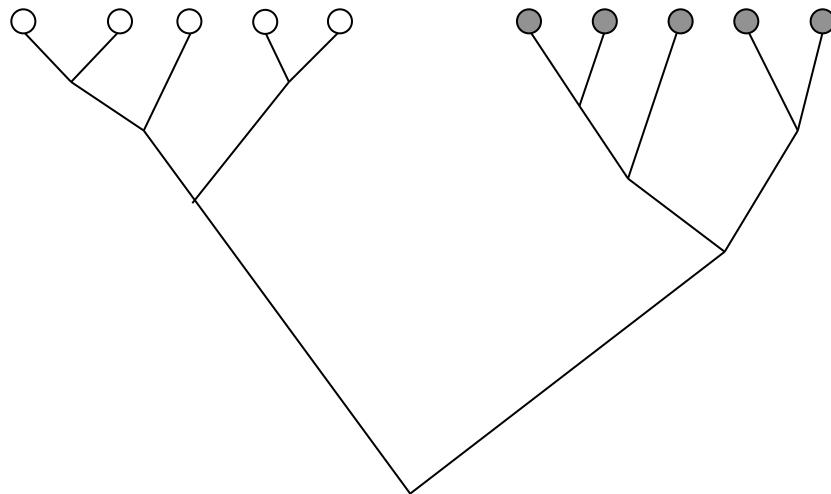
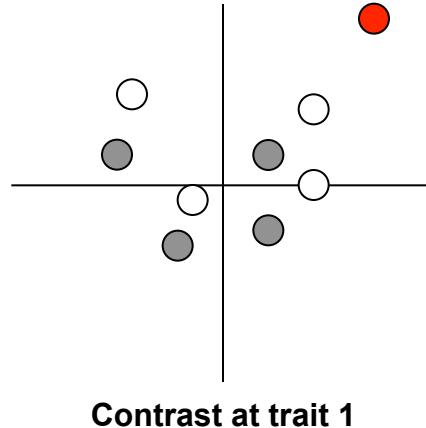


An introduction to PGLS

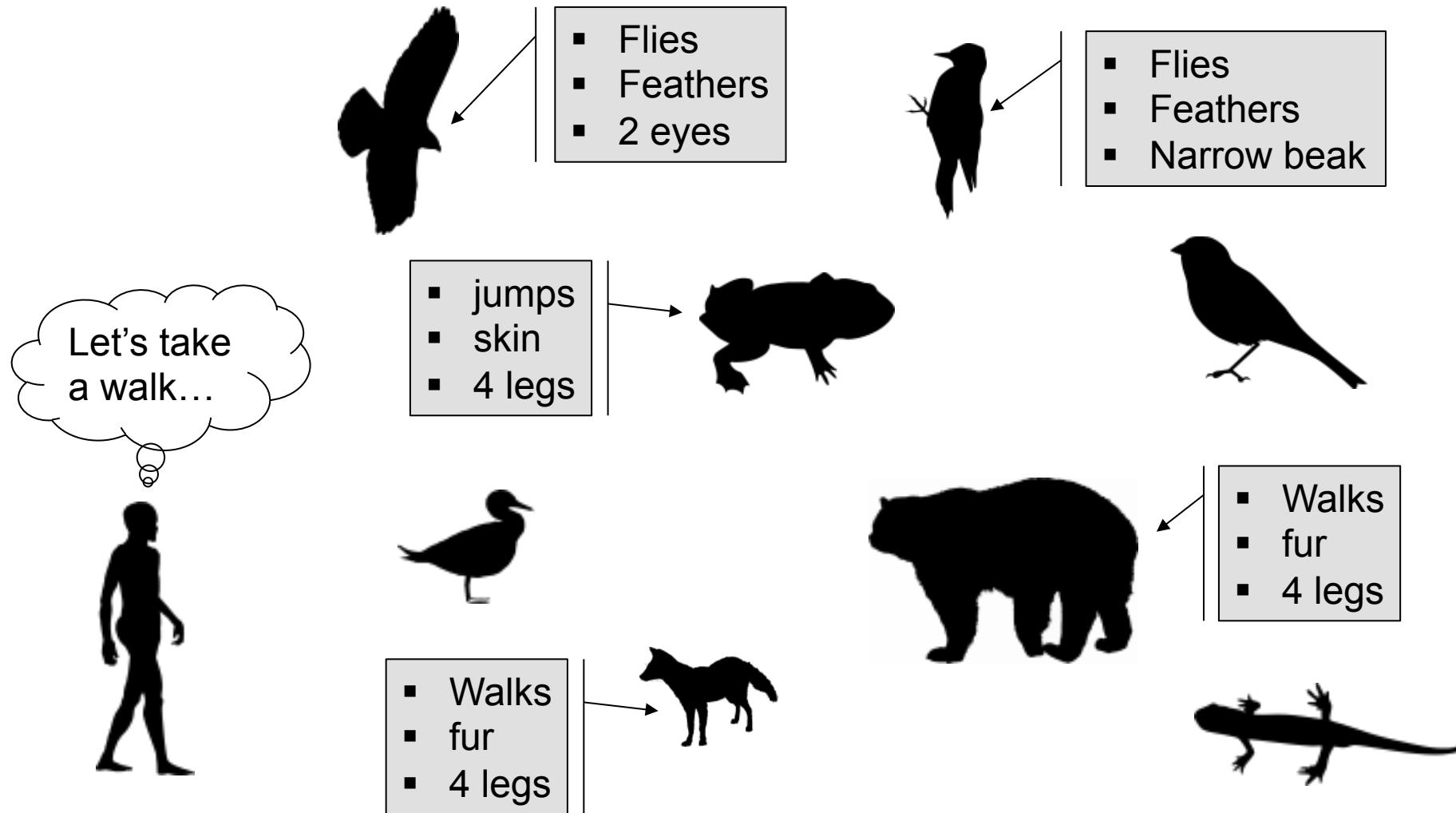


Contrast at trait 2



An introduction

An introduction



An introduction

	feathers	other
Animal flies	9	0
Animal doesn't fly	0	7

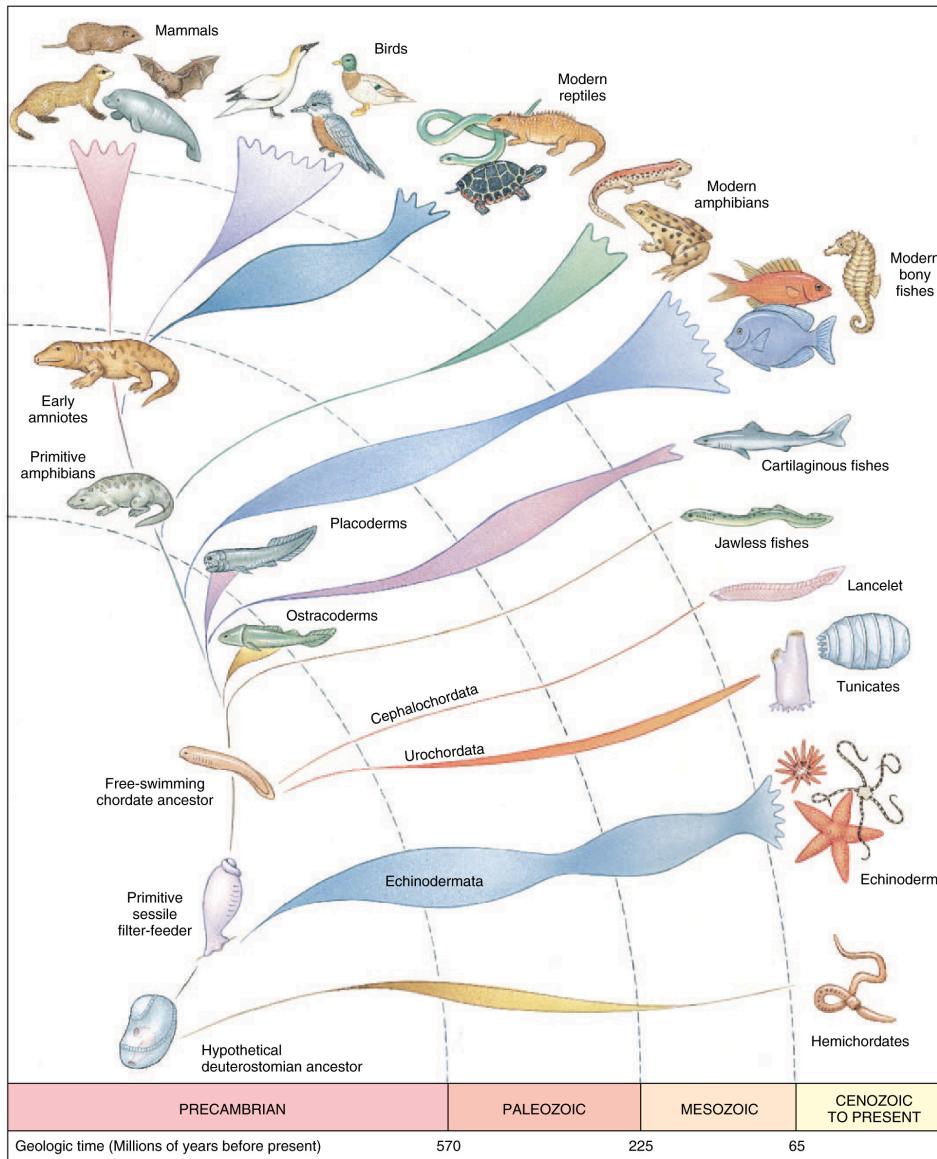


```
> fisher.test(matrix(c(9,0,0,7),2))
```

