Tests of Phylogenetic Signal on Networks

Cécile Ané[1,2], Paul Bastide[3,4], Mahendra Mariadassou[4], Stéphane Robin[3], Claudia Solís-Lemus[1]

[1] Department of Statistics, University of Wisconsin–Madison, WI, 53706, USA
[2] Department of Botany, University of Wisconsin–Madison, WI, 53706, USA
[3] INRA - AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France
[4] INRA, UR1404 Unité MaIAGE, F78352 Jouy-en-Josas, France.

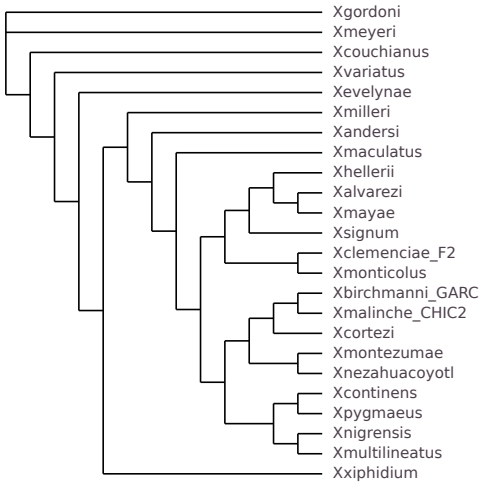
30 May 2017

AgroParisTech

INRA
SCIENCE & IMPACT

MaIAGE

Introduction: Phylogenetic "Networks"



Phylogenetic Tree

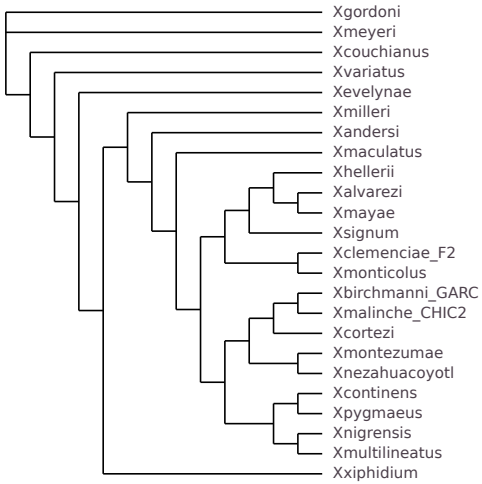
X. Couchianus

X. Nezajuacoyotl

Network: Solís-Lemus and Ané (2016)

Images: Xiphophorus Genetic Stock Center, Texas State University

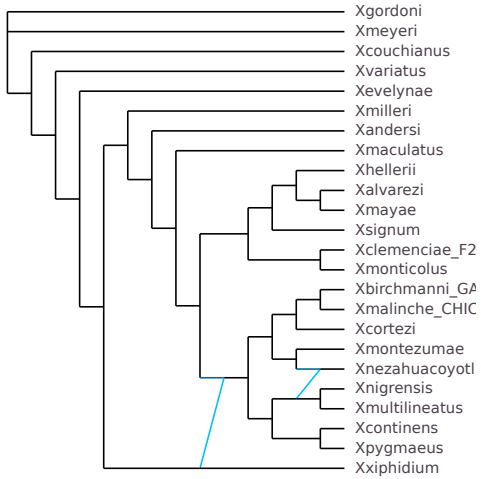
Introduction: Phylogenetic "Networks"



Phylogenetic Tree

X. Couchianus

X. Nezajuacoyotl

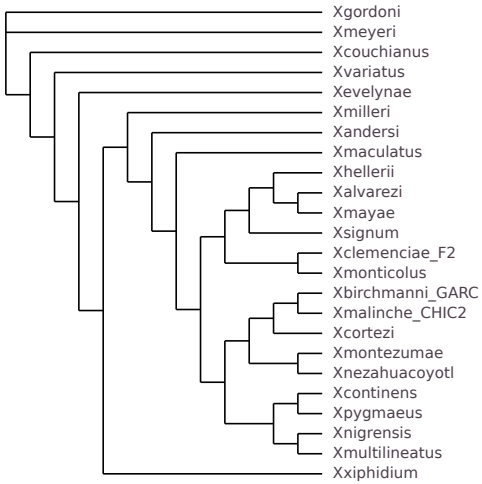


Phylogenetic "Network"

Network: Solís-Lemus and Ané (2016)

Images: Xiphophorus Genetic Stock Center, Texas State University

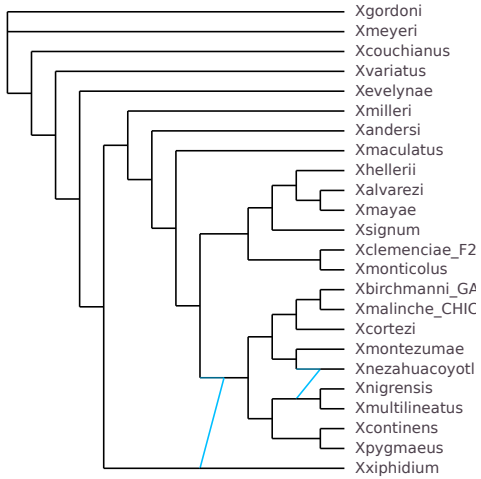
Introduction: Phylogenetic "Networks"



Phylogenetic Tree

X. Couchianus

X. Nezajuacoyotl

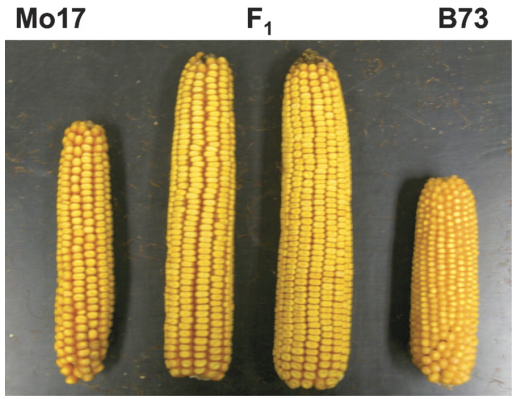


Phylogenetic "Network"

Question: Can we see the effects of ancestral hybridization on the trait measured for present-day species ?

