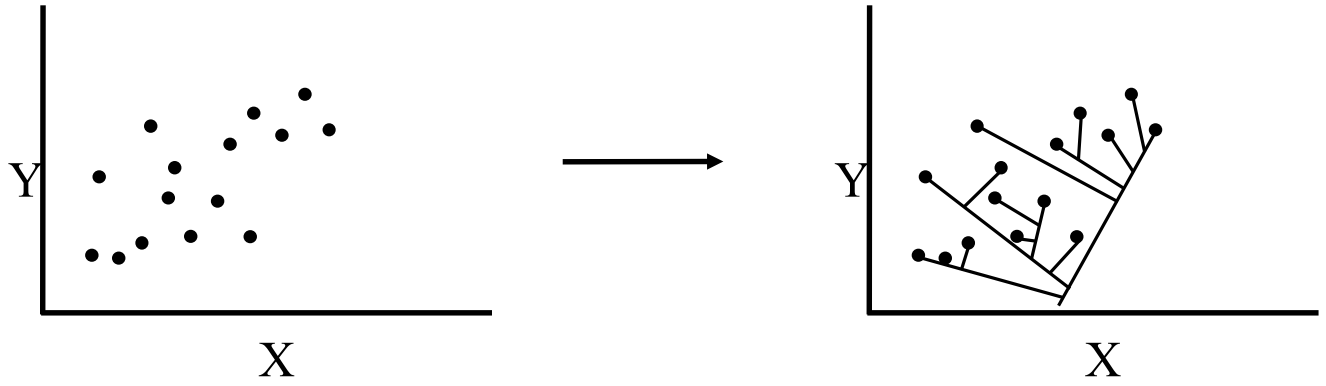
Phylogenetic Comparative Methods: II

Last time:

-Taxa are not independent

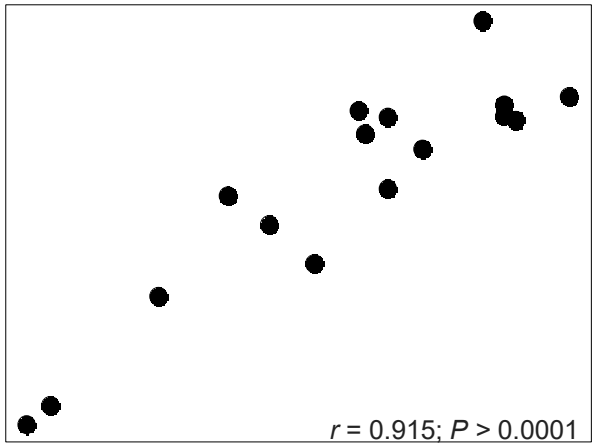
-Ignore evolutionary history AT ONE'S PERIL!



Phylogenetic comparative methods *condition* the data on the phylogeny to account for lack of independence during the analysis

Continuous Data: The Problem

Issues arises with both discrete and continuous traits

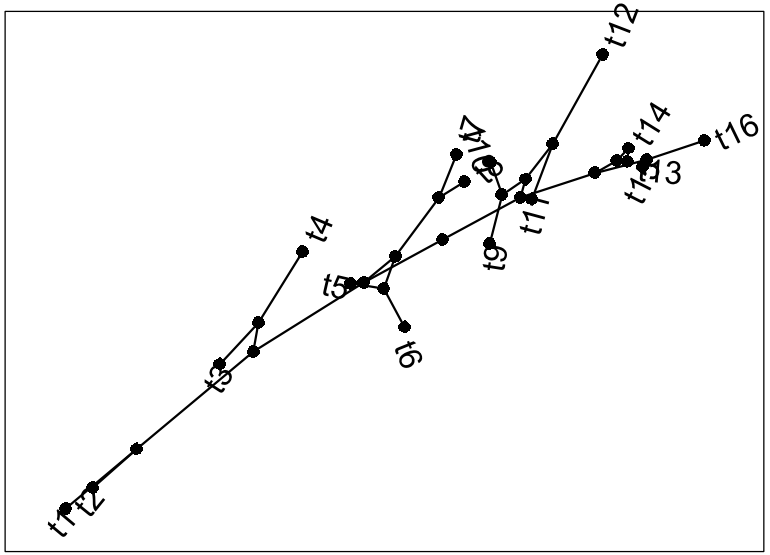


Say we have this correlation,

Here is the same pattern with the phylogeny superimposed

Clearly, closely related taxa are similar.

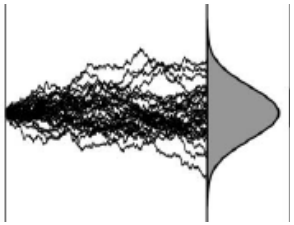
How can we account for this?



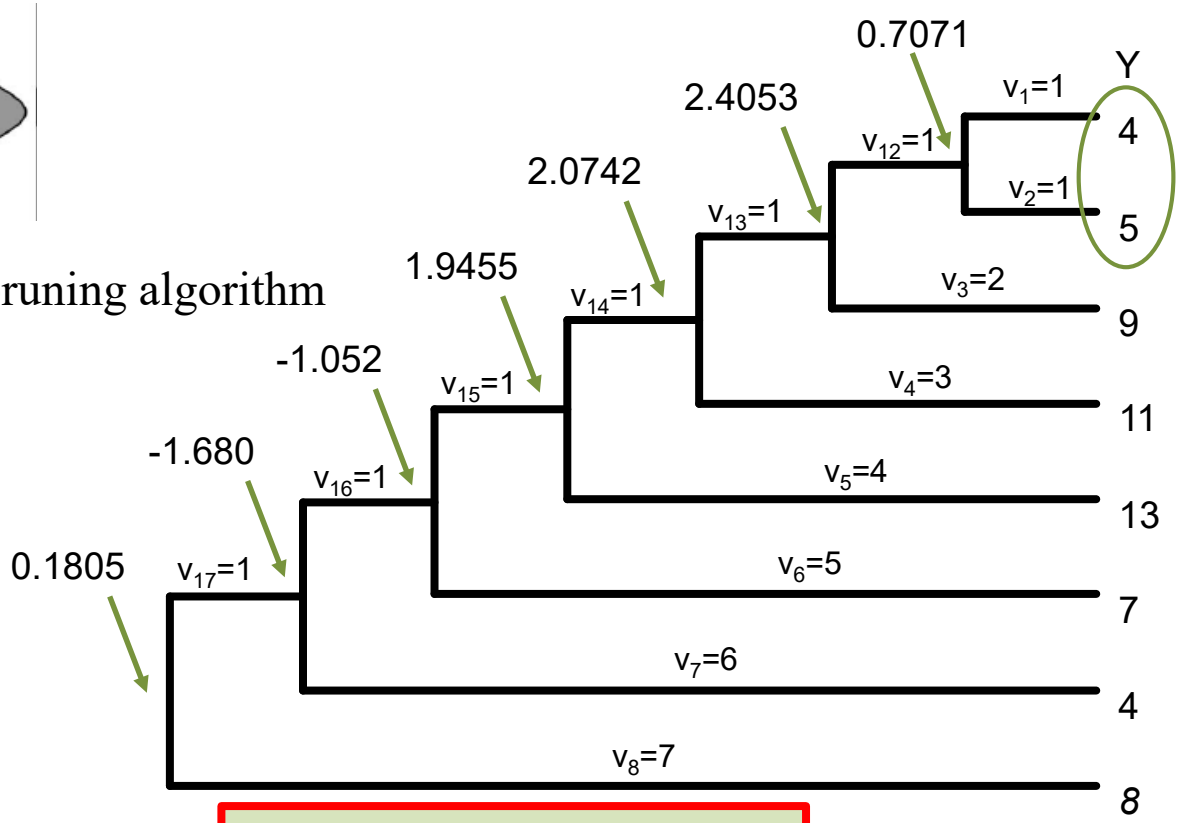
Phylogenetically Independent Contrasts (PIC)