Fisher's Exact Test for Count Data

```
data: matrix(c(9, 0, 0, 7), 2)
p-value = 8.741e-05
```

Evolution of flight (and feathers)



Credit: University Of Kelaniya, Sri Lanka

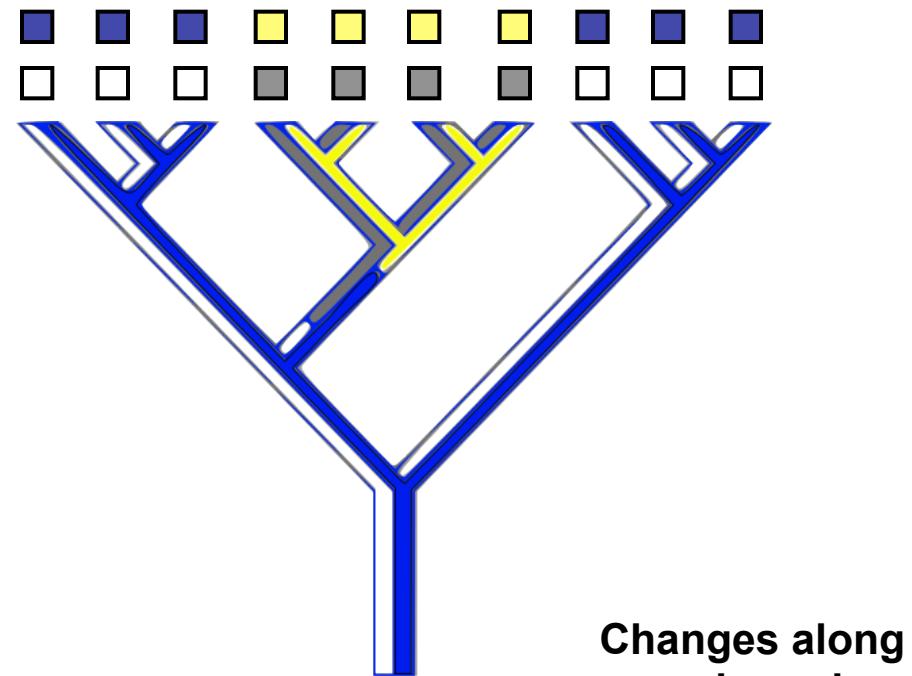
The comparative method

The comparative method

- Definition
 - It is the name given to a family of methods that aims at correcting for the phylogenetic non-independence of species when comparing them
- Why ?
 - Species cannot be considered without the context of evolution; the observations are not independent

Another qualitative example

species	flower	stem
1	blue	spiny
2	blue	spiny
3	blue	spiny
4	blue	spiny
5	yellow	smooth
6	yellow	smooth
7	yellow	smooth
8	yellow	smooth
9	blue	spiny
10	blue	spiny