Network: Solís-Lemus and Ané (2016) Images: Xiphophorus Genetic Stock Center, Texas State University

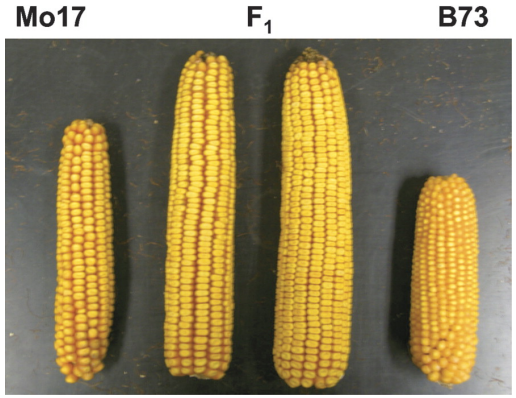
Introduction: Heterosis



Heterosis: hybrid vigor

Image: Springer and Stupar (2007)

Introduction: Heterosis



Heterosis: hybrid vigor

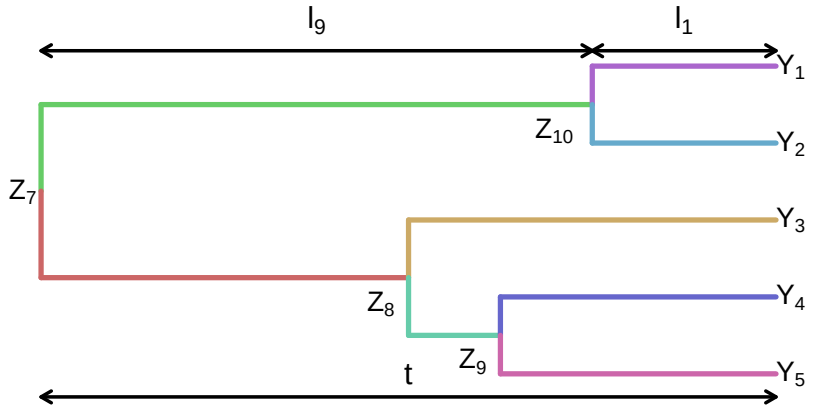
Question: Can we see the effects of ancestral heterosis on the trait measured for present-day species ?

Image: Springer and Stupar (2007)

Outline

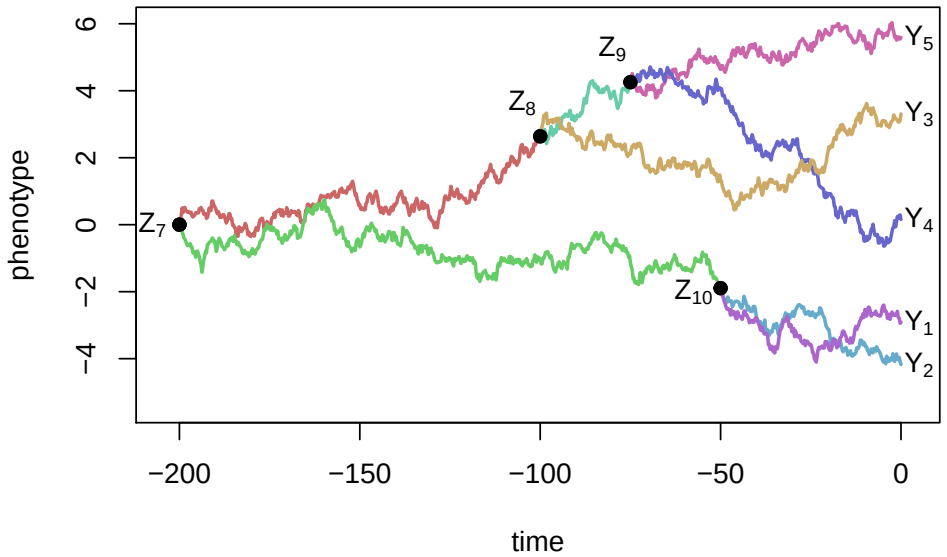
Stochastic Process on a Tree

(Felsenstein, 1985)



The tree is known.

Only *tip* values are observed



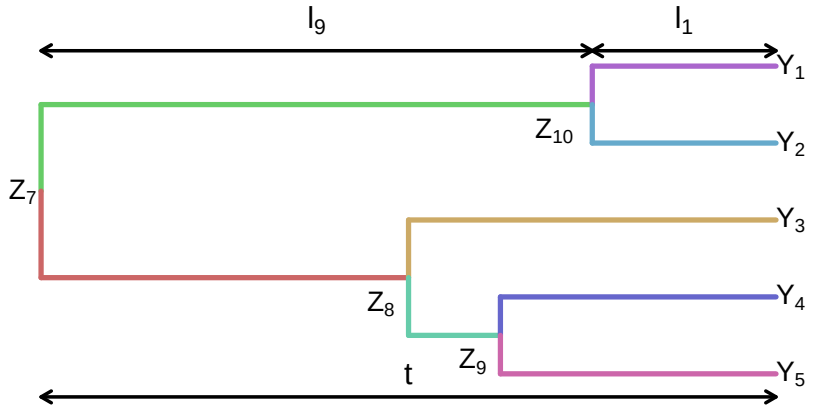
Brownian Motion:

$$\mathbb{V}\mathrm{ar}\left[Y_1\right] = \sigma^2 t = \sigma^2(\ell_9 + \ell_1)$$

$$\mathbb{C}\mathrm{ov}\left[Y_1; Y_2\right] = \sigma^2 \ell_9$$

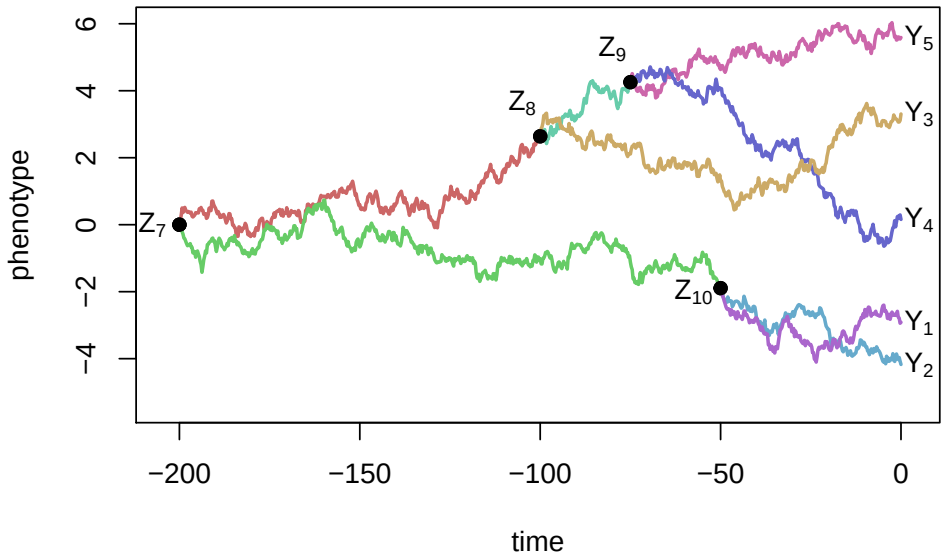
Stochastic Process on a Tree

(Felsenstein, 1985)



The tree is known.