Construct contrast scores at tree nodes, which are *evolutionarily independent of one another*

-Based on a Brownian motion null model of trait evolution



-Find contrasts via recursive pruning algorithm



Contrast scores

$$Y_{ij} = \frac{Y_i - Y_j}{\sqrt{v_i + v_j}}$$

$$c_1 = \frac{5-4}{\sqrt{1+1}}$$

Internal nodes: weighted average

$$Y_{n1} = \frac{{}^{1}/_{v_1} Y_1 + {}^{1}/_{v_2} Y_2}{{}^{1}/_{v_1} + {}^{1}/_{v_2}}$$

NOTE: Internal branches adjusted as:

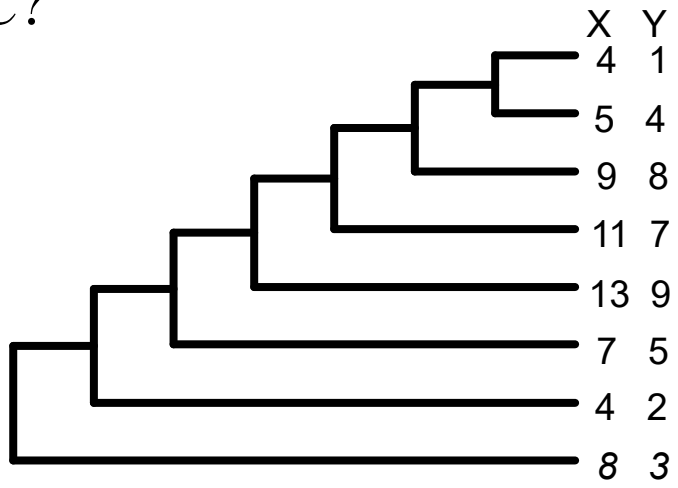
$$v_{ij}^{*} = v_{ij} + \left(\frac{1}{1/v_i + 1/v_j} \right)$$

CHALLENGE: See if you can calculate the same PICs using the Y data and branch lengths!

Felsenstein (1985)

Testing Associations: Independent Contrasts

How do we test trait correlations with PIC?



1: Calculate X_{pic} & Y_{pic}

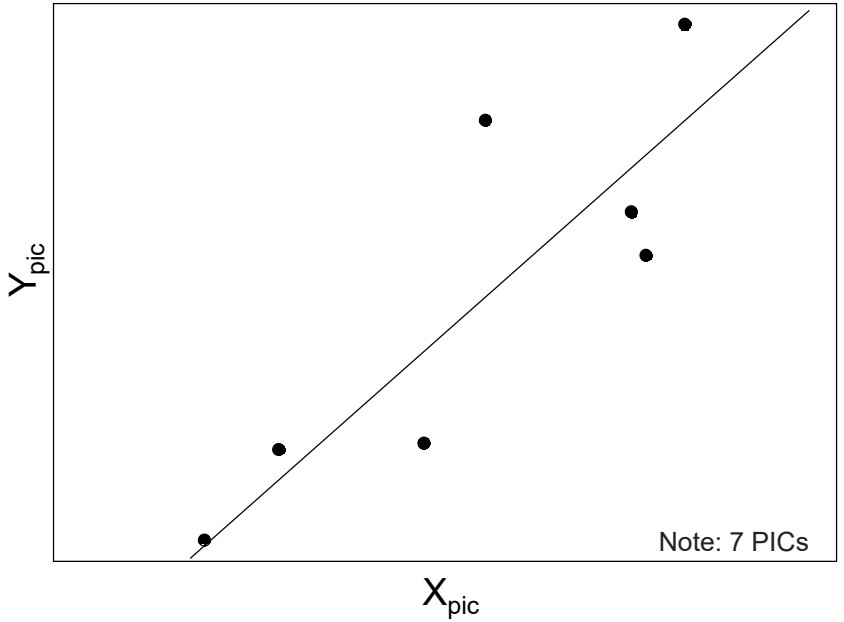
pic.x	pic.y
0.1805090	-0.6229475
-1.6805047	-1.4447758
-1.0520183	-0.6755435
1.9455001	1.3404970
2.0742519	0.9723056
2.4053512	2.9398737
0.7071068	2.1213203

2: Test association via correlation and regression
(note: regression through origin as sign of contrasts arbitrary)

$r_{PIC} = 0.848$

```
> anova(lm(pic.y~pic.x + 0))
>      Df Sum Sq Mean Sq F value Pr(>F)
> pic.x     1 14.3519 14.3519 19.285 0.00461 **
  Residuals 6 4.4651 0.7442 ---
```

$\beta_{PIC} = 0.8846$

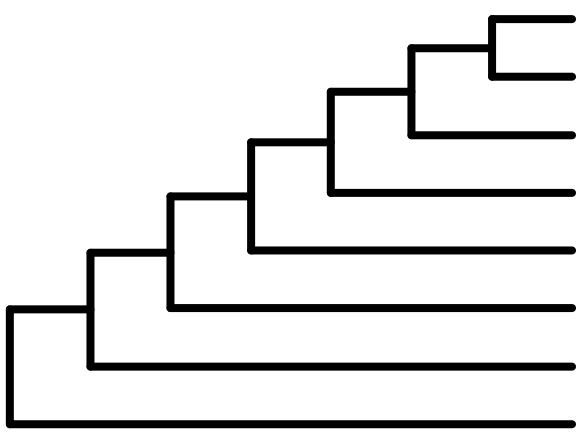


Felsenstein (1985)

PIC Challenges: ANOVA Models

PIC works great for regression/correlation

What about ANOVA (Y~gp)?



Y	Gp	Gp
1	A	1
4	A	1
8	A	1
7	A	1
9	B	0
5	B	0
2	B	0
3	B	0

Could use binary coding to 'trick' PIC algorithm

GP_{PIC}
-0.09767595
-0.14980902
-0.23925500
-0.40333538
0.00000000
0.00000000
0.00000000

Now Y_{PIC}~ GP_{PIC} is anova

But what about 3 or more groups?

Solutions possible, but far from obvious*

*NOTE: Algebraic solution is to use *k*-1 binary columns for *k* groups. PICs are obtained for each, and the set of GP_{PIC} columns treated as the independent variables. This will work, but algebra must be done by hand, as canned functions will not generate proper null model for test (DCA has never seen this done)

Testing Associations: "Phylogenetic" ANOVA

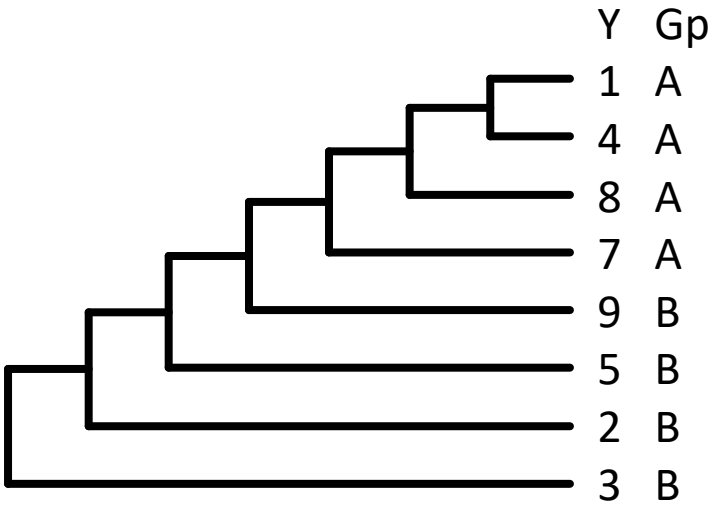
For Phy-ANOVA, a simulation approach can be used