blue yellow

□ spiny

■ smooth

6	0
0	4

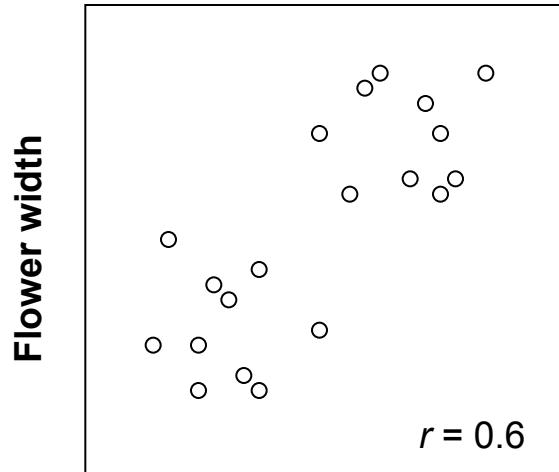
Fisher exact test
 $p = 0.00476$

Fisher exact test
 $p = 0.056$

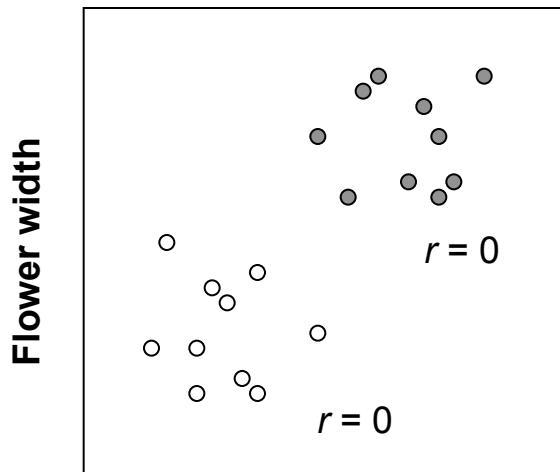
stem

flower		stem	
yes	no	yes	no
1	0	1	0
0	17	0	17

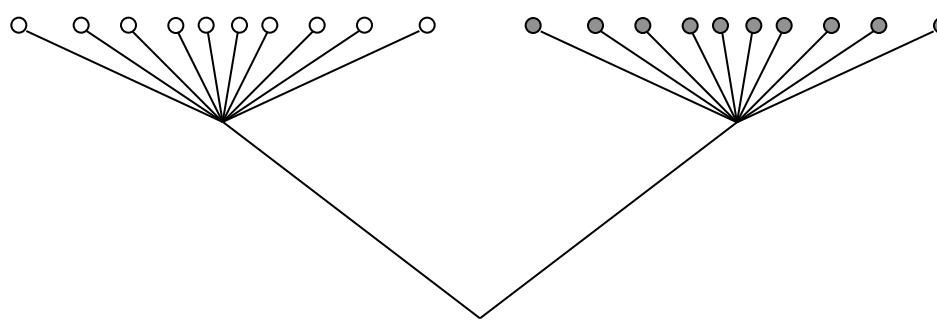
A quantitative example



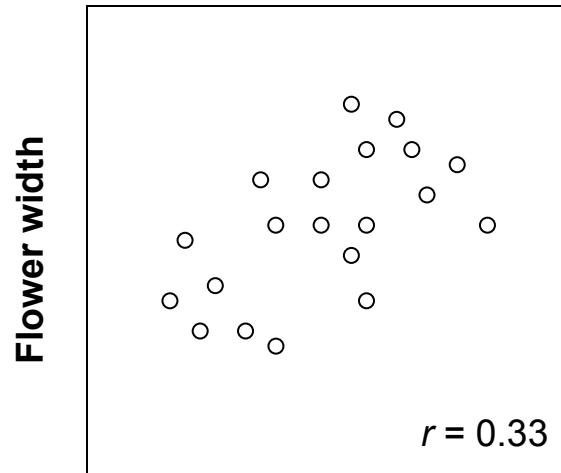
Trichome number
on the leaves



Trichome number
on the leaves

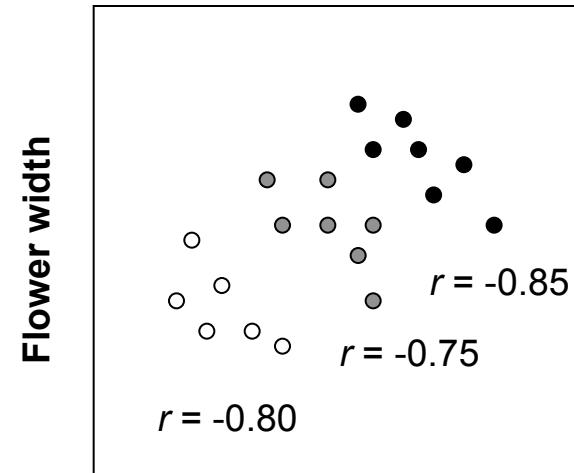


Another quantitative example



Trichome number
on the leaves

$r = 0.33$

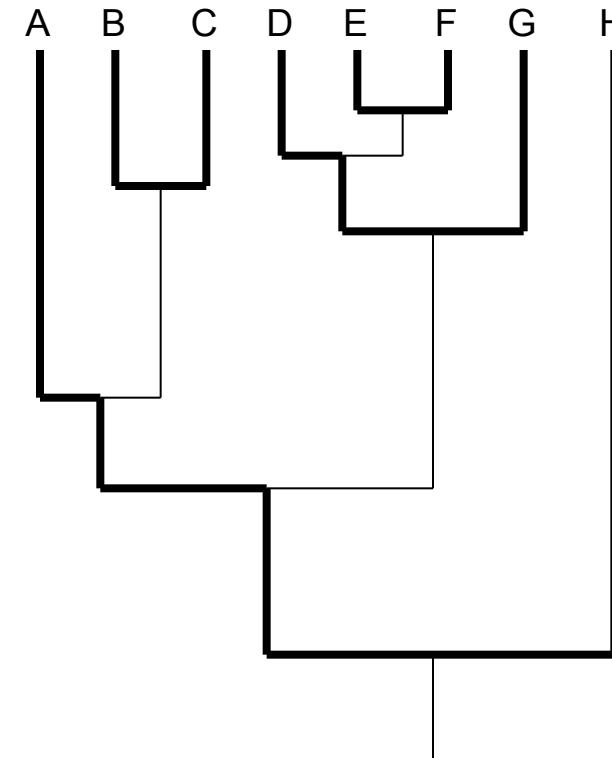
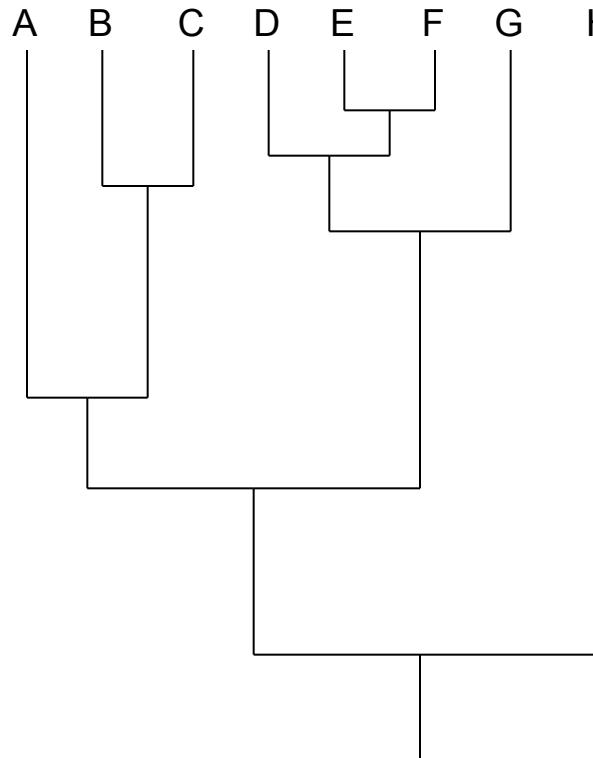


Comparative methods

Different approaches

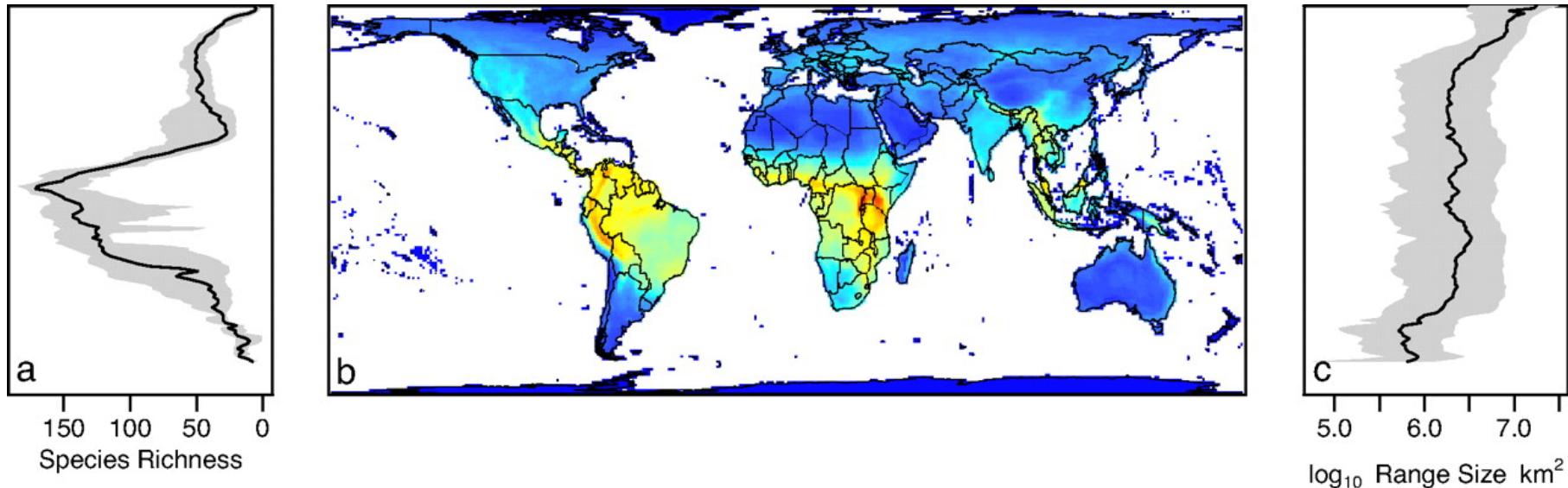
- Independent contrasts
- Pairs of species
- Phylogenetic independent contrasts

Pairs of species (sometimes sister clades)