Only *tip* values are observed



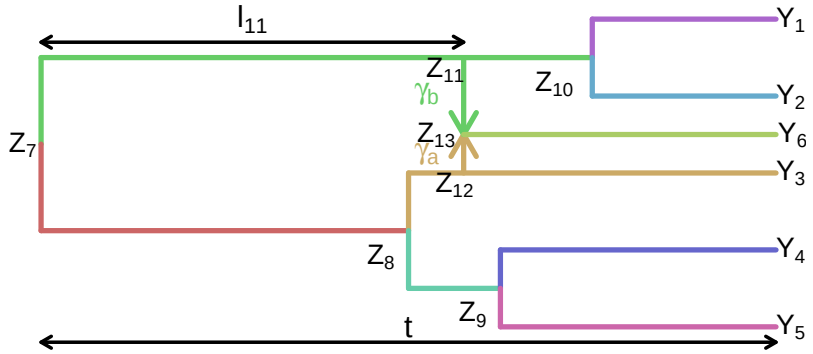
Brownian Motion:

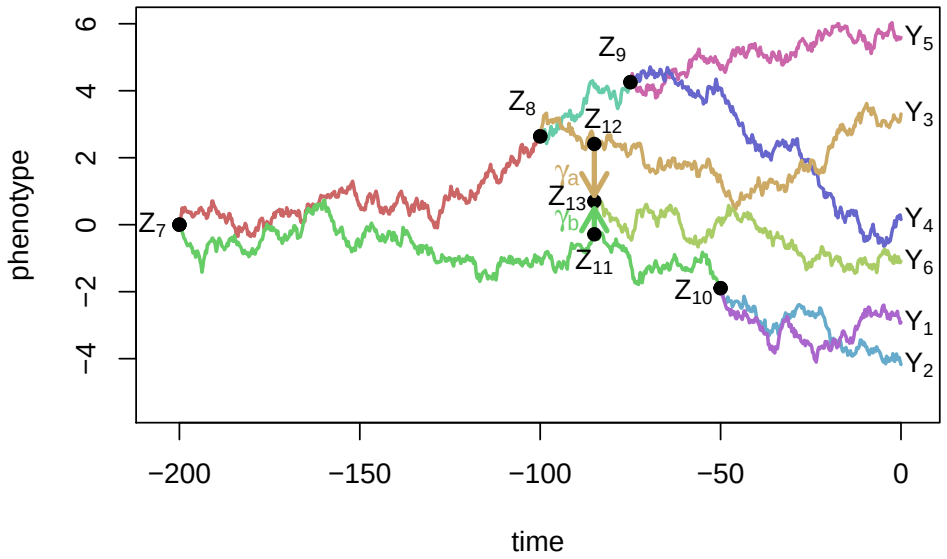
$$V_{ij}^{\text{tree}} = \sum_{e \in p_i \cap p_j} \ell_e$$

Sum over shared edges.

p_i: path from root to tip i

Stochastic Process on a Network





The *network* is known.

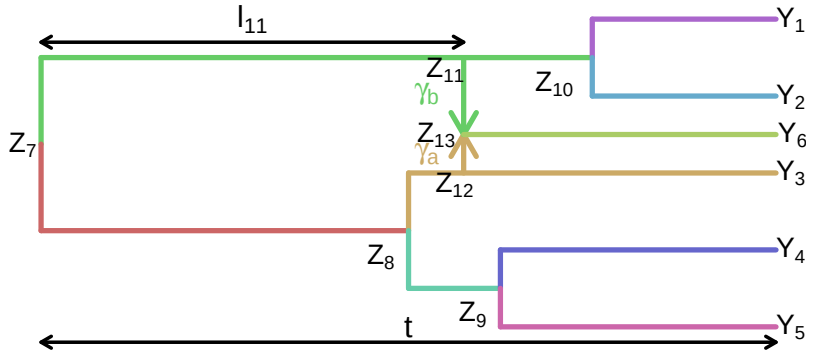
Only *tip* values are observed

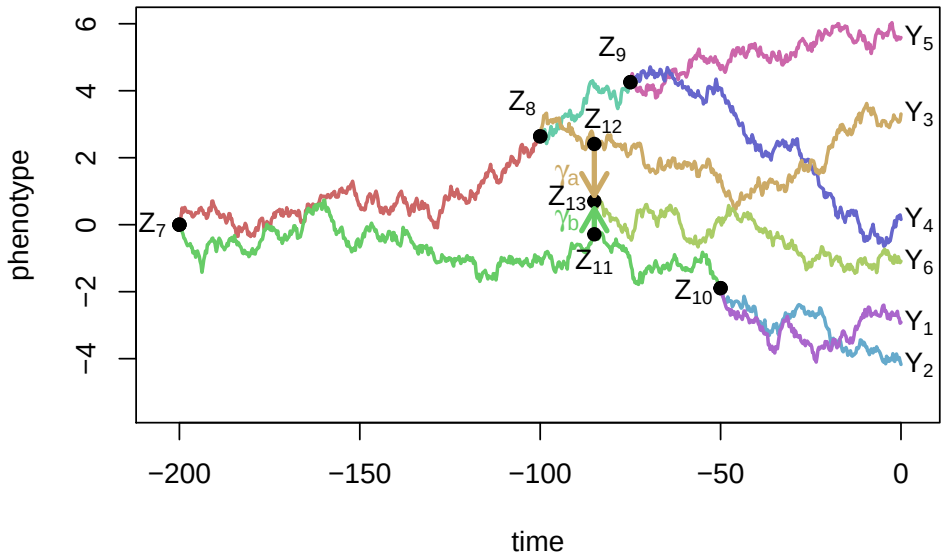
Hybrid:

$$Z_{13} = \gamma_a Z_{12} + \gamma_b Z_{11}$$

$$\gamma_a + \gamma_b = 1$$

Stochastic Process on a Network





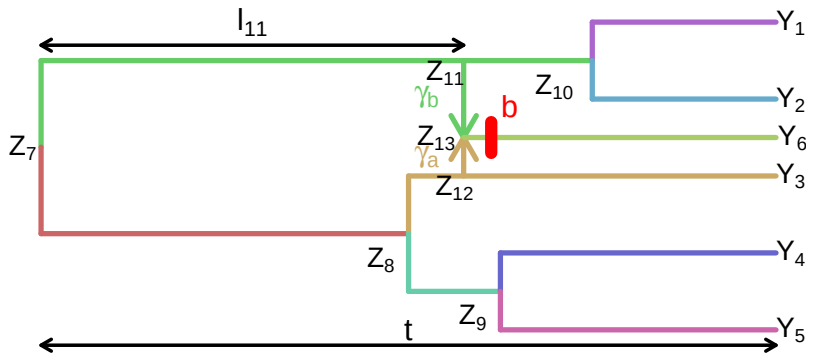
The *network* is known.