1: Perform NON-PHYLOGENETIC ANOVA

2: Simulate many Y datasets on phylogeny under BM

3: Evaluate simulated datasets

4: Compare observed ANOVA output to distribution



Garland et al. (1993)

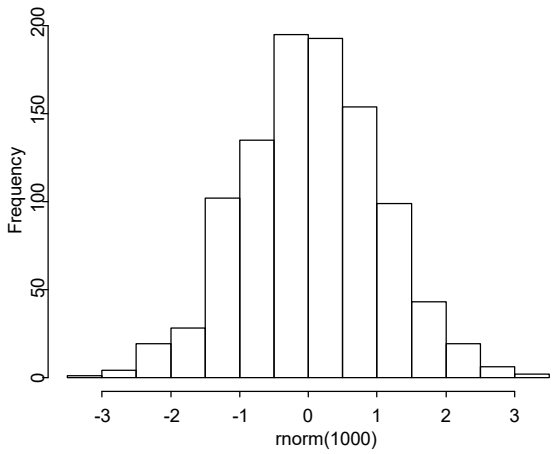
Digression: BM Simulation

How does one simulate data?

Point-estimates:
random draws from a distribution

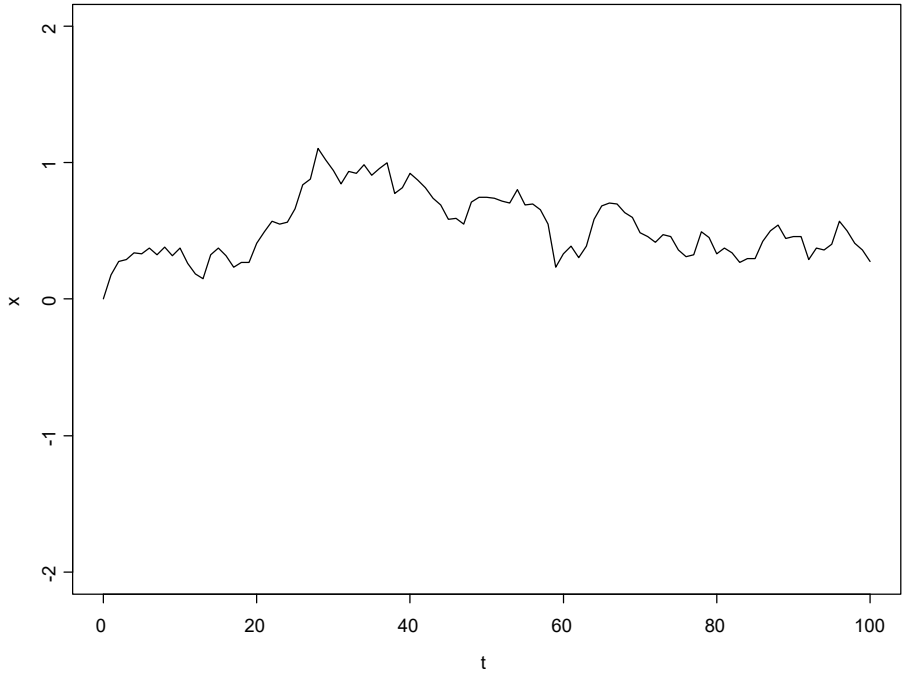
```
➢ rnorm(1)
-0.9698378

> hist(rnorm(1000))
```



BM over time:
Cumulative sum of random draws
for many time steps

```
t <- 0:100 # time
sig2 <- 0.01 ## first, simulate a set of random deviates
x <- rnorm(n = length(t) - 1, sd = sqrt(sig2))
 ## now compute their cumulative sum
x <- c(0, cumsum(x)) plot(t, x, type = "l", ylim = c(-2, 2))
```



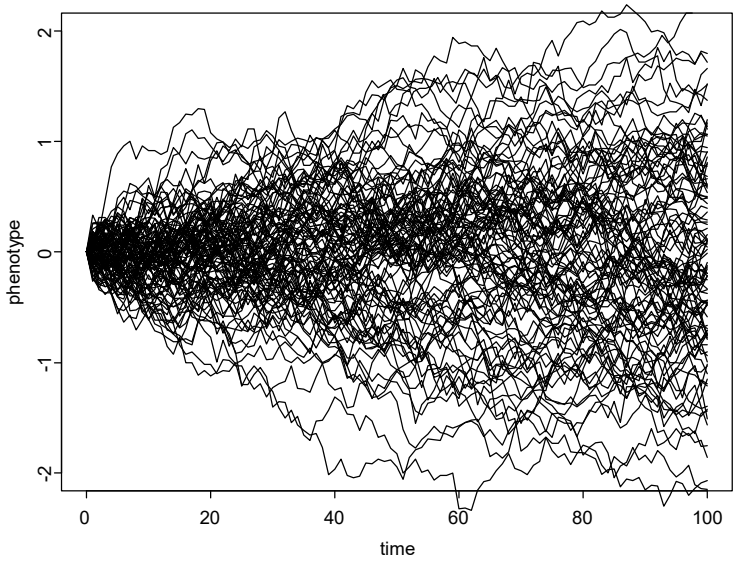
example from L. Revell
http://www.phytools.org/eqg/Exercise_4.1/

Digression: BM Simulation

How does one simulate data?

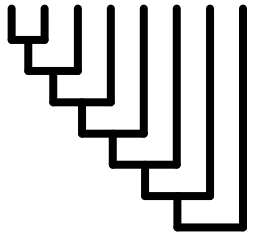
Many BM runs over time:

```
nsim <- 100
X <- matrix(rnorm(n = nsim * (length(t) - 1), sd = sqrt(sig2)), nsim,
length(t) - 1)
X <- cbind(rep(0, nsim), t(apply(X, 1, cumsum)))
plot(t, X[1, ], xlab = "time", ylab = "phenotype", ylim = c(-2, 2), type =
"l") apply(X[2:nsim, ], 1, function(x, t) lines(t, x), t = t)
```



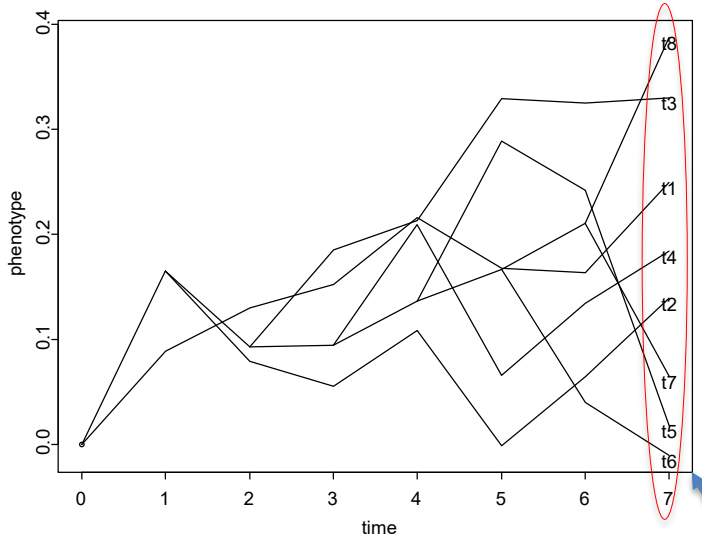
BM on phylogeny:

Same idea, but keep track of branches



<<Code skipped here>>

Many R functions do this: sim.char, fast.BM, etc.