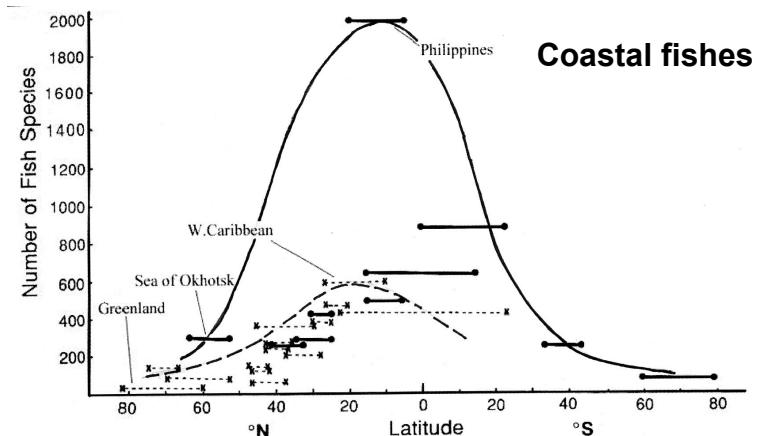


We can thus treat each comparison as independent from each other

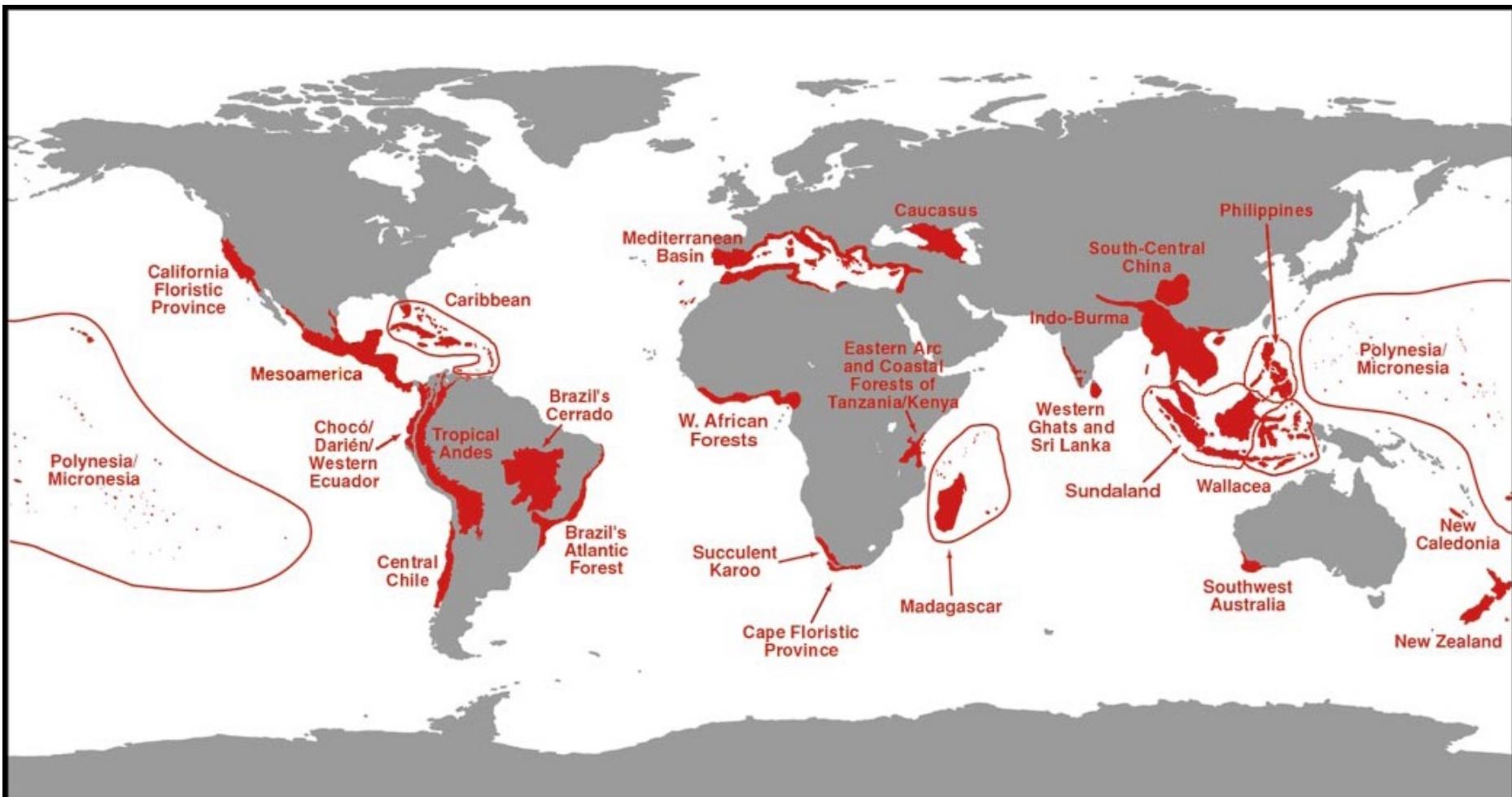
Biodiversity latitudinal gradient



Mammals - Davies T J et al. PNAS 2008;105:11556-11563

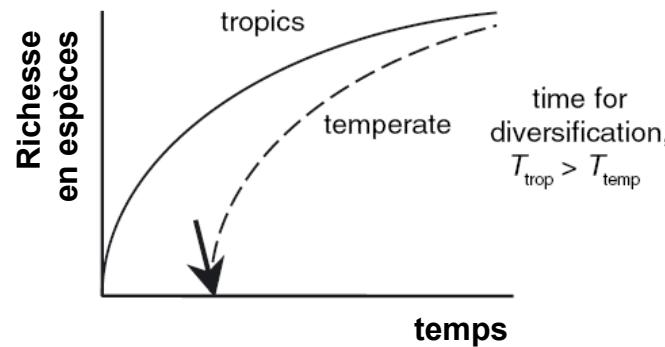


25 biodiversity hotspots

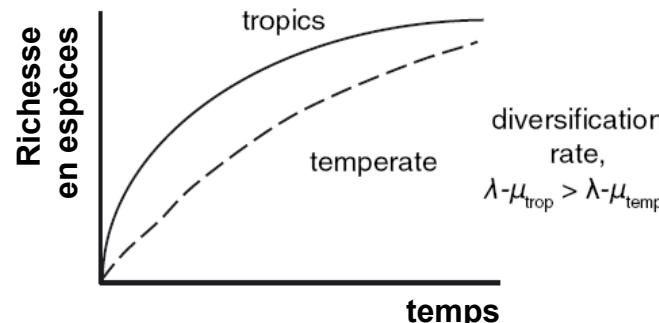


Some evolutionary hypotheses

- More time for speciation in the tropics (older environments)



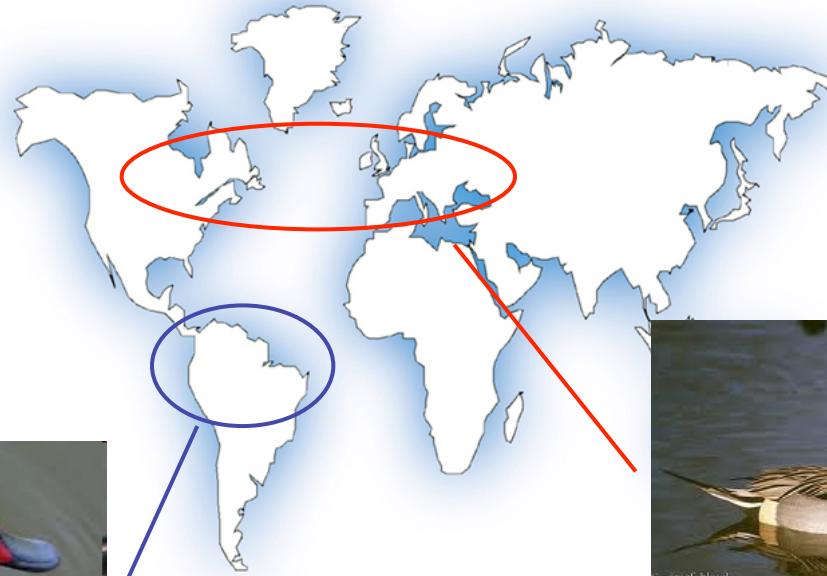
- Higher diversification rates in the tropics, due to:
 - Higher speciation rates in the tropics (Cradle theory)
 - Lower extinction rates in the tropics (Museum theory)



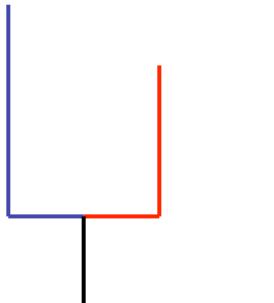
Test for higher diversification rate in the tropics