Only *tip* values are observed

Hybrid:

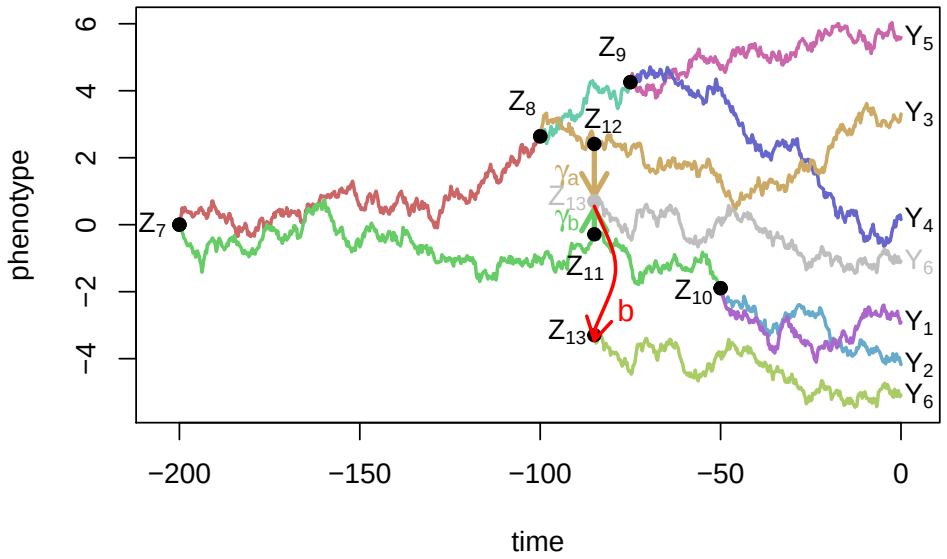
$$V_{ij}^{\text{net}} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \left(\prod_{e \in p_i} \gamma_e \right) \left(\prod_{e \in p_j} \gamma_e \right) \sum_{e \in p_i \cap p_j} \ell_e$$

Heterosis



The *network* is known.

Only *tip* values are observed

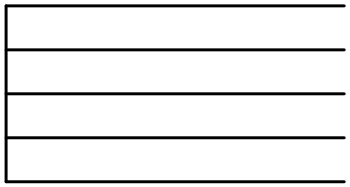
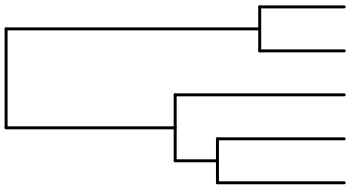
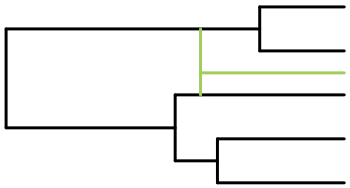
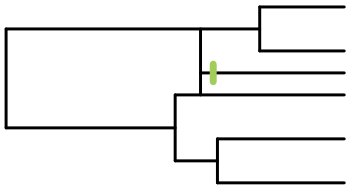


Hybrid:

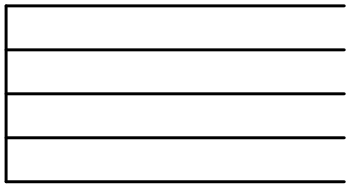
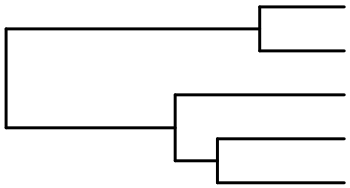
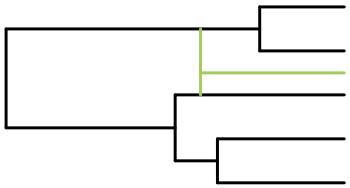
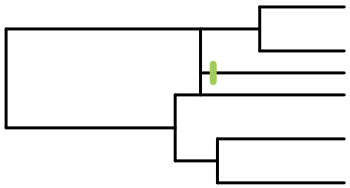
$$Z_{13} = \gamma_a Z_{12} + \gamma_b Z_{11} + b$$

b : Heterosis.

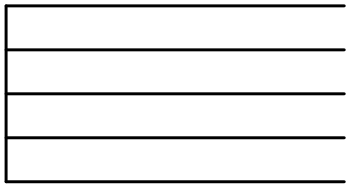
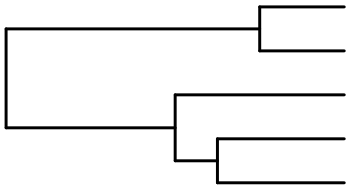
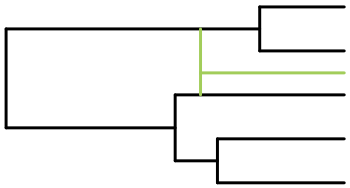
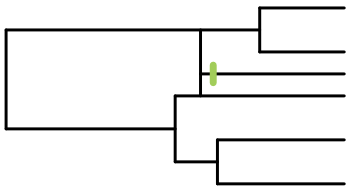
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

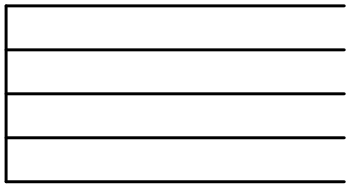
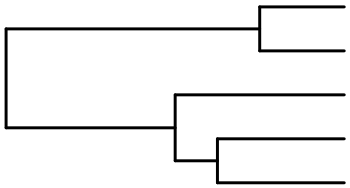
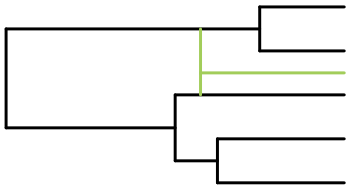
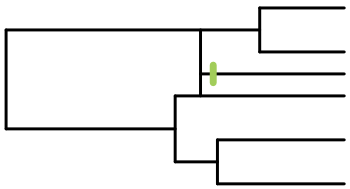
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	$\updownarrow$?
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

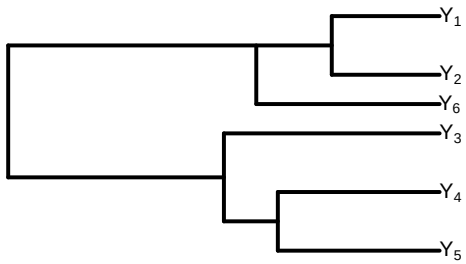
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-) (Pagel, 1999)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

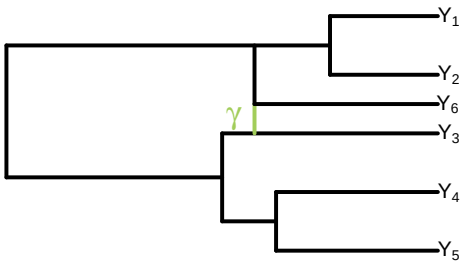
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-) (Pagel, 1999)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	↕ ?
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