Final values at tips

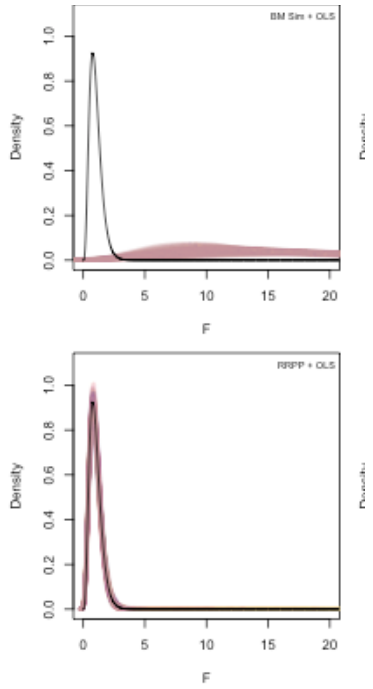
DCA Programming note: one can also simulate MVN data and post-multiply (project) by the phylogenetic covariance matrix

examples from L. Revell
http://www.phytools.org/eqg/Exercise_4.1/

“Phylogenetic” ANOVA”: Problem

PROBLEM: Typical approach does NOT condition data on phylogeny:
It is OLS-based ANOVA with BM simulations (wrong model, evolutionary relationships ignored!)

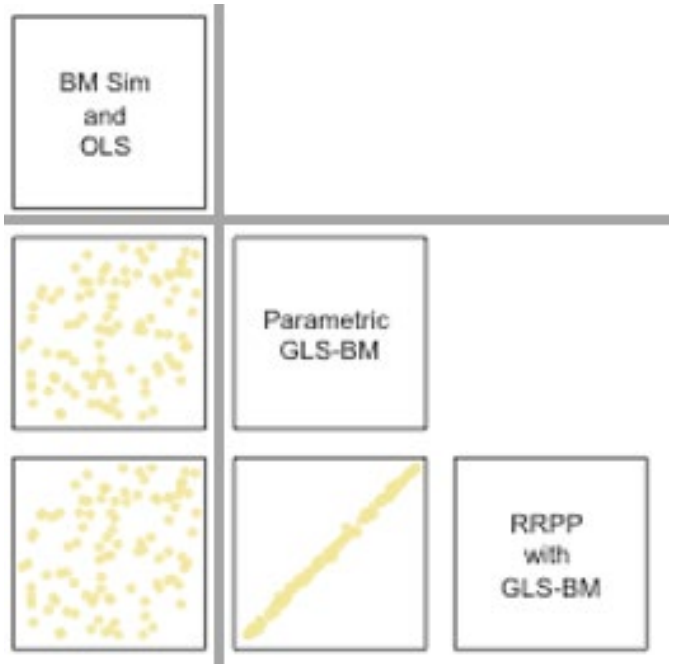
F-distributions



Simulation-based with OLS

Permutation-based (RRPP)

P-values from simulation



Need a better approach (PGLS: next)

Adams and Collyer (2022)

Testing Associations: PGLS

Phylogenetic Generalized Least Squares

Follows the GLS model

(a 'correlated observations method)

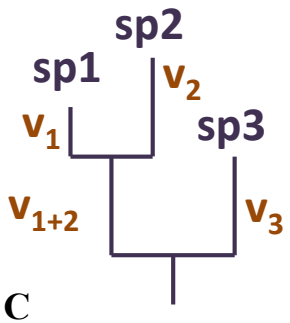
$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \varepsilon$$

Data — The Design — Error: $\mathcal{N}(0, \mathbf{V})$

-error (ε) is *not iid*, but contains expected covariation as described by phylogeny (**V**)

V = *phylogenetic covariance matrix**

-Describes amount of evolutionary time species *share* via common ancestors (and thus how similar their trait values are expected to be)



$$\mathbf{V} = \begin{array}{c} \begin{array}{ccc} \text{sp1} & \text{sp2} & \text{sp3} \end{array} \\ \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix} \end{array} \begin{array}{l} \text{sp1} \\ \text{sp2} \\ \text{sp3} \end{array}$$

*Matrix is also called **C**

Grafen (1989)

For review see Adams and Collyer (2018a)

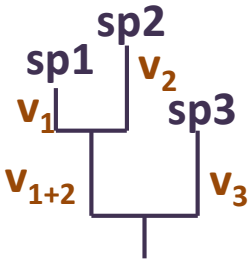
Solving PGLS: must find $\hat{\boldsymbol{\beta}}$

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \varepsilon$$

Parameters obtained using standard GLS approach:

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathbf{t}}\mathbf{V}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{t}}\mathbf{V}^{-\mathbf{1}}\mathbf{Y}$$

$$V = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$



-PGLS is a 'weighted' GLS, where weights are inverse of structured error

But how did we get this equation?

Grafen (1989)

Statistical Digression: Linear Models

Ordinary Least Squares (OLS) models (e.g., ANOVA, regression) are of the form:

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \varepsilon$$

Matrix of dependent values

Model matrix of independent variables

Model coefficients

Matrix of unexplained values (the error)

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & \vdots \\ 1 & X_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Parameters found as: $$\hat{\mathbf{B}} = \left(\mathbf{X}^{\mathbf{t}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{t}}\mathbf{Y}$$

Why?

$$\mathbf{Y} = \mathbf{XB}$$

$$\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \mathbf{X}^{\mathbf{T}}\mathbf{XB}$$

$$\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{XB}$$

$$\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \mathbf{IB}$$

Statistical Digression: Linear Models

OLS models *assume* error is iid

(independent, identically distributed error)

-More formally, error drawn from: $\mathcal{N}(0, \mathbf{V}_{iid})$

$$\mathbf{V}_{iid} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$\mathbf{V_{iid}}$ is an *N* x *N* identity matrix.

-It describes the fact that each specimen has identical expected variance (the 1s on diagonal), and is independent of other specimens (0s on off-diagonal)

Linear model solved as: $\boldsymbol{\beta} = (\mathbf{X^t X})^{-1}\mathbf{X^t Y}$

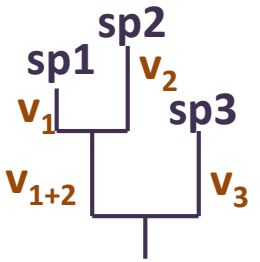
But since $\mathbf{V_{iid}}$ does nothing to matrix computations, this is the same as: $\boldsymbol{\beta} = (\mathbf{X^t V_{iid}^{-1} X})^{-1}\mathbf{X^t V_{iid}^{-1} Y}$

From OLS to GLS

With species-level data, OLS with $\mathbf{V_{iid}}$ is wrong model

-Assumes independence when the data are not

-Non-independence (species correlations) described by phylogeny



-Assuming Brownian motion, we obtain:

$$V = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$

-Using **V** yields: $\boldsymbol{\beta = (X^t V^{-1} X)^{-1} X^t V^{-1} Y}$

-This is a *weighted* model via **generalized least squares (GLS)**

Why OLS is Incorrect Here

OLS comparative model: $\boldsymbol{\beta = (X^t X)^{-1} X^t Y}$

OLS is an unweighted model: $\boldsymbol{\beta = (X^t V_{iid}^{-1} X)^{-1} X^t V_{iid}^{-1} Y}$

$$V_{id} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

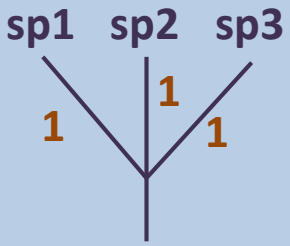
PGLS is a weighted model: $\boldsymbol{\beta = (X^t V^{-1} X)^{-1} X^t V^{-1} Y}$

$$V = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$

In PGLS, the weights are the phylogenetic distances, which describe the phylogenetic non-independence