Anas bahamensis



Anas acuta

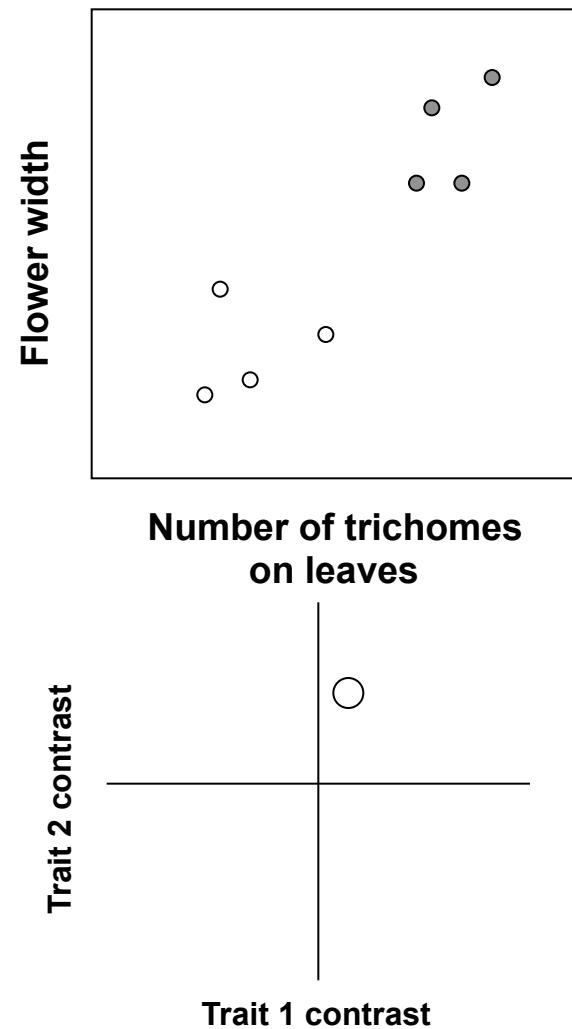
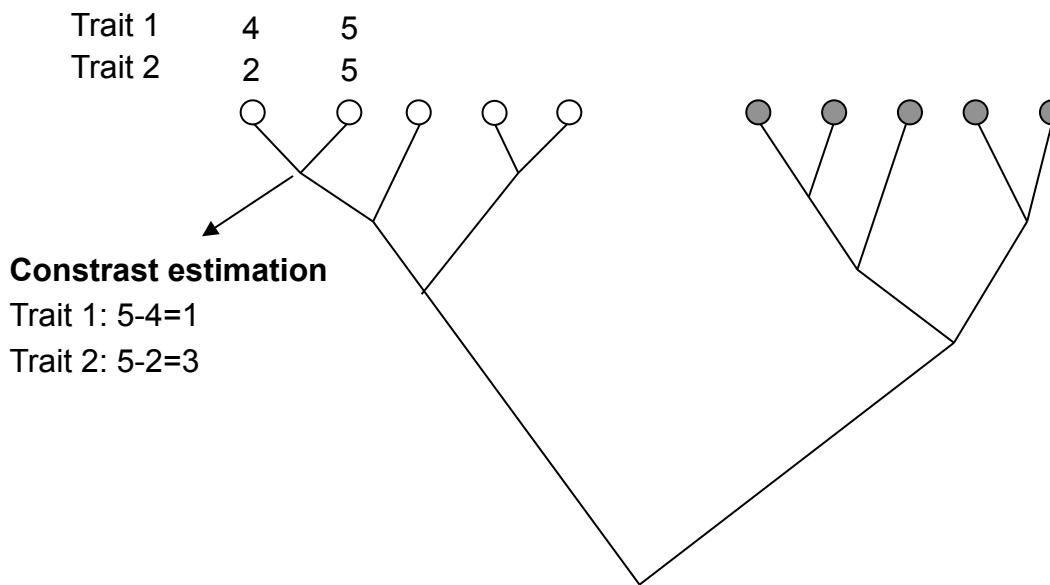


Number of contrasts	Number of positive contrasts	Sign test	Wilcoxon signed rank test
33	15	0.76	0.36

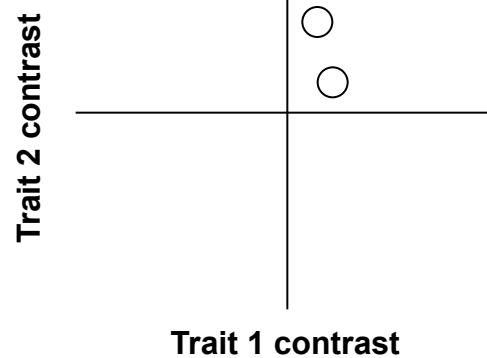
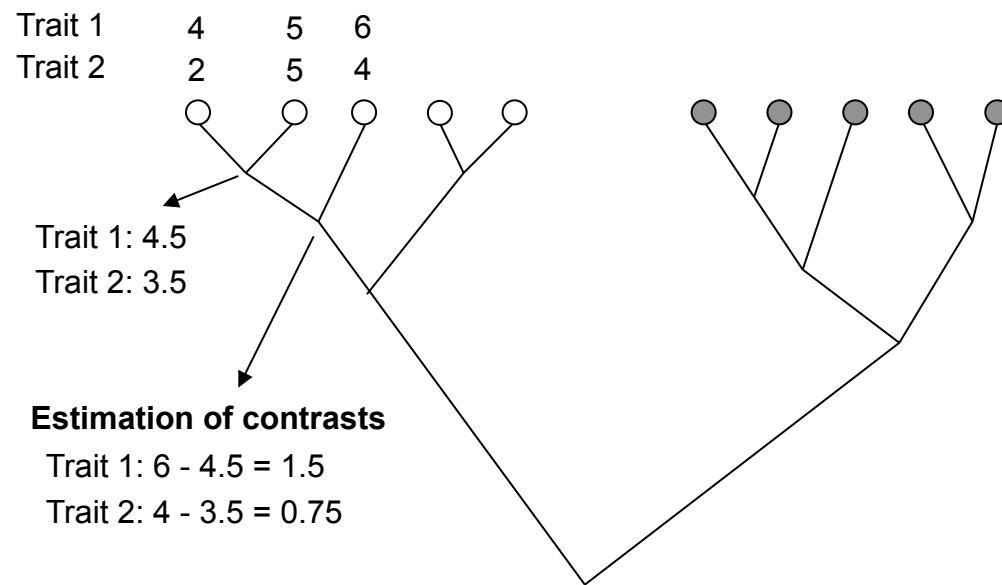
Not significant

Independent contrasts

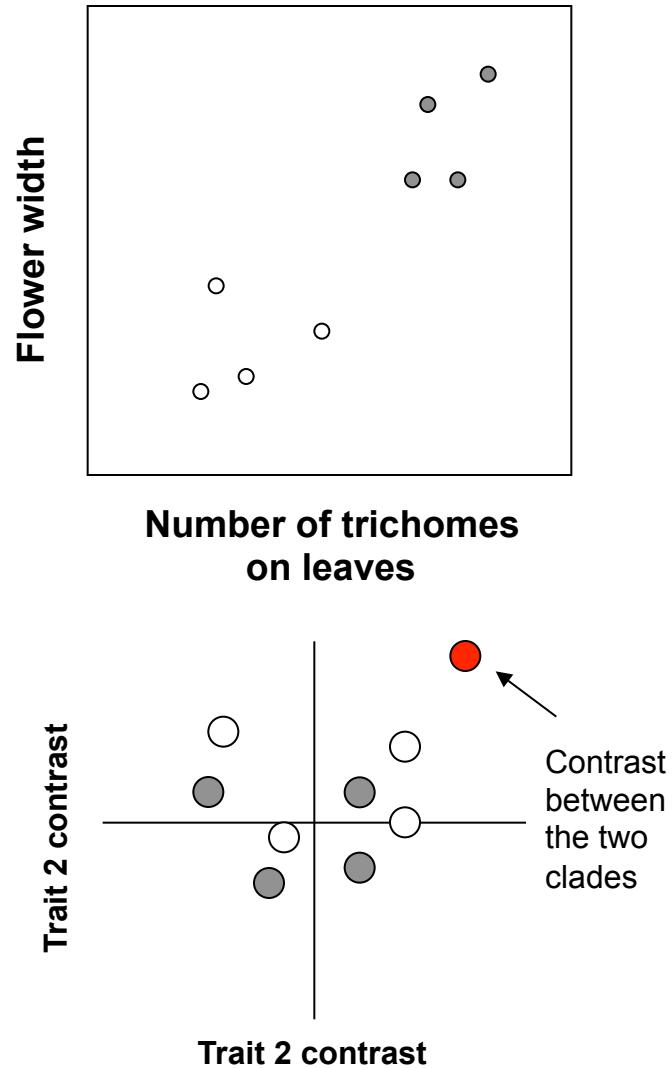
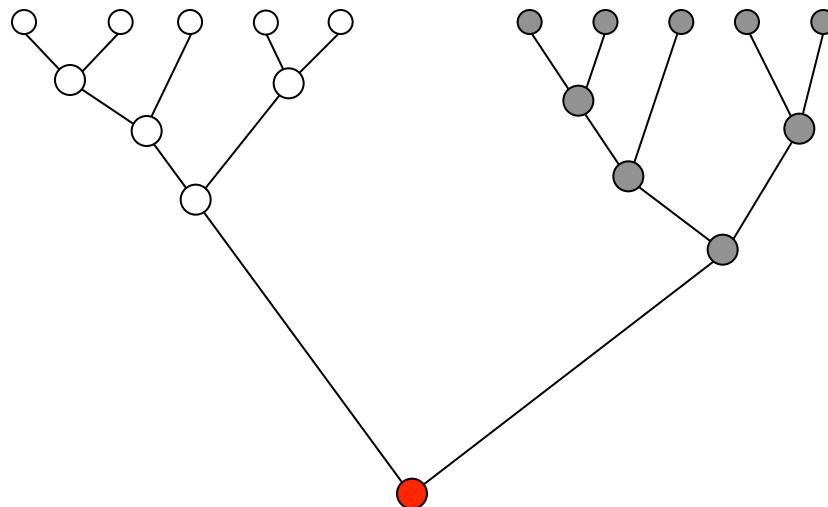
Objective: Test the correlation between characters while taking into account the evolutionary history of species