Tree vs Network



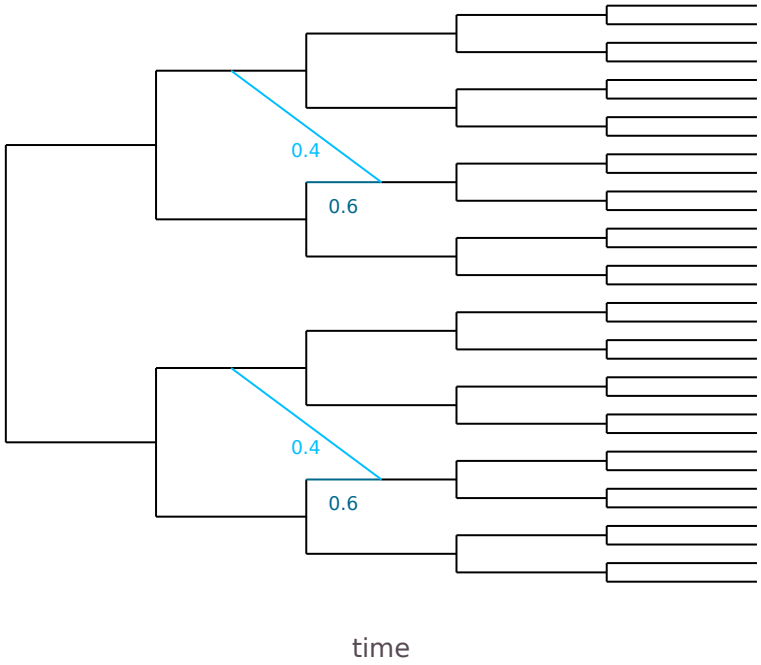
Tree: $\gamma = 0$

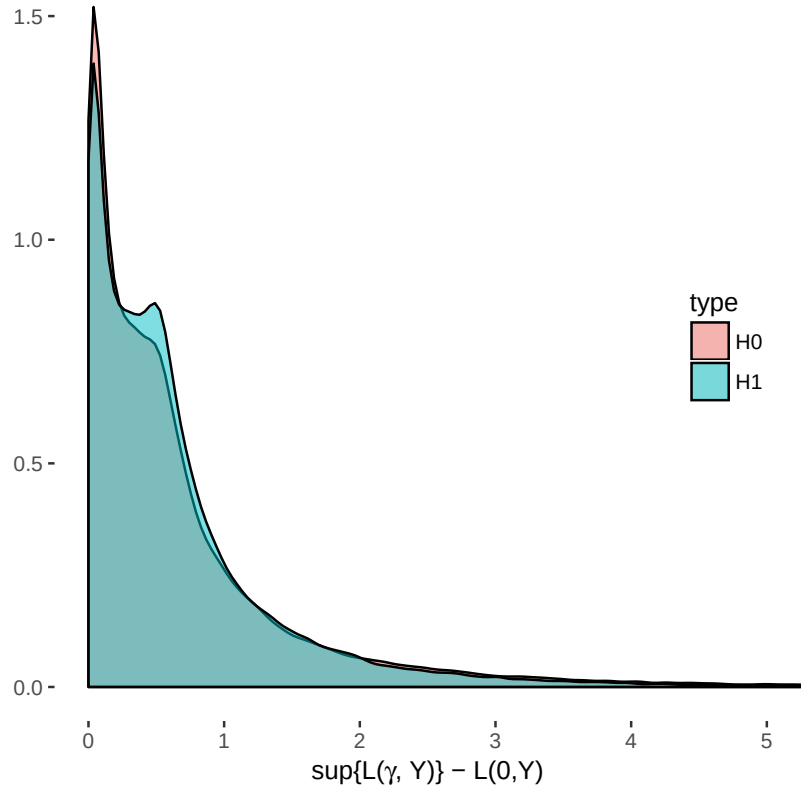


Network: $\gamma \neq 0$

$$\text{LRT:} \quad T = 2\left(\sup_{\gamma \neq 0}\{L(\gamma, Y)\} - L(0, Y)\right)$$

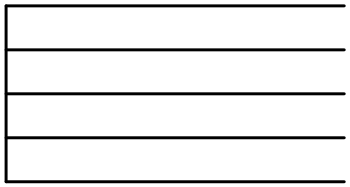
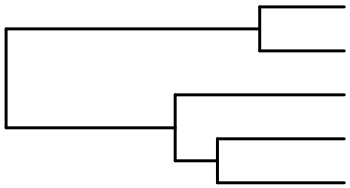
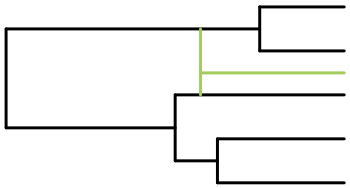
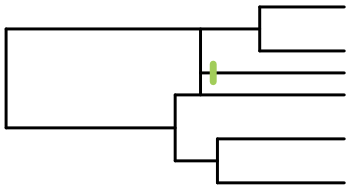
Tree vs Network



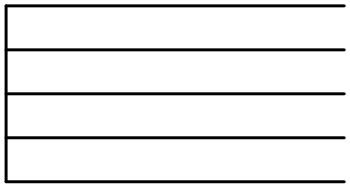
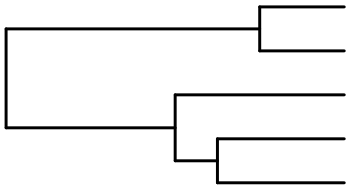
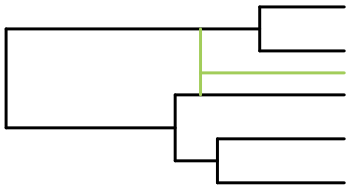
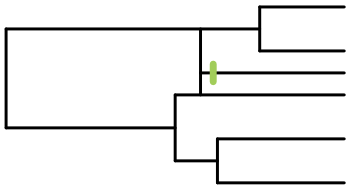


Empirical Distribution, 100000 *simulations.*

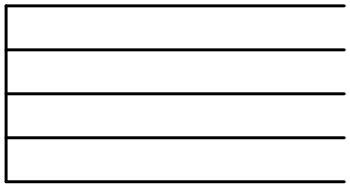
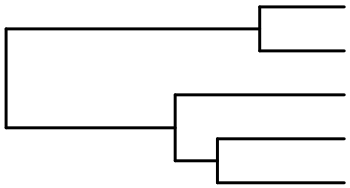
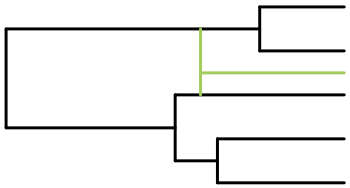
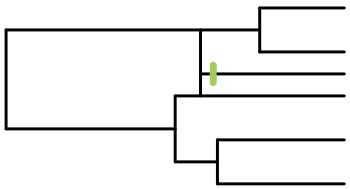
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	↕ ?
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