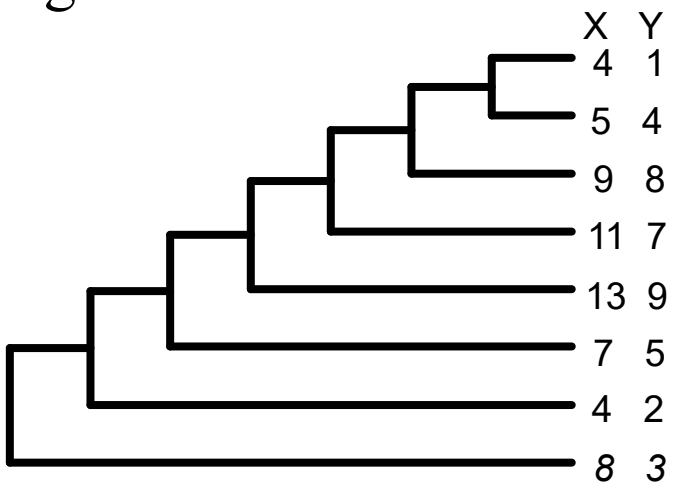
Attention! Not taking phylogeny into account, corresponds to assuming a star phylogeny

$$V_{id} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



Testing Associations: PGLS

Using PGLS:



$$\mathbf{V} = \begin{array}{c|cccccccc} & \mathbf{t1} & \mathbf{t2} & \mathbf{t3} & \mathbf{t4} & \mathbf{t5} & \mathbf{t6} & \mathbf{t7} & \mathbf{t8} \\ \hline \mathbf{t1} & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{t2} & 0 & 7 & 1 & 1 & 1 & 1 & 1 & 1 \\ \mathbf{t3} & 0 & 1 & 7 & 2 & 2 & 2 & 2 & 2 \\ \mathbf{t4} & 0 & 1 & 2 & 7 & 3 & 3 & 3 & 3 \\ \mathbf{t5} & 0 & 1 & 2 & 3 & 7 & 4 & 4 & 4 \\ \mathbf{t6} & 0 & 1 & 2 & 3 & 4 & 7 & 5 & 5 \\ \mathbf{t7} & 0 & 1 & 2 & 3 & 4 & 5 & 7 & 6 \\ \mathbf{t8} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array}$$

$$\boldsymbol{\beta} = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{Y}$$

	numDF	F-value	p-value
(Intercept)	1	12.87792	0.0115
X	1	19.28544	0.0046

$\beta_{PGLS} = 0.8846$

Recall from PICs:

```
> anova(lm(pic.y~pic.x + 0))
>       Df Sum Sq Mean Sq F value  Pr(>F)
> pic.x     1 14.3519 14.3519 19.285 0.00461 **
  Residuals 6 4.4651 0.7442 ---
```

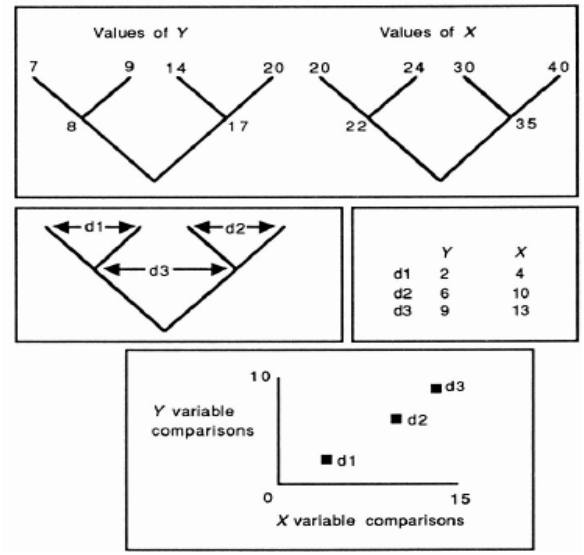
$\beta_{PIC} = 0.8846$

PIC & PGLS yield identical results!

PIC vs. PGLS

PIC and PGLS seem *VERY* different computationally

PIC (Felsenstein 1985)

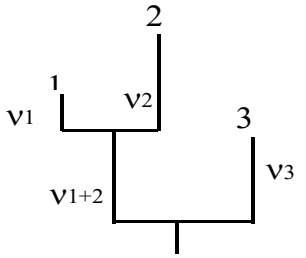


PGLS (Grafen 1989; Martins and Hansen 1997)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Error: $\mathcal{N}(0, \mathbf{V})$

$$\mathbf{V} = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$



$$\boldsymbol{\beta} = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{Y}$$

Both condition the data on the phylogeny (one via contrasts and one during the regression)

If implemented correctly, both yield identical **β** and model p-values

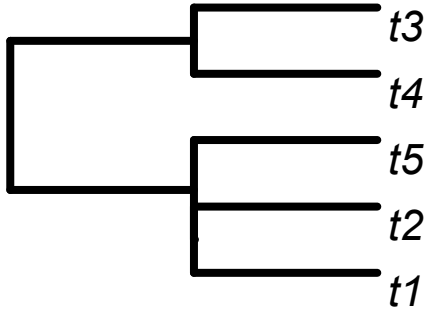
-PIC a special case of PGLS

Garland and Ives (2000)
Rohlf (2001)
Blomberg (2012)
Adams and Collyer (2018a)

PGLS: Greater Flexibility

PGLS is more flexible than PIC

-Polytomies easily accommodated

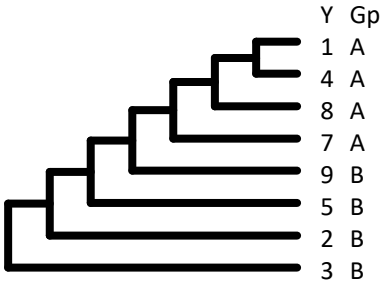


Blocks with identical values

$\mathbf{V} =$

	t1	t2	t5	t4	t3
t1	2	1	1	0	0
t2	1	2	1	0	0
t5	1	1	2	0	0
t4	0	0	0	2	1
t3	0	0	0	1	2

-Regression, ANOVA, and factorial models possible ($\mathbf{Y} \sim \mathbf{X_1} + \mathbf{X_2}$)



	numDF	F-value	p-value
(Intercept)	1	3.0624998	0.1307
grp	**1**	**0.0131349**	**0.9125**

-Other evolutionary covariance models (e.g., OU) may be implemented by changing expected covariance in **V** (later in semester)

PGLS is preferred implementation over PIC

Phylogenetic Transformation

Statistically OLS and GLS are the same algebra: $\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \varepsilon$

$$\boldsymbol{\beta}_{OLS} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = \mathbf{X}^t\mathbf{V}_{iid}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}_{iid}^{-1}\mathbf{Y}$$

$$\boldsymbol{\beta}_{GLS} = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{Y}$$

The difference is in the error covariance structure:

$$V_{id} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$V = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$

In statistics, there is a standard GLS→OLS transformation procedure based on error covariance transformation