Independent contrasts

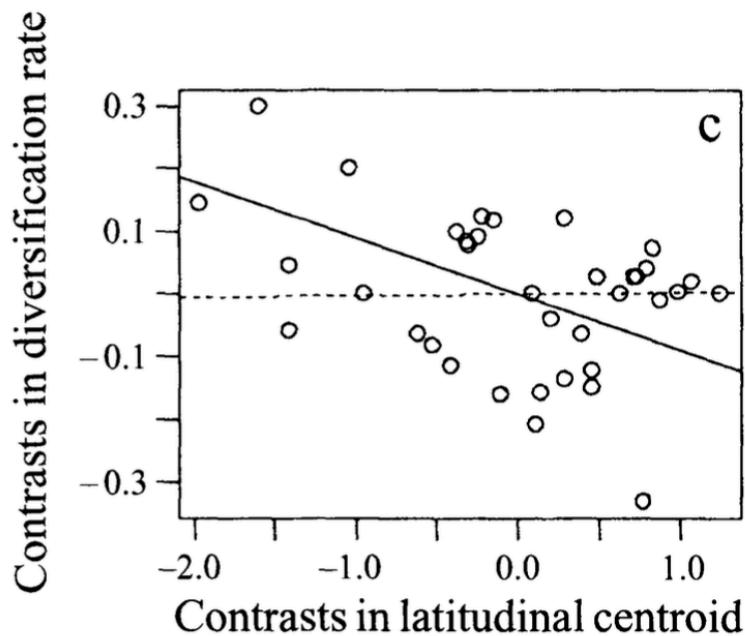


Independent contrasts



Test for higher diversification rate in the tropics

Test: Using a bird genera phylogeny, Cardillo et al. (2005) estimated for each genera their diversification rate and their mean latitude. They then estimated the PIC to test if there is a correlation between latitude and the diversification rate.



$P\text{-value} = 0.023$

Higher diversification rates
in the tropics

Phylogenetic generalized least squares

Linear model (ordinary least squares – OLS)

$$Y_i = \alpha + \beta_i X_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Many assumptions

- Normality, homogeneity, fixed X , independence, and correct model specification
- Independence:
 - The value of Y_i at X_i should not be influenced by other values of X_i

An example with real data

Seed plant functional traits (Paquette et al. 2015)

```
##   Code      Species.name Occurrence maxH   Wd    Sm Shade   N
## 1 ABBA     Abies balsamea    7759    25 0.34   7.6  5.0 1.66
## 2 ACNE     Acer negundo      0    20 0.44  34.0  3.5 2.50
## 3 ACNI     Acer nigrum       1    30 0.52  65.0  3.0 1.83
## 4 ACPE Acer pensylvanicum   665    10 0.44  41.0  3.5 2.22
## 5 ACPL     Acer platanoides     0    15 0.51 172.0  4.2 1.99
## 6 ACRU     Acer rubrum      3669    25 0.49  20.0  3.4 1.91
```

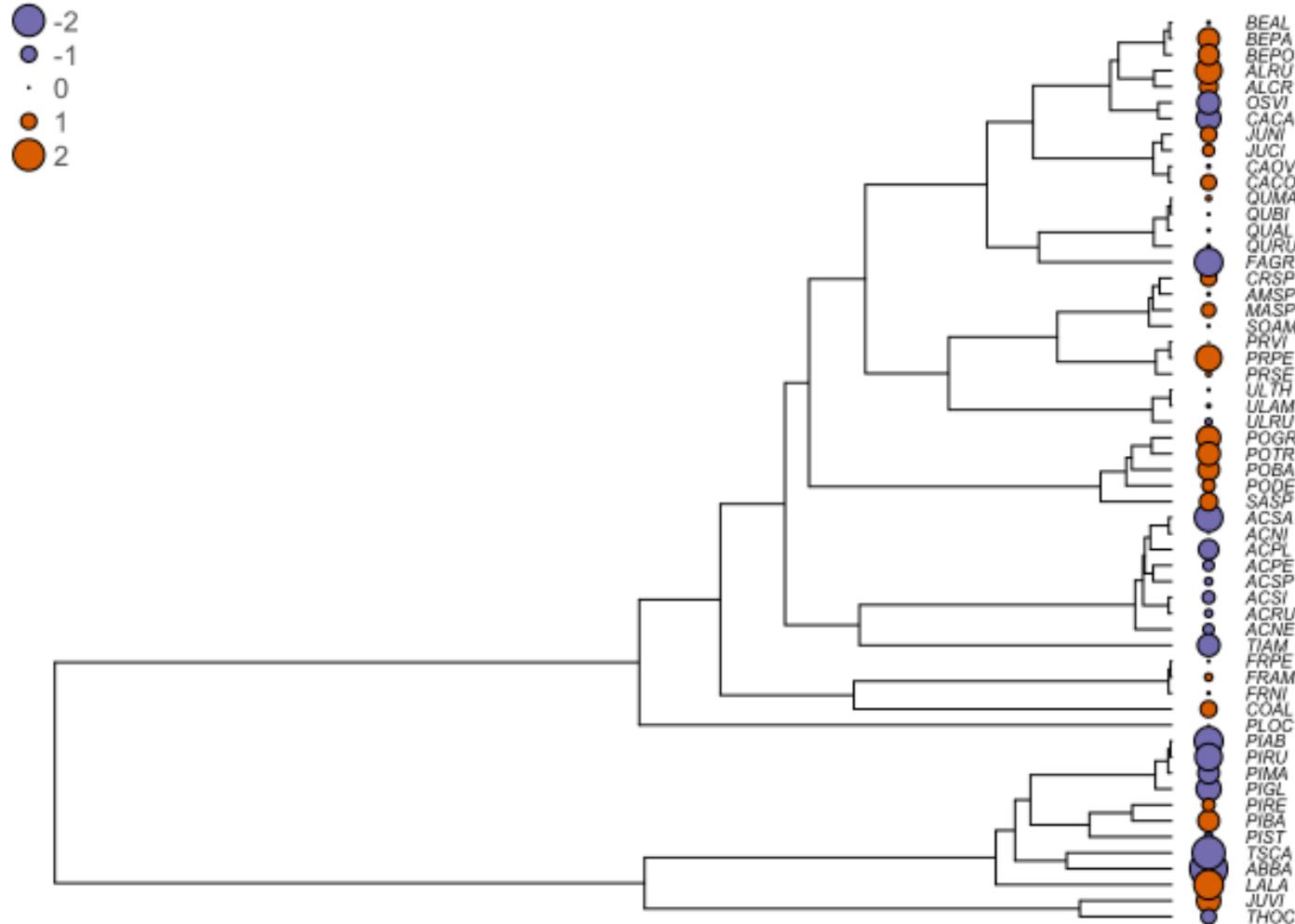
Linear model in R

```
# Fit a linear model
shade.lm <- lm(Shade ~ Wd, data = seedplantsdata)
summary(shade.lm)
```

Get model residuals

```
# Extract residuals
shade.res <- residuals(shade.lm)
```