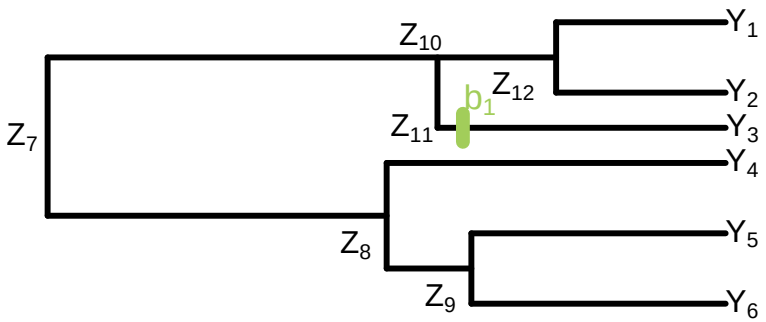
Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	
			↕ :-)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	
			↕ :-(
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	↕ :-(
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	↕ ?
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

Linear Regression Model

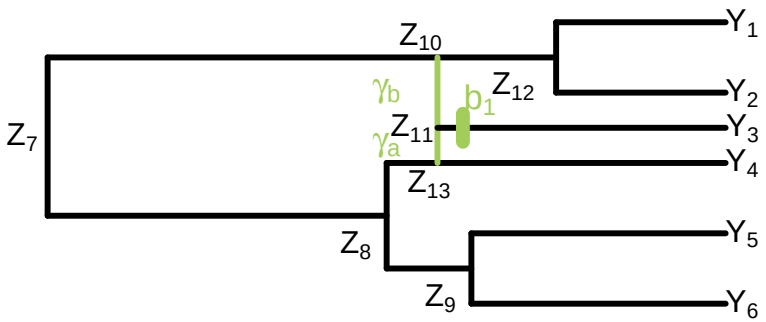


$$\mathbf{b} = \begin{matrix} 7 \\ 8 \\ 9 \\ 10 \\ 12 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{Tb} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \begin{pmatrix} \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

$$\mathbf{T} = \begin{matrix} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{matrix} \begin{matrix} \begin{matrix} Z_7 & Z_8 & Z_9 & Z_{10} & Z_{12} & Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{matrix} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{matrix}$$

$$\textit{Model}: \quad Y = \mu\mathbf{1} + \mathbf{Tb} + E^{\text{tree}}$$

Linear Regression Model



$$\mathbf{b} = \begin{matrix} 7 \\ 8 \\ 9 \\ 10 \\ 12 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{Tb} = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \begin{pmatrix} \cdot \\ \cdot \\ b_1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

$$\mathbf{T} = \begin{matrix} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{matrix} \begin{matrix} \begin{matrix} Z_7 & Z_8 & Z_9 & Z_{10} & Z_{12} & Z_{13} \end{matrix} & \begin{matrix} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{matrix} \\ \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & \gamma_b & \cdot & \gamma_a \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 1 & 1 & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & \cdot & \cdot & \cdot \end{pmatrix} & \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{matrix}$$

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

$$\textit{Model}: \quad Y = \mu \mathbf{1} + \mathbf{Tb} + E^{\text{net}}$$

Heterosis: Testing Effect

Model:

$$\mathbf{Y} = \mu\mathbf{1} + \mathbf{Tb} + \sigma^2\mathbf{E} \quad , \quad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

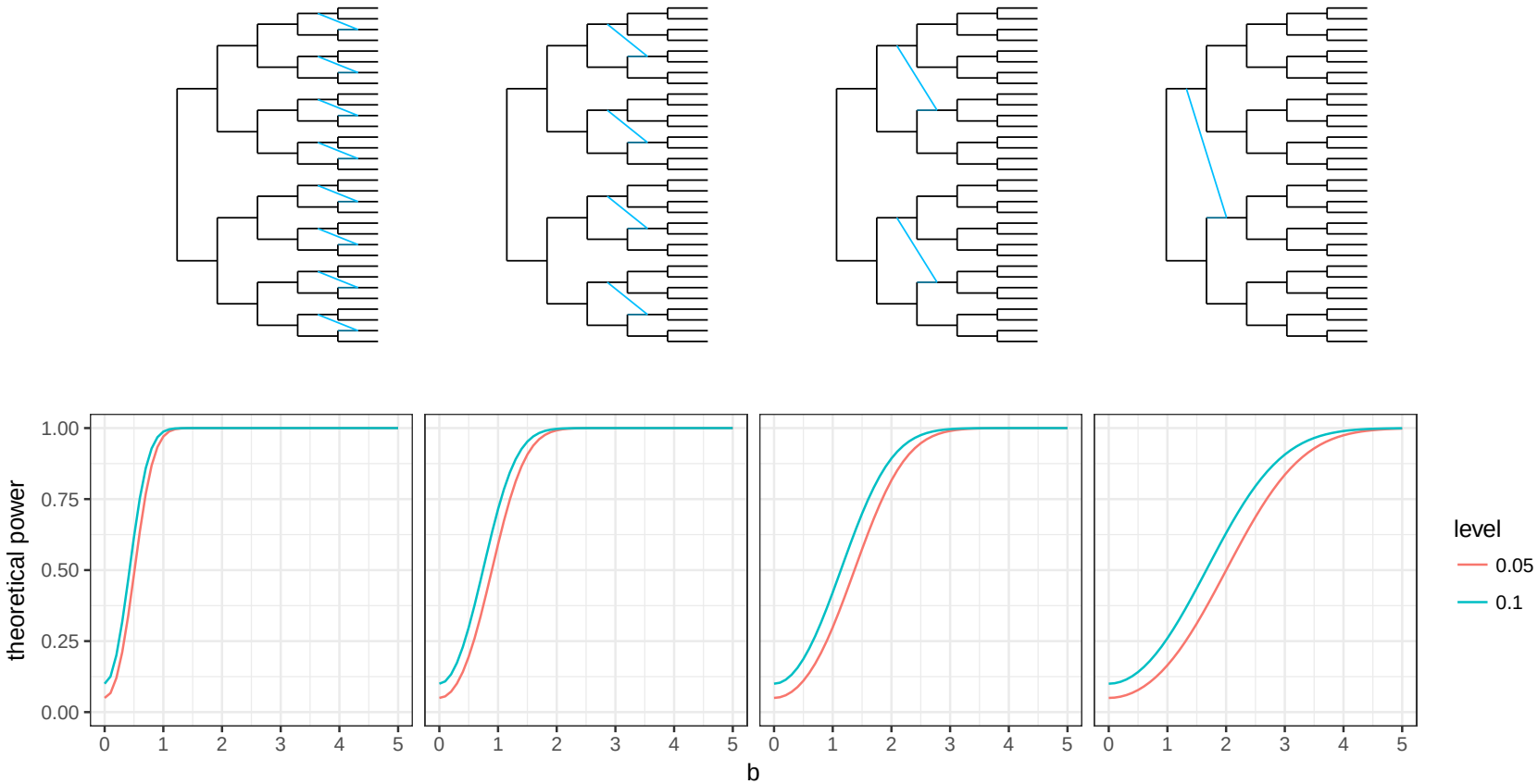
Tests:

$\mathcal{H}_0$: No heterosis
$\mathcal{H}_1$: Heterosis with one single effect
$\mathcal{H}_2$: Heterosis with heterogeneous effects

Stats.:

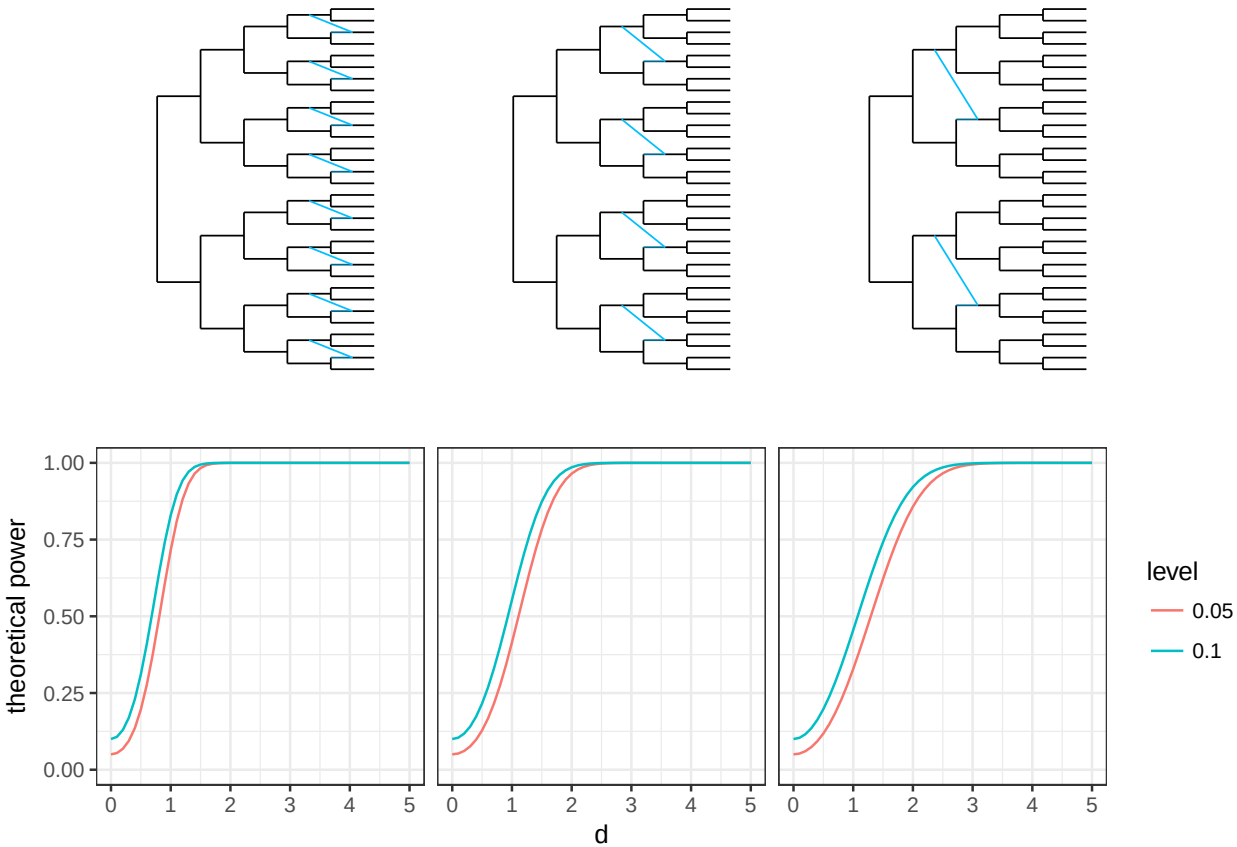
$$F_{10} \sim \mathcal{F}\left(1, n-2, \Delta_{10}(b, \sigma^2)\right)$$
$$F_{21} \sim \mathcal{F}\left(h-1, n-h-1, \Delta_{21}(\mathbf{b}, \sigma^2)\right)$$

Heterosis: Single Effect



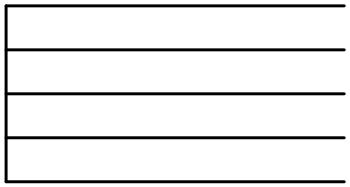
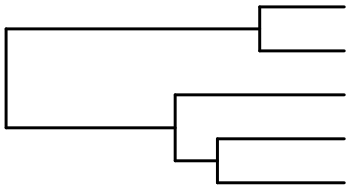
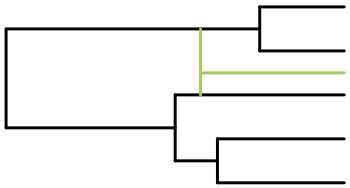
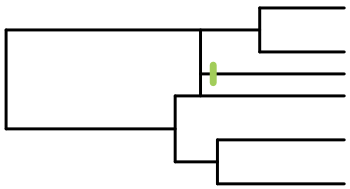
Detection Power ($\sigma^2 = 1$)

Heterosis: Several Effects

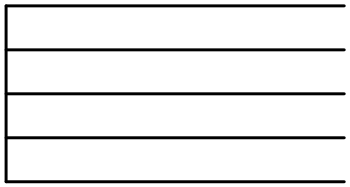
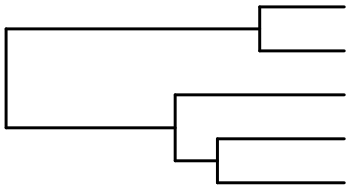
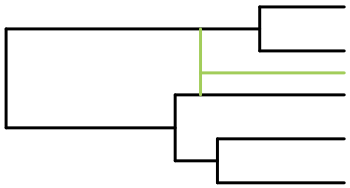
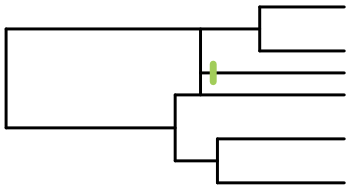


Detection Power ($\sigma^2 = 1$)

Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	↕ :-(
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	↕ ?
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

Linear Models

	Structure	Model	Test
Independent traits		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{I})$	↕ :-)
Phylogenetic Tree		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{tree}})$	↕ :-(
Phylogenetic Network		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	↕ :-)
Heterosis		$\mathbf{Y} \sim \mathcal{N}(\mu\mathbf{1}+\mathbf{Tb}, \sigma^2\mathbf{V}^{\text{net}}(\gamma))$	

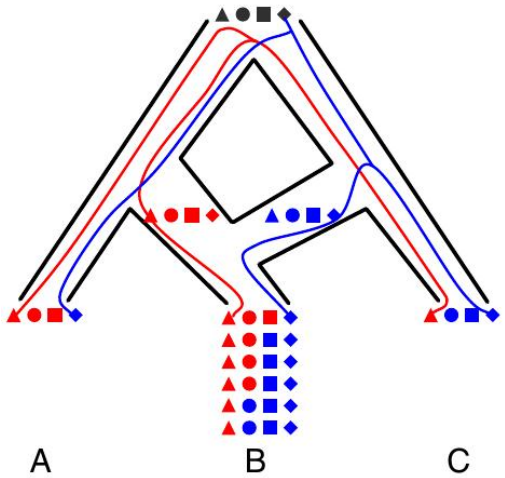
Phylogenetic Networks

- Inference is hard Kubatko (2009); Yu et al. (2012, 2014); Solís-Lemus and Ané (2016).
- Main issue: deal with Incomplete Lineage Sorting (ILS).
- State of the art methods: up to ~ 15 taxa, with ~ 3 hybrids.

