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ \mu + b \end{pmatrix}$

$$T = \begin{pmatrix} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 & Z_6 & Z_7 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \\ Y_1 & 1 & \cdot & \cdot & 1 & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ Y_2 & 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ Y_3 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ Y_4 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ Y_5 & 1 & 1 & 1 & \cdot & 1 & \cdot \\ Y_6 & 1 & \cdot & \cdot & \gamma_b & 1 & \cdot & \gamma_a & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

$$\mathbb{E}[\mathbf{Y}] = T\Delta$$

$$T_{ij} = \sum_{p \in \mathcal{P}_{j \rightarrow i}} \prod_{e \in p} \gamma_e$$

Linear Regression Model



$$\Delta = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}$$

$$T\Delta = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}$$

$$\begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \cdot \\ \textcolor{red}{b} \end{pmatrix}$$

$$T = \begin{pmatrix} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 & Z_6 & Z_7 & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & \textcolor{red}{Y_6} \\ Y_1 & 1 & \cdot & \cdot & 1 & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ Y_2 & 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ Y_3 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot \\ Y_4 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ Y_5 & 1 & 1 & 1 & \cdot & 1 & \cdot \\ Y_6 & 1 & \cdot & \cdot & \gamma_b & 1 & \cdot & \gamma_a & \cdot & \cdot & \cdot & \cdot & \textcolor{red}{1} \end{pmatrix}$$

$$\mathbb{E}[\mathbf{Y}] = \mu \mathbf{1} + \mathbf{Nb}$$

$$T_{ij} = \sum_{p \in \mathcal{P}_{j \rightarrow i}} \prod_{e \in p} \gamma_e$$

Transgressive Evolution: Testing Effect(s)

Model:

$$\mathbf{Y} = \mu \mathbf{1} + \mathbf{Nb} + \sigma \mathbf{E} \quad , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

Tests:

- | | |
|---|-------------------------------|
| \mathcal{H}_0 : No TE | $\mathbf{b} = \mathbf{0}$ |
| \mathcal{H}_1 : TE with one single effect | $\mathbf{b} = b_1$ |
| \mathcal{H}_2 : TE with heterogeneous effects | $\mathbf{b} \in \mathbb{R}^h$ |

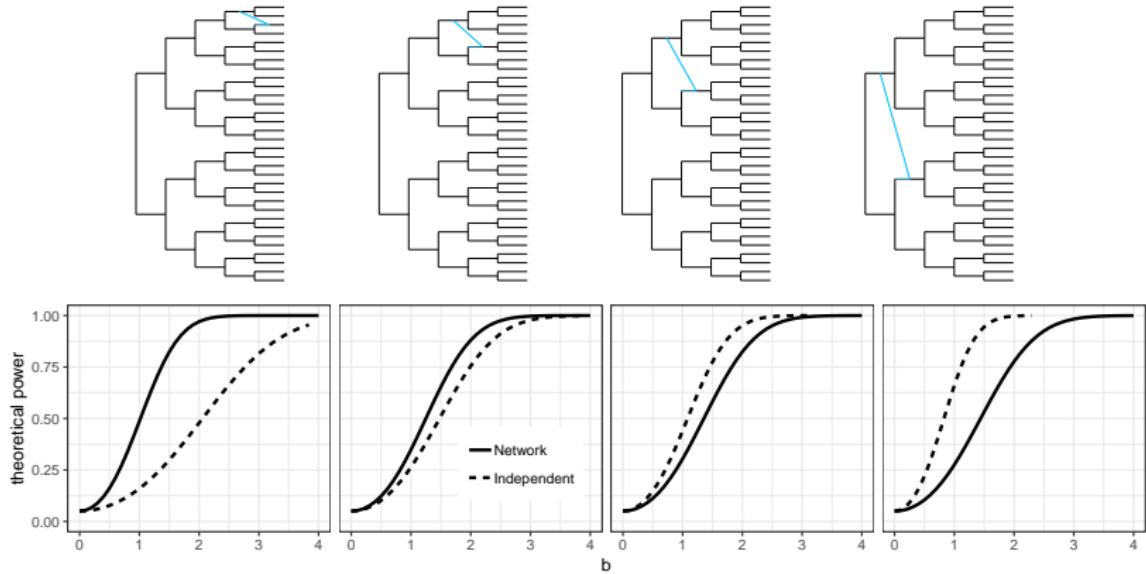
Fisher:

$$F_{10} \sim \mathcal{F}_{1,n-2} (\Delta_{10}(b, \sigma^2))$$

$$F_{21} \sim \mathcal{F}_{h-1,n-h-1} (\Delta_{21}(\mathbf{b}, \sigma^2))$$

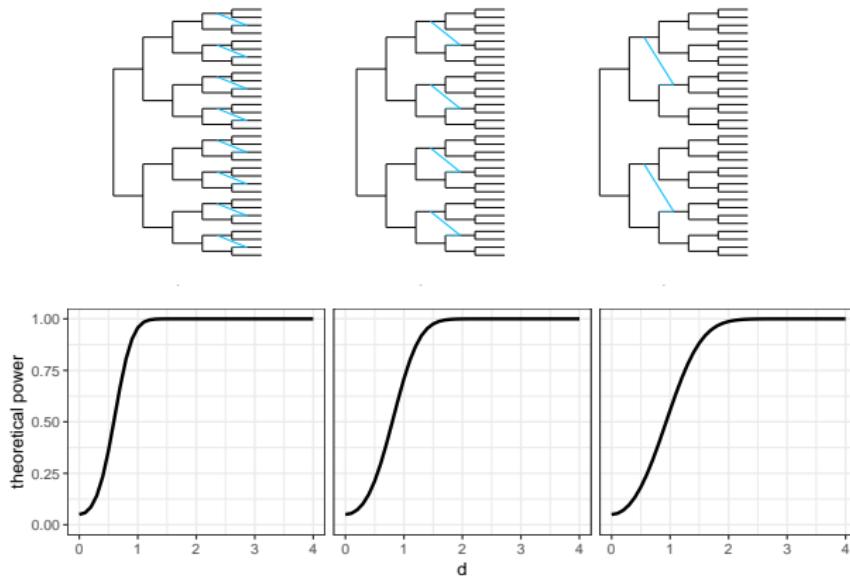


TE: Single Effect



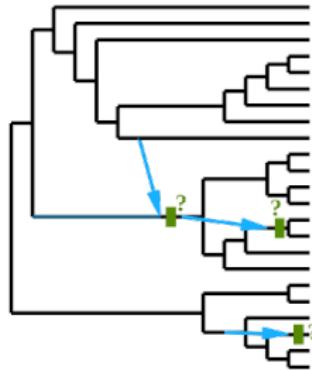
Detection power ($\sigma^2 = 1$, level 0.05)

TE: Several Effects



Detection power ($\sigma^2 = 1$, level 0.05)

Xiphophorus Fish Dataset



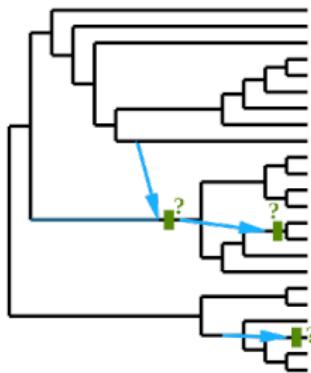
Regressor Matrix N

```
N = regressorHybrid(net);
```

```
dat = join(dat, N, on = :tipNames);
```

○

Xiphophorus Fish Dataset



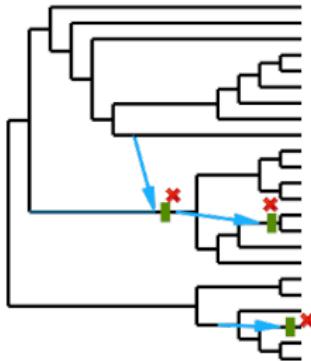
Regressor Matrix N

```
N = regressorHybrid(net);  
  
dat = join(dat, N, on = :tipNames);
```

Sword Index ?

```
fit0 = phylоНetworklm(@formula(sword_index ~ 1), dat, net)  
  
fit1 = phylоНetworklm(@formula(sword_index ~ sum), dat, net)  
  
fit2 = phylоНetworklm(@formula(sword_index ~ shift_24 + shift_37 + shift_45),  
                      dat, net)  
  
ftest(fit2, fit1, fit0)
```

Xiphophorus Fish Dataset



Regressor Matrix \mathbf{N}

```
N = regressorHybrid(net);
```

```
dat = join(dat, N, on = :tipNames);
```