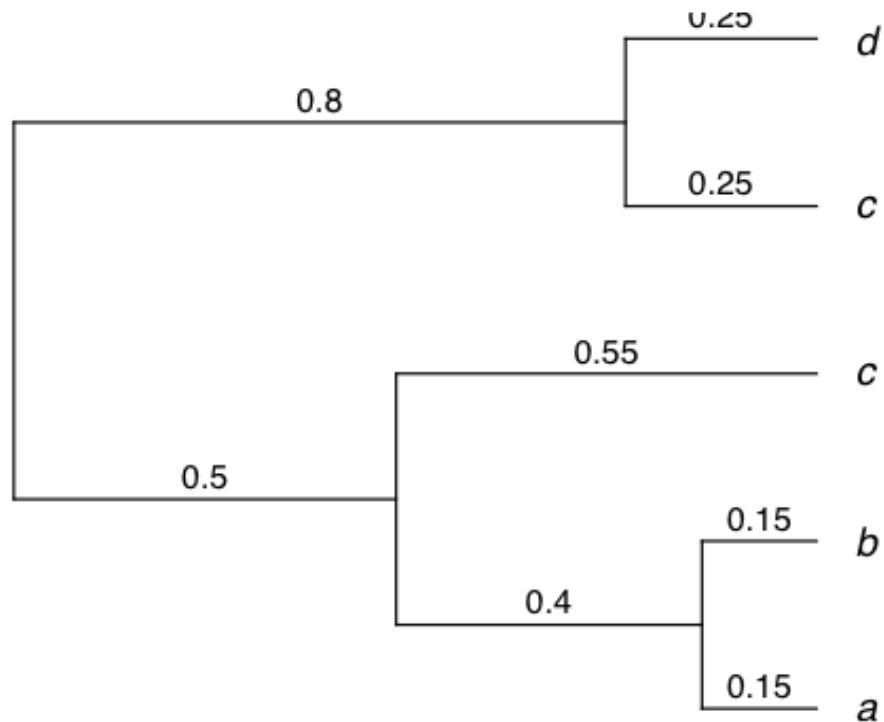
Correlation in the residuals



Phylogenetic generalized least squares (PGLS)

- A special case of generalized least squares (GLS)
 - Allow the residuals of the model to be correlated in a specific way
 - Used for spatial correlation, time series, phylogenetic relationships, etc.
- Need to have a correlation structure!

From a tree to a variance-covariance structure



	a	b	c	c	d
a	1.05	0.90	0.50	0.00	0.00
b	0.90	1.05	0.50	0.00	0.00
c	0.50	0.50	1.05	0.00	0.00
c	0.00	0.00	0.00	1.05	0.80
d	0.00	0.00	0.00	0.80	1.05

... to a correlation matrix

	a	b	c	c	d
a	1.05	0.90	0.50	0.00	0.00
b	0.90	1.05	0.50	0.00	0.00
c	0.50	0.50	1.05	0.00	0.00
c	0.00	0.00	0.00	1.05	0.80
d	0.00	0.00	0.00	0.80	1.05

VCV matrix



	a	b	c	c	d
a	1.000	0.857	0.476	0.000	0.000
b	0.857	1.000	0.476	0.000	0.000
c	0.476	0.476	1.000	0.000	0.000
c	0.000	0.000	0.000	1.000	0.762
d	0.000	0.000	0.000	0.762	1.000

correlation matrix

Back to PGLS

$$\epsilon_i^{GLS} \sim MVN(0, \sigma^2 \mathbf{C})$$

$$\sigma^2 \mathbf{C} = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1i} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2i} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{i1} & \rho_{i2} & \rho_{i3} & \dots & 1 \end{pmatrix}$$

correlation matrix

Let's look at the residuals

Run the pgls

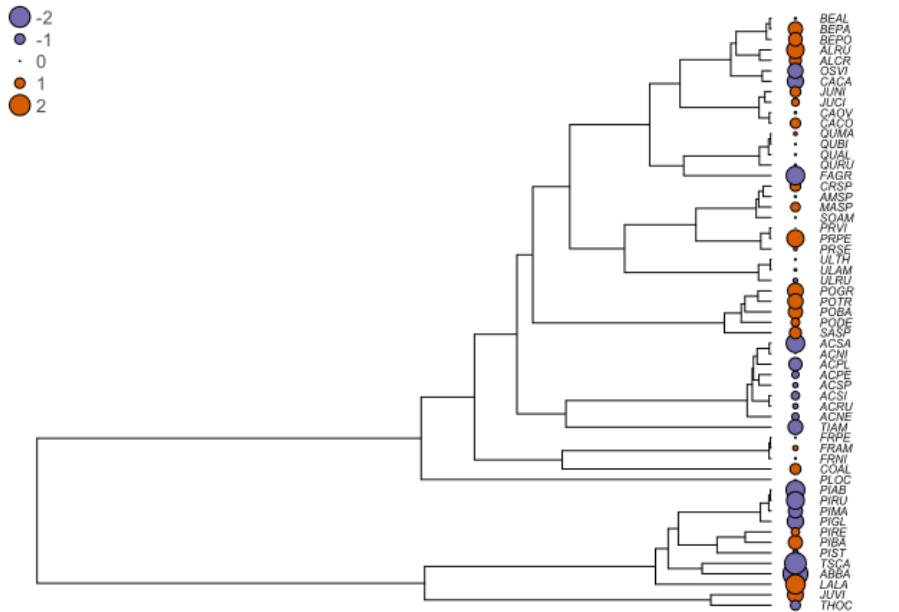
```
# Get the correlation structure
bm.corr <- corBrownian(phy=seedplantstree)
# PGLS
shade.pgls1b <- gls(Shade ~ Wd, data = seedplantsdata, correlation=bm.corr)
summary(shade.pgls1b)
```

Extract the residuals

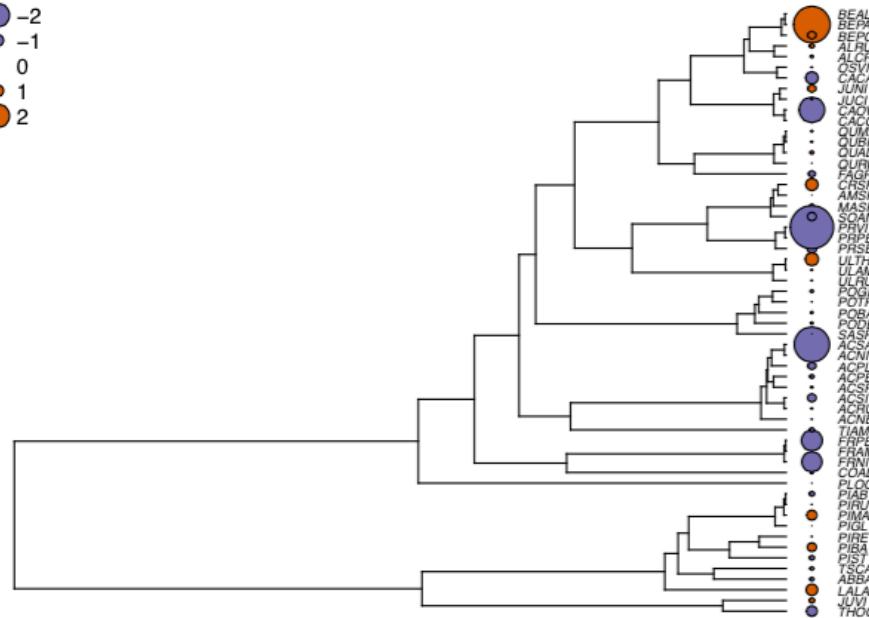
```
# Extract the residuals corrected by the correlation structure
pgls1.res <- residuals(shade.pgls1b, type="normalized")
```

Let's look at the residuals

Ordinary least squares



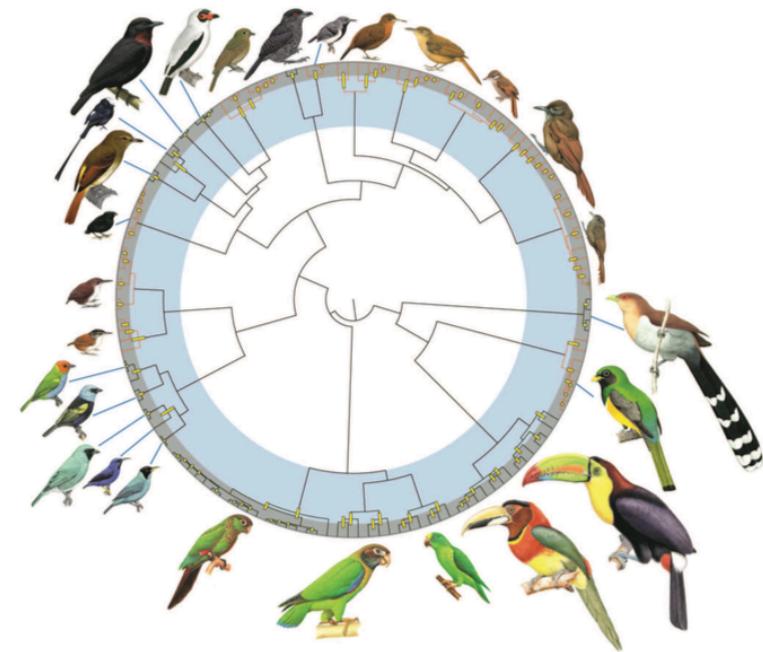
Generalized least squares



The residuals are not correlated anymore!