-Phylogenetic transformation uses this approach

GLS→OLS in statistics:
Judge et al. (1985)
Johnson and DiNardo (1997)
Rencher (2000)

Phylogenetic Transformation

-*Condition* data on phylogeny *prior* to statistical evaluation

1: Obtain phylogenetic transformation matrix, **P**:

a) Eigen-decomposition of **V**:

$$\mathbf{V} = \mathbf{U}\mathbf{W}\mathbf{U}^{-1}$$

-This represents the characteristic information found in **V**, expressed in different way (generates a set of 'basis' vectors which express variation in V orthogonally)

b) Generate **P** as:

$$\mathbf{P} = (\mathbf{U}\mathbf{W}^{1/2}\mathbf{U}^{\mathbf{T}})^{-1}$$

P is an *N* x *N* matrix

-This expresses the information in **V** using orthogonal axes, **U**

Garland and Ives (2000)
Adams (2014)
Adams and Collyer (2018a)

Phylogenetic Transformation

-*Condition* data on phylogeny *prior* to statistical evaluation

2: Project data (**X** & **Y**) on **P**:

$$\tilde{\mathbf{X}} = \mathbf{PX} \quad \text{and} \quad \tilde{\mathbf{Y}} = \mathbf{PY}$$

$$X_R = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad X_F = \begin{bmatrix} 1 & X_1 \\ 1 & \vdots \\ 1 & X_n \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

This conditions **X** & **Y** on phylogeny, rendering the values independent of evolutionary history

*NOTE: Need to do for both $\mathbf{X_F}$ and $\mathbf{X_R}$ to compare models

Garland and Ives (2000)
Adams (2014)
Adams and Collyer (2018a)

Phylogenetic Transformation

-*Condition* data on phylogeny *prior* to statistical evaluation

3: Find β using OLS

$$\hat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}^{\boldsymbol{t}}\ \tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{\boldsymbol{t}}\ \tilde{\mathbf{Y}}$$

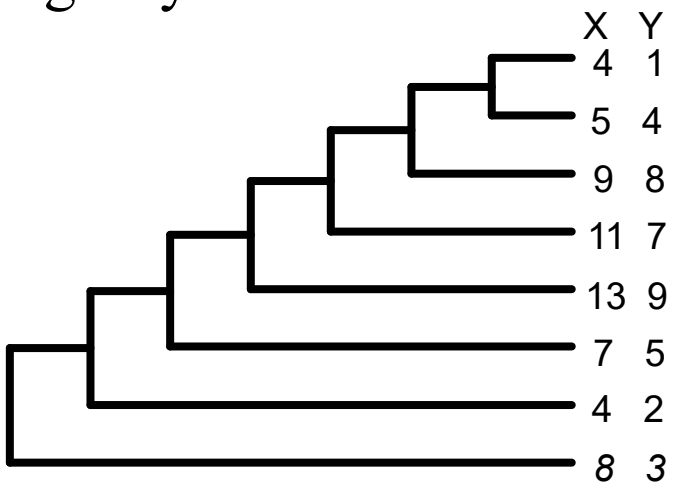
-ε of model now $\mathcal{N}(0, \mathbf{I})$

4: Significance evaluated using parametric methods (LRT and/or F-ratios), or permutation

Garland and Ives (2000)
Adams (2014)
Adams and Collyer (2018a)

Testing Associations: Phylo-Transform

Using Phylo-Transform:



V =

	t1	t2	t3	t4	t5	t6	t7	t8
t1	7	0	0	0	0	0	0	0
t2	0	7	1	1	1	1	1	1
t3	0	1	7	2	2	2	2	2
t4	0	1	2	7	3	3	3	3
t5	0	1	2	3	7	4	4	4
t6	0	1	2	3	4	7	5	5
t7	0	1	2	3	4	5	7	6
t8	0	1	2	3	4	5	6	7

P =

	t1	t2	t3	t4	t5	t6	t7	t8
t1	0.378	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.00000
t2	0.000	0.3845	-0.0184	-0.0143	-0.0111	-0.0087	-0.0071	-0.0071
t3	0.000	-0.0184	0.4027	-0.0343	-0.0265	-0.0207	-0.0168	-0.0168
t4	0.000	-0.0143	-0.0343	0.4332	-0.0511	-0.0394	-0.0316	-0.0316
t5	0.000	-0.0111	-0.0265	-0.0511	0.4801	-0.0739	-0.0581	-0.0581
t6	0.000	-0.0087	-0.0207	-0.0394	-0.0739	0.5552	-0.1164	-0.1164
t7	0.000	-0.0071	-0.0168	-0.0316	-0.0581	-0.1164	0.6982	-0.3018
t8	0.000	-0.0071	-0.0168	-0.0316	-0.0581	-0.1164	-0.3018	0.6982

$$\boldsymbol{\beta} = \left(\tilde{\mathbf{X}}^t\ \tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^t\ \tilde{\mathbf{Y}}$$

	Df	SS	MS	Rsq	F	Z	Pr(>F)
X	1	14.3519	14.3519	0.76271	19.285	1.734	0.006 **
Residuals	6	4.4651	0.7442	0.23729			

$\beta_{PTrans} = 0.8846$

Recall from PIC & PGLS:

```
> anova(lm(pic.y~pic.x + 0))
>       Df Sum Sq Mean Sq F value Pr(>F)
> pic.x  1 14.3519 14.3519 19.285 0.00461 **
  Residuals 6 4.4651 0.7442 ---
```

$\beta_{PIC} = 0.8846$

Same as PIC & PGLS!

PIC, PGLS, Phylo-Transformation

PIC, PGLS, Phylo-transform

-3 implementations of phylogenetic comparative methods

-All condition data on phylogeny

-All yield identical regression coefficients & parameters

Which to use*?

-PIC restricted primarily to regression

-PGLS & Phylo-Transform more general: ANOVA, regression, etc.

-PGLS: BM & non-BM models (e.g., OU, EB)

-Phylo-Transform: can better accommodate multivariate **Y**

Garland and Ives (2000)
Rohlf (2001)
Blomberg (2012)
Adams and Collyer (2018a)

*Note: phylogenetic mixed model was also proposed (Lynch 1991)

PCM: Assessing Significance

PIC, PGLS, and phylo-transform yield identical parameters and coefficients

$$\widehat{\mathbf{B}} = \left(\mathbf{X^t V^{-1} X}\right)^{-1} \mathbf{X^t V^{-1} Y}$$
$$= \left(\mathbf{X}_{\boldsymbol{pic}}^{\boldsymbol{t}} \mathbf{X}_{\boldsymbol{pic}}\right)^{-1} \mathbf{X}_{\boldsymbol{pic}}^{\boldsymbol{t}} \mathbf{Y}_{\boldsymbol{pic}}$$
$$= \left(\widetilde{\mathbf{X}}^{\boldsymbol{T}} \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^{\boldsymbol{T}} \widetilde{\mathbf{Y}}$$

How do we evaluate them statistically?