Sword Index ? No Transgressive Evolution

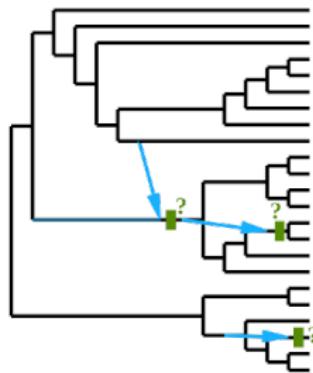
```
fit0 = phylоНetworklm(@formula(sword_index ~ 1), dat, net)

fit1 = phylоНetworklm(@formula(sword_index ~ sum), dat, net)

fit2 = phylоНetworklm(@formula(sword_index ~ shift_24 + shift_37 + shift_45),
                      dat, net)

ftest(fit2, fit1, fit0)
```

Xiphophorus Fish Dataset



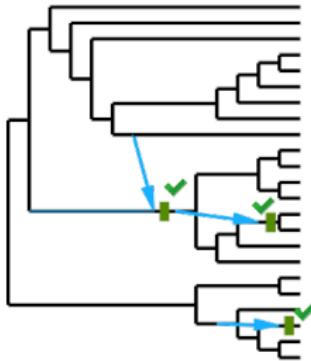
Regressor Matrix N

```
N = regressorHybrid(net);  
  
dat = join(dat, N, on = :tipNames);
```

Female Preference ?

```
fit0 = phylоНetworklm(@formula(preference ~ 1), dat, net)  
  
fit1 = phylоНetworklm(@formula(preference ~ sum), dat, net)  
  
fit2 = phylоНetworklm(@formula(preference ~ shift_24 + shift_37 + shift_45),  
                      dat, net)  
  
ftest(fit2, fit1, fit0)
```

Xiphophorus Fish Dataset



Regressor Matrix \mathbf{N}

```
N = regressorHybrid(net);
```

```
dat = join(dat, N, on = :tipNames);
```

Female Preference ? Heterogeneous TE

```
fit0 = phylоНetworklm(@formula(preference ~ 1), dat, net)
```

```
fit1 = phylоНetworklm(@formula(preference ~ sum), dat, net)
```

```
fit2 = phylоНetworklm(@formula(preference ~ shift_24 + shift_37 + shift_45),
                      dat, net)
```

```
ftest(fit2, fit1, fit0)
```

Julia package PhyloNetworks



“Julia combines the **ease of use of R** with the **speed of C++**.”

Source: Julia press release.

Julia package PhyloNetworks



"Julia combines the **ease of use of R** with the **speed of C++**."

Source: Julia press release.

Solís-Lemus et al. (2017). **PhyloNetworks**: a Package for Phylogenetic Networks. *MBE*, 34(12), 3292–3298.

Manipulation, visualization and inference of phylogenetic networks.

Julia package PhyloNetworks



"Julia combines the **ease of use of R** with the **speed of C++**."

Source: Julia press release.

Solís-Lemus et al. (2017). **PhyloNetworks**: a Package for Phylogenetic Networks. *MBE*, 34(12), 3292–3298.

Manipulation, visualization and inference of phylogenetic networks.



github.com/crs14/PhyloNetworks.jl



Automated documentation with Documenter.jl



Continuous Integration on Travis.



Unitary tests.

Caveats and Perspectives

Remarks

- Power: the more recent the better
- Highly dependent on the network.

Caveats and Perspectives

Remarks

- Power: the more recent the better
- Highly dependent on the network.

Other Features

- Trait simulation
- Ancestral State Reconstruction
- Pagel's λ
- Link with pedigrees



Caveats and Perspectives

Remarks

- Power: the more recent the better
- Highly dependent on the network.

Other Features

- Trait simulation
- Ancestral State Reconstruction
- Pagel's λ
- Link with pedigrees

Perspectives

- Early Burst, Ornstein-Uhlenbeck, ...
- Other merging rules ?

Caveats and Perspectives

Remarks

- Power: the more recent the better
- Highly dependent on the network.

Other Features

- Trait simulation
- Ancestral State Reconstruction
- Pagel's λ
- Link with pedigrees

Perspectives

- Early Burst, Ornstein-Uhlenbeck, ...
- Other merging rules ?

Conclusion

A general framework for trait evolution on networks.