Example

- What drove bird diversification in the neotropics?
 - Lineage age
 - Foraging behaviour (stratum: understorey or canopy)
 - Ancestral area
 - Niche breadth
- Studied the diversity of 27 lineages of birds



Example

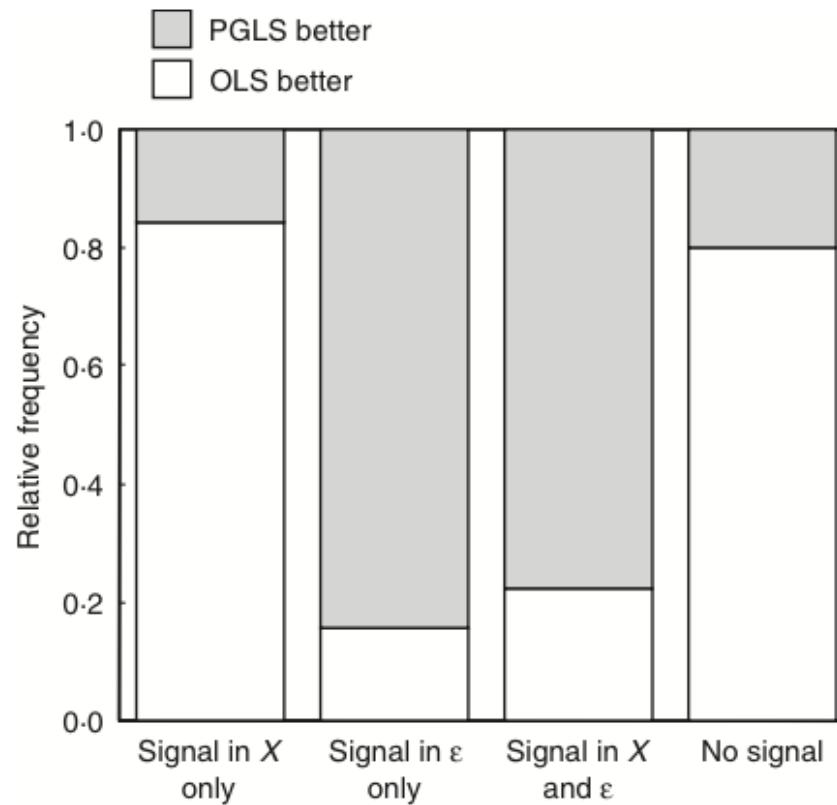
- What drove bird diversification in the neotropics?

Table 1 | Phylogenetic generalized least-squares regression showing the effects of historical and ecological variables on species diversity

Effect	Estimate	Standard error	t value	P	ΔAIC_c
Lineage age	0.1187	0.0283	4.1907	0.0004	6.9586
Foraging stratum	0.5188	0.2025	2.5623	0.0178	4.0122
Ancestral origin	-0.1921	0.2023	-0.9495	0.3527	-1.9546
Niche breadth	1.0097	1.0658	0.9473	0.3538	-1.9595

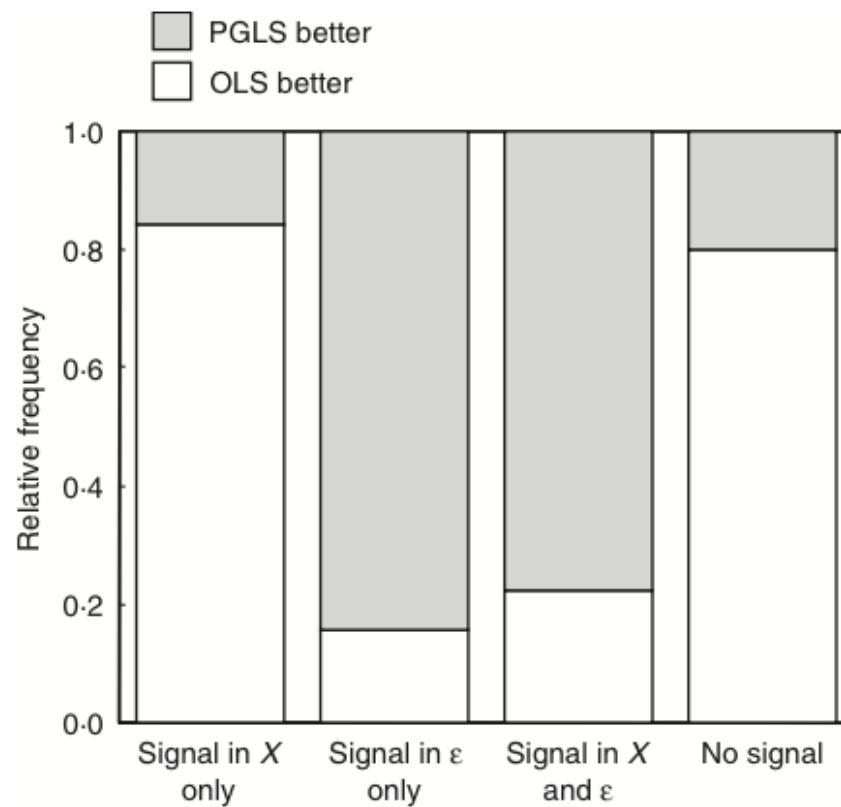
Why use PGLS ?

- High type I error
- Poor parameter estimates



But ...

- OLS is better than PGLS when there are no phylogenetic signal in the residuals.

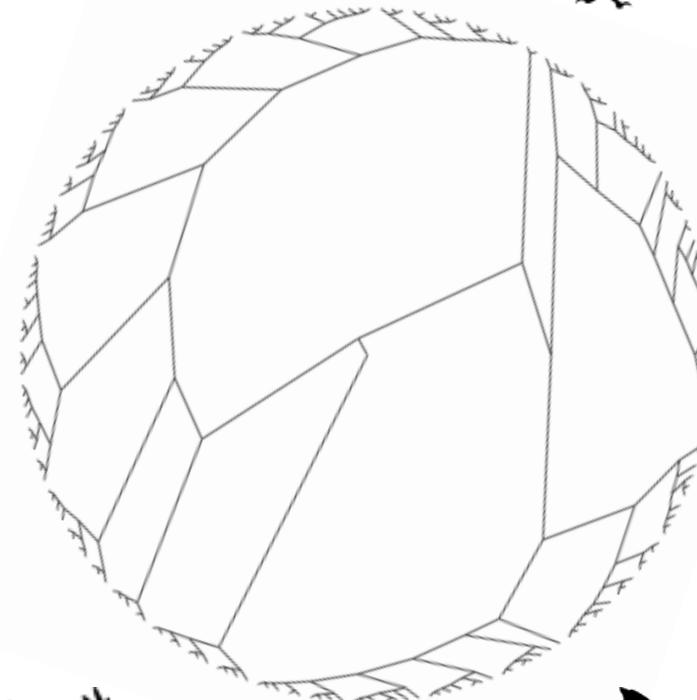


The solution

- Estimate a parameter λ in the GLS that determines the strength of the phylogenetic signal in the residuals.
 - $\lambda = 0$: no phylogenetic signal (equivalent to OLS)
 - $\lambda = 1$: perfect phylogenetic signal

$$\sigma^2 \mathbf{C} = \sigma^2 \begin{pmatrix} 1 & \lambda\rho_{12} & \lambda\rho_{13} & \dots & \lambda\rho_{1i} \\ \lambda\rho_{21} & 1 & \lambda\rho_{23} & \dots & \lambda\rho_{2i} \\ \lambda\rho_{31} & \lambda\rho_{32} & 1 & \dots & \lambda\rho_{3i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda\rho_{i1} & \lambda\rho_{i2} & \lambda\rho_{i3} & \dots & 1 \end{pmatrix}$$

Any questions?



Class material

Contains

- This presentation
- Tutorial containing practical exercises in R on PGLS
- Links to other resources

[http://github.com/simjoly/CourseComparativeMethods/tree/master/
WorkshopPGLS_halfday](http://github.com/simjoly/CourseComparativeMethods/tree/master/WorkshopPGLS_halfday)