$\rightarrow$ Need fast algorithms, and fast implementation

Phylogenetic Networks

- Inference is hard Kubatko (2009); Yu et al. (2012, 2014); Solís-Lemus and Ané (2016).
- Main issue: deal with Incomplete Lineage Sorting (ILS).



ILS (Yu et al., 2014)

- State of the art methods: up to ~ 15 taxa, with ~ 3 hybrids.

$\rightarrow$ Need fast algorithms, and fast implementation

Phylogenetic Networks

- Inference is hard Kubatko (2009); Yu et al. (2012, 2014); Solís-Lemus and Ané (2016).
- Main issue: deal with Incomplete Lineage Sorting (ILS).

- State of the art methods: up to ~ 15 taxa, with ~ 3 hybrids.

$\rightarrow$ Need fast algorithms, and fast implementation

julia package PhyloNetworks

julia

- "Julia is a **high-level**, **high-performance** dynamic programming language for numerical computing."
- "Julia use expanded dramatically in 2016, and 2017 is shaping up to be the year that Julia expands **from early adopters into the mainstream**."
- (source: julia website...)

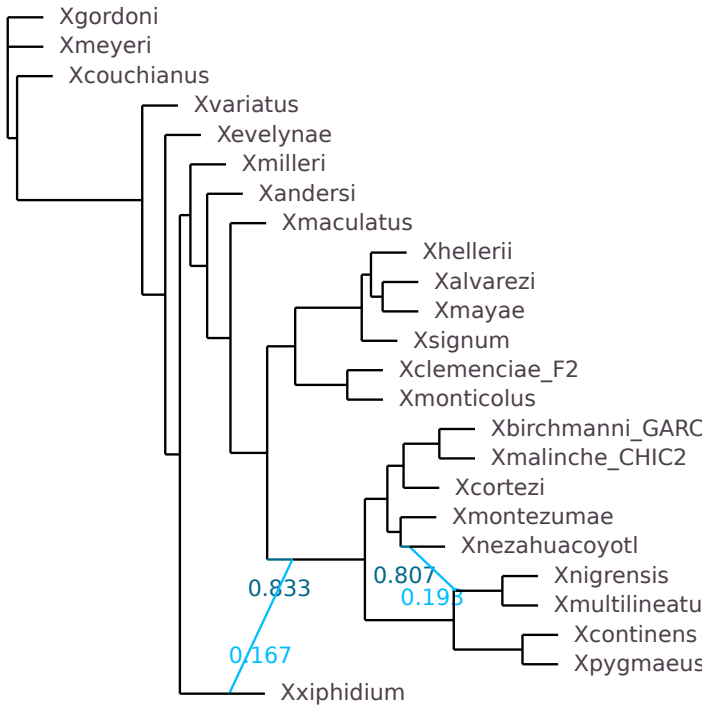
PhyloNetworks

- Inference of phylogenetic networks
- Interactive study of networks

Xiphophorus Fish Dataset

```
using PhyloNetworks

net = readTopology(fish)
plot(net, useEdgeLength=true, showGamma=true)
```



No branch length

Conclusion and Perspectives

A general inference framework for trait evolution on networks.

Conclusions

- BM model of evolution on networks
- Tests of hybridization / heterosis
- Phylogenetic regression

`julia` package

- Available on GitHub
- From network inference to trait evolution

Perspectives

- Getting a network
- Ornstein-Uhlenbeck
- Shifts detection

Bibliography

Felsenstein, J.
1985. Phylogenies and the Comparative Method. *The American Naturalist*, 125(1):1–15.

Hergé
1958. *Coke en Stock*.

Kubatko, L. S.
2009. Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology*, 58(5):478–488.

Pagel, M.
1999. Inferring the historical patterns of biological evolution. *Nature*, 401(6756):877–884.

Solís-Lemus, C. and C. Ané
2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, 12(3):e1005896.

Springer, N. M. and R. M. Stupar
2007. Allelic variation and heterosis in maize: How do two halves make more than a whole? *Genome Research*, 17(3):264–275.

Yu, Y., J. H. Degnan, and L. Nakhleh
2012. The Probability of a Gene Tree Topology within a Phylogenetic Network with Applications to Hybridization Detection. *PLoS Genetics*, 8(4):e1002660.

Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh
2014. Maximum likelihood inference of reticulate evolutionary histories. *PNAS*, 111(46):16448–16453.

Photo Credits:

- Xiphophorus Genetic Stock Center, Texas State University,

`http://www.xiphophorus.txstate.edu/resources/galleries/comprehensive.html`

Thank you for listening

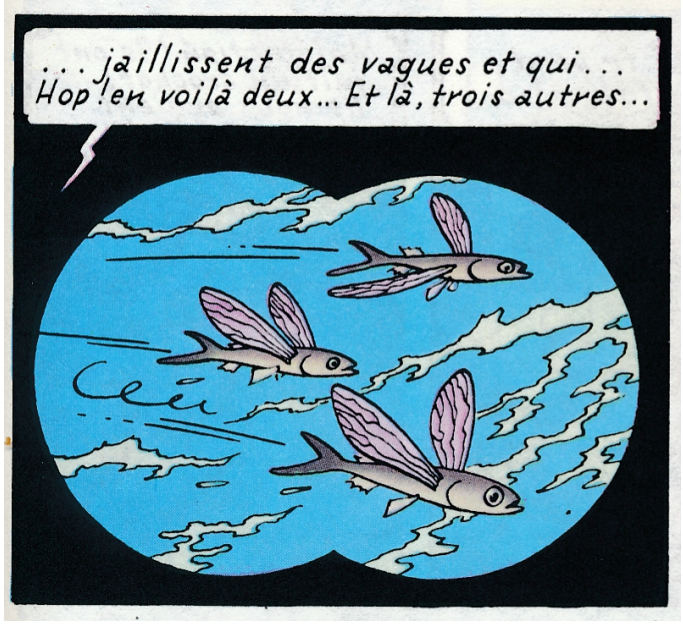


Image: Hergé (1958)

pbastide.github.io

Appendices