Bastide, Solís-Lemus, Kriebel, Sparks, Ané (2018). Phylogenetic Comparative Methods for Phylogenetic Networks with Reticulations. *Syst. Biol.*

Acknowledgements



Claudia Solís-Lemus



Ricardo Kriebel

Mahendra Mariadassou



K. William Sparks

Stéphane Robin

Cécile Ané



KU LEUVEN



Philippe Lemey's Lab

Bibliography

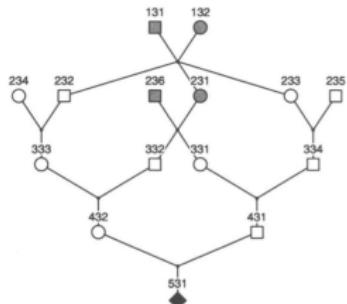
- Bastide P, Solís-Lemus C, Kriebel R, Sparks KW, Ané C. 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Systematic Biology*. 67:800–820.
- Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013. Phylogenomics Reveals Extensive Reticulate Evolution in Xiphophorus Fishes. *Evolution*. 67:2166–2179.
- Degnan JH. 2018. Modeling Hybridization Under the Network Multispecies Coalescent. *Systematic Biology*. 67:786–799.
- Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*. 12:e1005896.
- Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: A Package for Phylogenetic Networks. *Molecular Biology and Evolution*. 34:3292–3298.
- Thompson EA. 2000. Statistical Inference from Genetic Data on Pedigrees, volume 6. Beachwood: Institute of Mathematical Statistics, nsf-cbms r edition.

Photo Credits:

Xiphophorus Genetic Stock Center, Texas State University.

Appendices

Link with Pedigrees Models

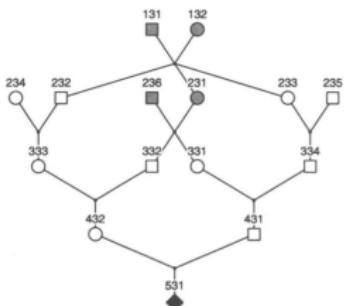


Thompson (2000)

Link with Pedigrees Models

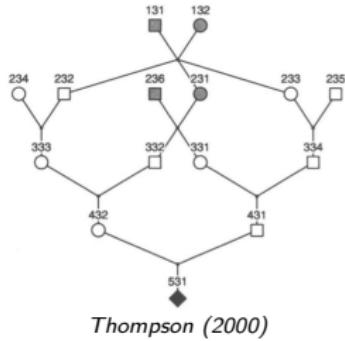
Animal model in Quantitative Genetics

$$X_i = \frac{1}{2}X_a + \frac{1}{2}X_b + \epsilon \quad \epsilon \sim \mathcal{N}(0, 1 - \frac{1}{4}(A_{aa} + A_{bb}))$$



Thompson (2000)

Link with Pedigrees Models

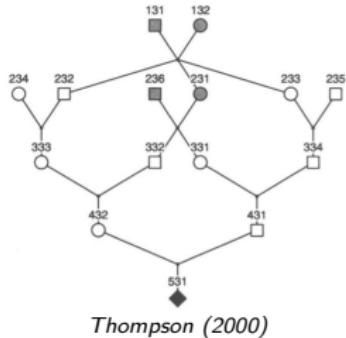


Animal model in Quantitative Genetics

$$X_i = \frac{1}{2}X_a + \frac{1}{2}X_b + \epsilon \quad \epsilon \sim \mathcal{N}(0, 1 - \frac{1}{4}(A_{aa} + A_{bb}))$$

$$\frac{1}{2}A_{ij} = \mathbb{P} \left[\begin{array}{l} \text{two genes from } i \text{ and } j \\ \text{are identical by descent} \end{array} \right] \\ = \text{kinship coefficient}$$

Link with Pedigrees Models



Animal model in Quantitative Genetics

$$X_i = \frac{1}{2}X_a + \frac{1}{2}X_b + \epsilon \quad \epsilon \sim \mathcal{N}(0, 1 - \frac{1}{4}(A_{aa} + A_{bb}))$$

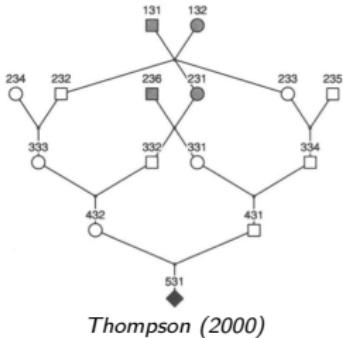
$$\frac{1}{2}A_{ij} = \mathbb{P} \left[\begin{array}{l} \text{two genes from } i \text{ and } j \\ \text{are identical by descent} \end{array} \right]$$