-Standard approach: parametric methods

A) F-ratios: MS_F / MS_R (LS solutions: equivalent to LRT comparing models)

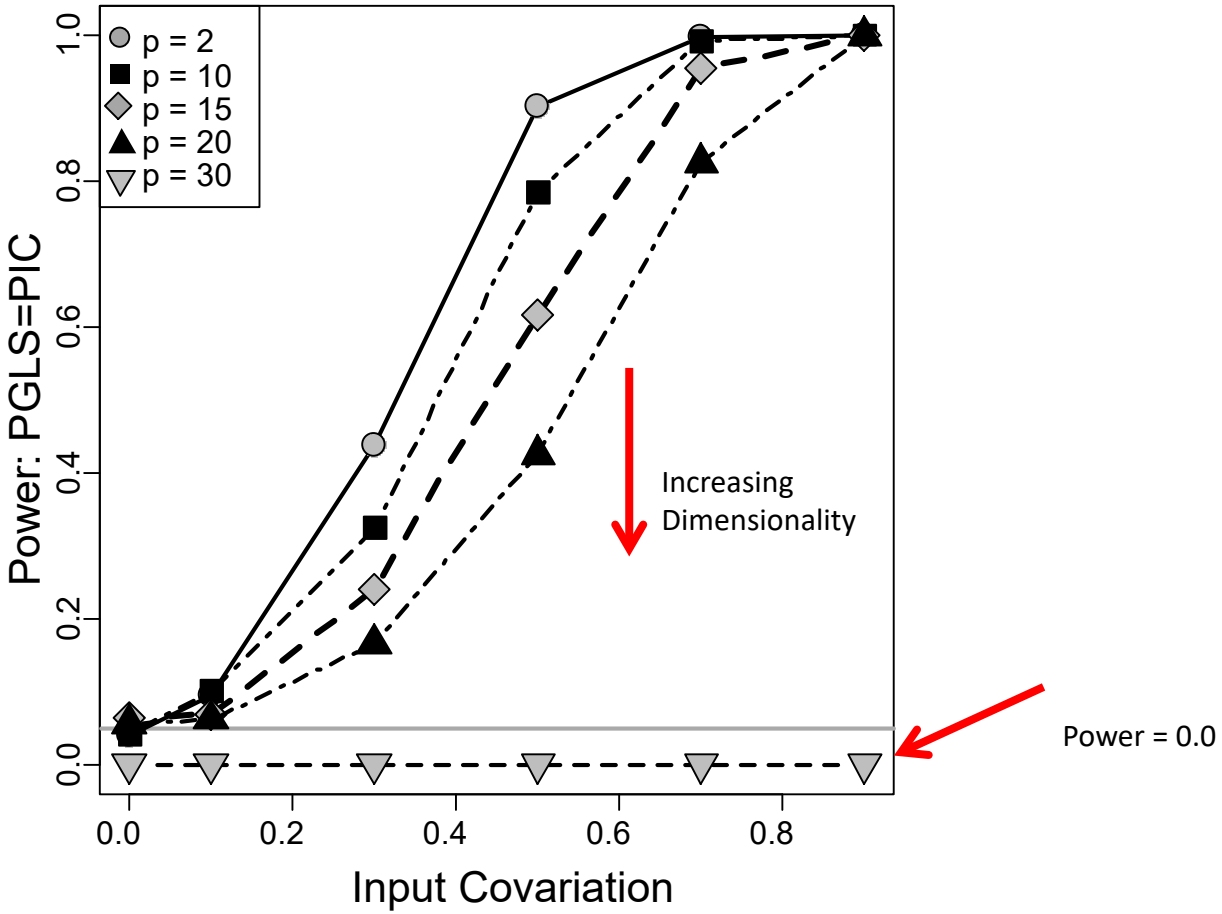
B) Optimize log$\mathcal{L}$ for model (popular)

$$\log\mathcal{L} = \log\left[\frac{\exp\left(-\frac{1}{2}(\mathbf{Y}-E(\mathbf{Y}))^t(\mathbf{V})^{-1}(\mathbf{Y}-E(\mathbf{Y}))\right)}{\sqrt{(2\pi)^N \times |\mathbf{V}|}}\right]$$

Adams and Collyer (2018a)

PROBLEM: Parametric PCMs suffer from Rao's paradox

-Reduced power with higher data dimensionality

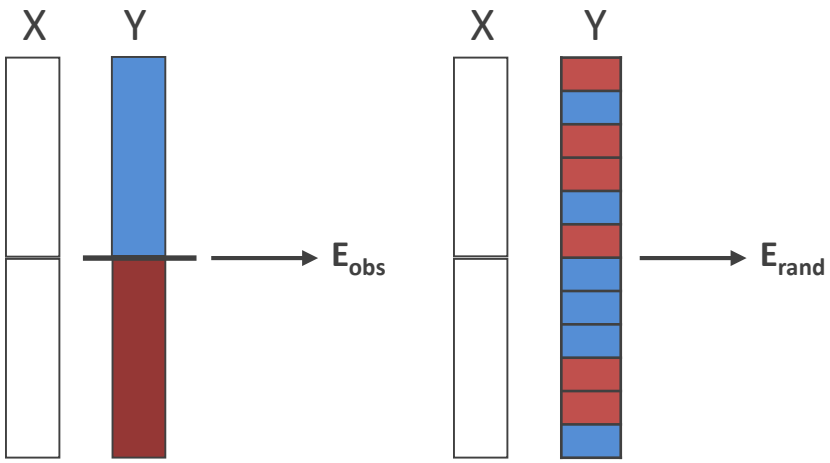


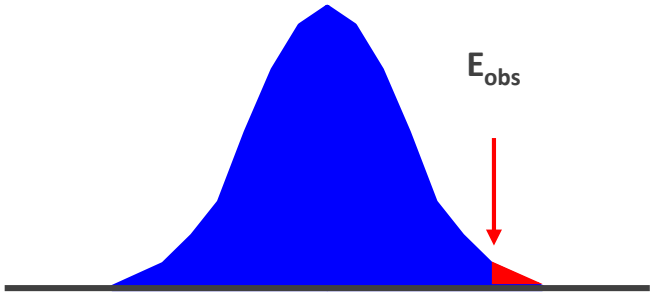
Adams (2014)
Adams and Collyer (2018a)

PCM: Assessing Significance

Alternative: Permutation methods

Permute data in some way to generate distribution of possible outcomes under H_0





For PCMs, what does one permute?

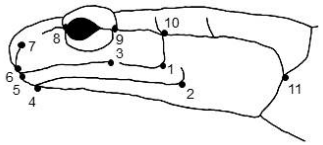
Must identify correct exchangeable units under H_0

Adams and Collyer (2018a)

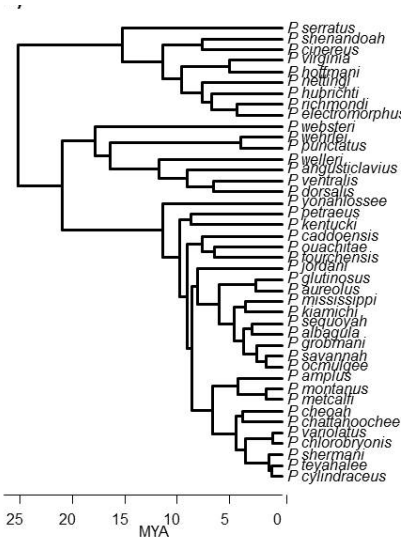
PCM: Assessing Significance

Permuting Y_{PIC} (e.g., Klingenberg & Marugán-Lobón [2013]) is incorrect

Results in elevated type I error rates

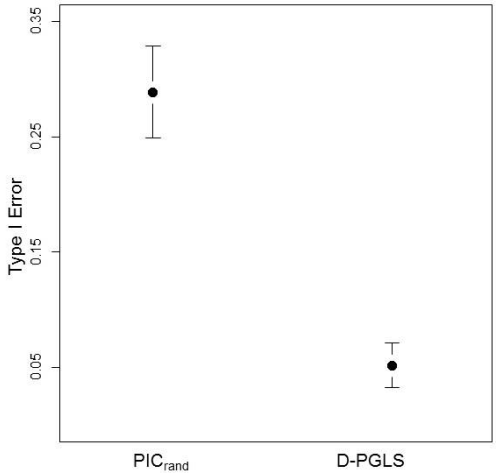


vs. SVL



D-PGLS	df	SS	MS	F	R^2	P_{Yrand}
SVL	1	0.0006586	0.0006586	3.0288	0.07039	0.221 NS
Residual	40	0.0086976	0.0086976			
Total	41	0.0093562	0.00021744			

PIC	df	SS	MS	F	R^2	$P_{PICrand}$	P_{Yrand}
SVL	1	0.0006586	0.0006586	3.0288	0.07039	0.026	0.221 NS
Residual	40	0.0086976	0.0086976				
Total	41	0.0093562	0.00021744				



Method should not be used

Adams and Collyer (2015)
Adams and Collyer. (2018a)

PCM: Assessing Significance

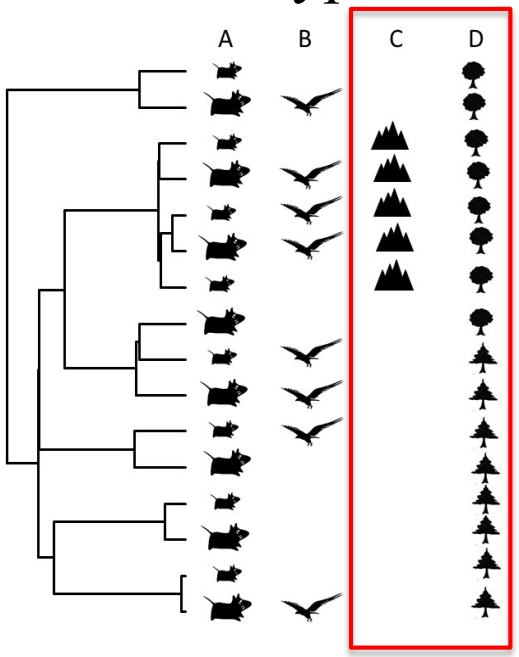
One could shuffle Y, then perform phylo-transform

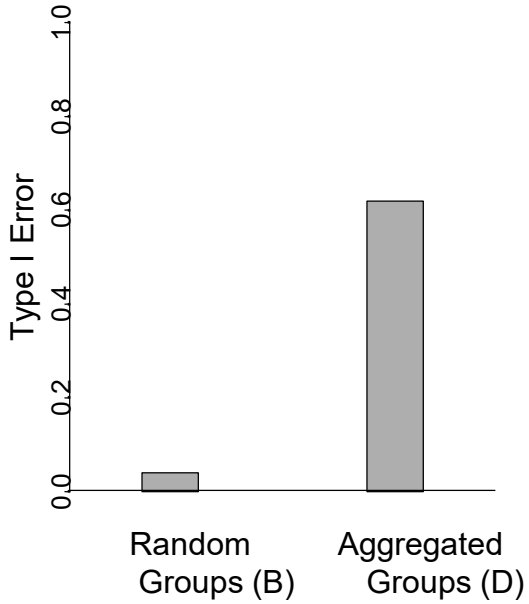
-Appropriate type I error and power for regression
-Slightly elevated type I error in some circumstances

Adams (2014)
Adams and Collyer (2015)

Goolsby (2016)
Adams and Collyer (2018)

-VERY HIGH type I error for ANOVA of aggregated groups





Method not a general solution for all statistical designs

Adams (2014)
Adams and Collyer (2015)
Adams and Collyer. (2018a)
Adams and Collyer (2018b)

PCM: Assessing Significance

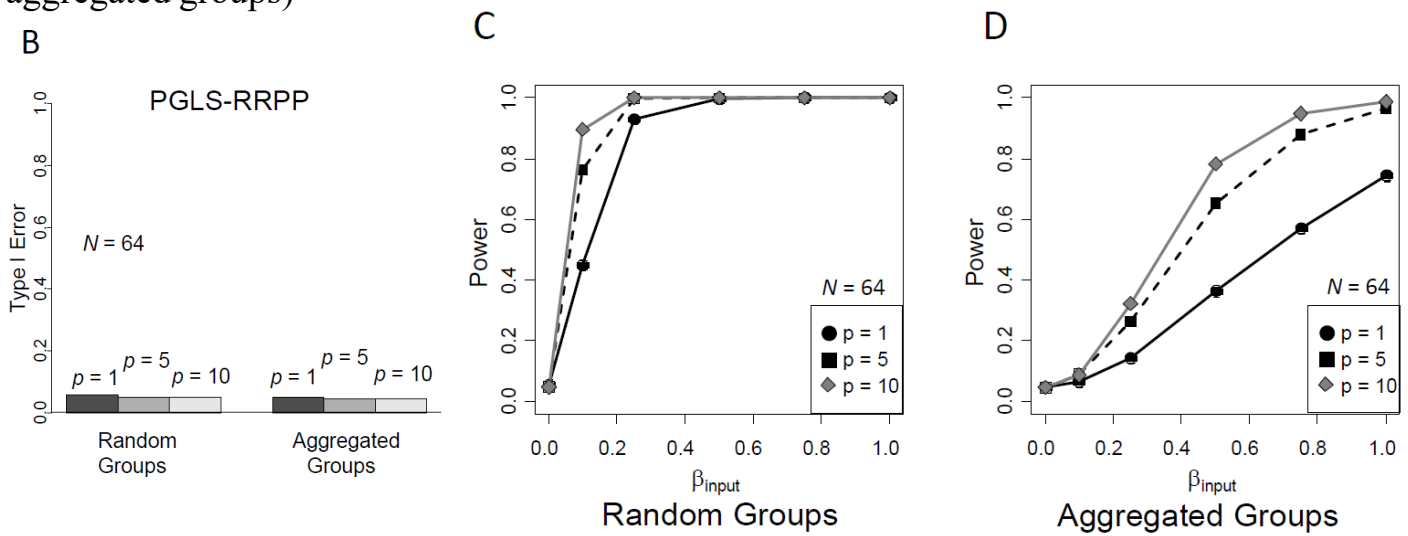
Phylo-transform first: RRPP*

-Transform data: $\tilde{\mathbf{X}} = \mathbf{PX}$ and $\tilde{\mathbf{Y}} = \mathbf{PY}$

-Run model: $\boldsymbol{\beta} = \left(\tilde{\mathbf{X}}^{t}\ \tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{t}\ \tilde{\mathbf{Y}}$

-Shuffle residuals ($\tilde{\boldsymbol{\varepsilon}}$) from reduced model; assess significance

-Appropriate type I error, power, bias, etc. (though note, correctly, that power decreases with aggregated groups)



Provides general solution for all phylogenetic linear models

-RULE: Transform data first, shuffle residuals second!

*RRPP: Residual Randomization Permutation Procedure: We'll discuss later in semester

Adams and Collyer (2019)

Testing Associations: Conclusions

PCMs condition data on phylogeny

3 implementations: PIC, PGLS, Phylo-transform

-All yield identical regression coefficients & parameters

Comparisons

-PIC restricted primarily to regression

-PGLS & Phylo-Transform more general: ANOVA, regression, etc.

-PGLS: BM & non-BM models (e.g., OU, EB)

-Parametric significance testing problematic with $\mathbf{Y}_{\mathbf{mult}}$

-Phylo-Transform + RRPP most flexible for $\mathbf{Y}_{\mathbf{mult}}$

Garland and Ives (2000)
Rohlf (2001)
Blomberg (2012)
Adams and Collyer (2018)