= kinship coefficient

Same as a Network with

- Several roots ("founders")
- $\gamma_e = \frac{1}{2}$ for all edges
- $\ell_{a \rightarrow i} = 2 - A_{aa}$

Link with Pedigrees Models



Animal model in Quantitative Genetics

$$X_i = \frac{1}{2}X_a + \frac{1}{2}X_b + \epsilon \quad \epsilon \sim \mathcal{N}(0, 1 - \frac{1}{4}(A_{aa} + A_{bb}))$$

$$\frac{1}{2}A_{ij} = \mathbb{P} \left[\begin{array}{l} \text{two genes from } i \text{ and } j \\ \text{are identical by descent} \end{array} \right] \\ = \text{kinship coefficient}$$

Same as a Network with

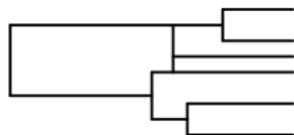
- Several roots ("founders")
- $\gamma_e = \frac{1}{2}$ for all edges
- $\ell_{a \rightarrow i} = 2 - A_{aa}$

Differences

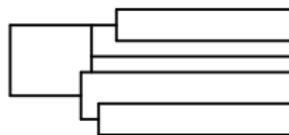
- Variance bounded ($\text{Var}[X_i] = A_{ii} \leq 2$)
- Time scale
- Data at nodes

back

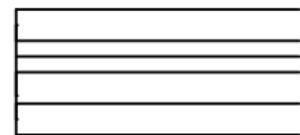
Pagel's λ



$\lambda = 1$ (*original net*)



$\lambda = 0.5$



$\lambda = 0$ (*star tree*)

$$\ell_i(\lambda) = \begin{cases} \lambda \ell_i & i \text{ internal node} \\ \ell_i + (1 - \lambda)t_{\text{pa}(i)} = \lambda \ell_i + (1 - \lambda)t_i & i \text{ tip} \end{cases}$$

back

TE: Single Effect

Model: $\mathbf{Y} = \mu \mathbf{1} + b \bar{\mathbf{N}} + \sigma \mathbf{E}$, $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$

Test: $\mathcal{H}_0 : b = 0$

Stat.: $F_{10} = \frac{\|\mathbf{Y} - \text{Proj}_{\mathbf{1}} \mathbf{Y}\|_{\mathbf{C}^{-1}}^2 - \|\mathbf{Y} - \text{Proj}_{[\mathbf{1} \ \bar{\mathbf{N}}]} \mathbf{Y}\|_{\mathbf{C}^{-1}}^2}{\|\mathbf{Y} - \text{Proj}_{[\mathbf{1} \ \bar{\mathbf{N}}]} \mathbf{Y}\|_{\mathbf{C}^{-1}}^2} \frac{n - r_{[\mathbf{1} \ \bar{\mathbf{N}}]}}{r_{[\mathbf{1} \ \bar{\mathbf{N}}]} - r_{\mathbf{1}}} \sim \mathcal{F}(1, n - 2, \frac{b^2}{2\sigma^2} \|\mathbf{I} - \text{Proj}_{\mathbf{1}}\|_{\mathbf{C}^{-1}}^2)$

TE: Several Effects

Model: $\mathbf{Y} = \mu \mathbf{1} + b \bar{\mathbf{N}} + \mathbf{Nd} + \sigma \mathbf{E}$, $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$

Test: $\mathcal{H}_1 : d_1 = \dots = d_h = 0$

Stat.: $F_{21} = \frac{\left\| \mathbf{Y} - \text{Proj}_{[1 \ N]} \mathbf{Y} \right\|_{\mathbf{C}^{-1}}^2 - \left\| \mathbf{Y} - \text{Proj}_{[1 \ N]} \mathbf{Y} \right\|_{\mathbf{C}^{-1}}^2}{\left\| \mathbf{Y} - \text{Proj}_{[1 \ N]} \mathbf{Y} \right\|_{\mathbf{C}^{-1}}^2} \frac{n - r_{[1 \ N]}}{r_{[1 \ N]} - r_{[1 \ \bar{N}]}}$

$$\sim \mathcal{F} \left(h - 1, n - h - 1, \frac{1}{2\sigma^2} \left\| (\mathbf{I} - \text{Proj}_{[1 \ \bar{N}]}) \mathbf{Nd} \right\|_{\mathbf{C}^{-1}}^2 \right)$$

back