

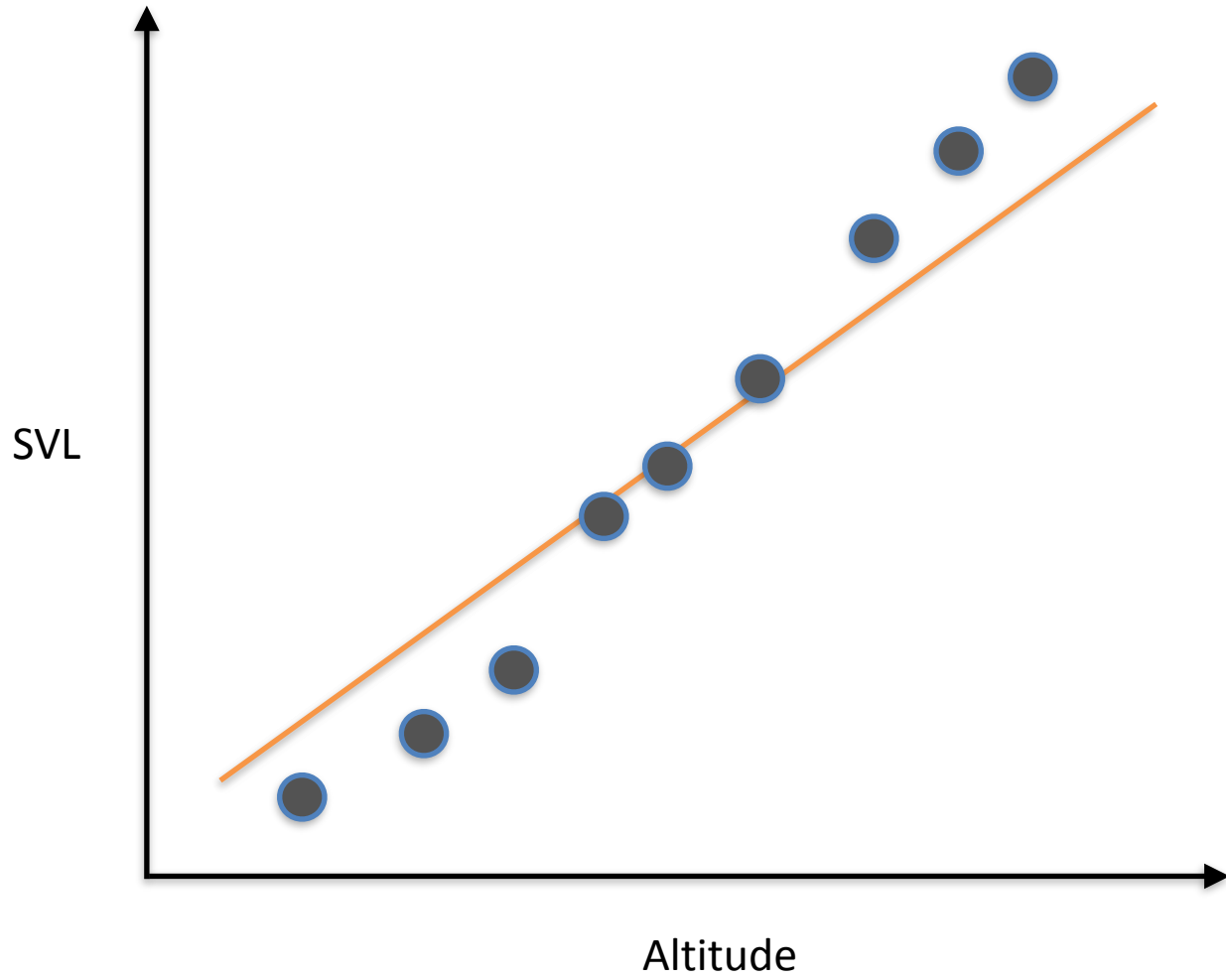
Latinamerican Macroevolution workshop

UNAM 2018



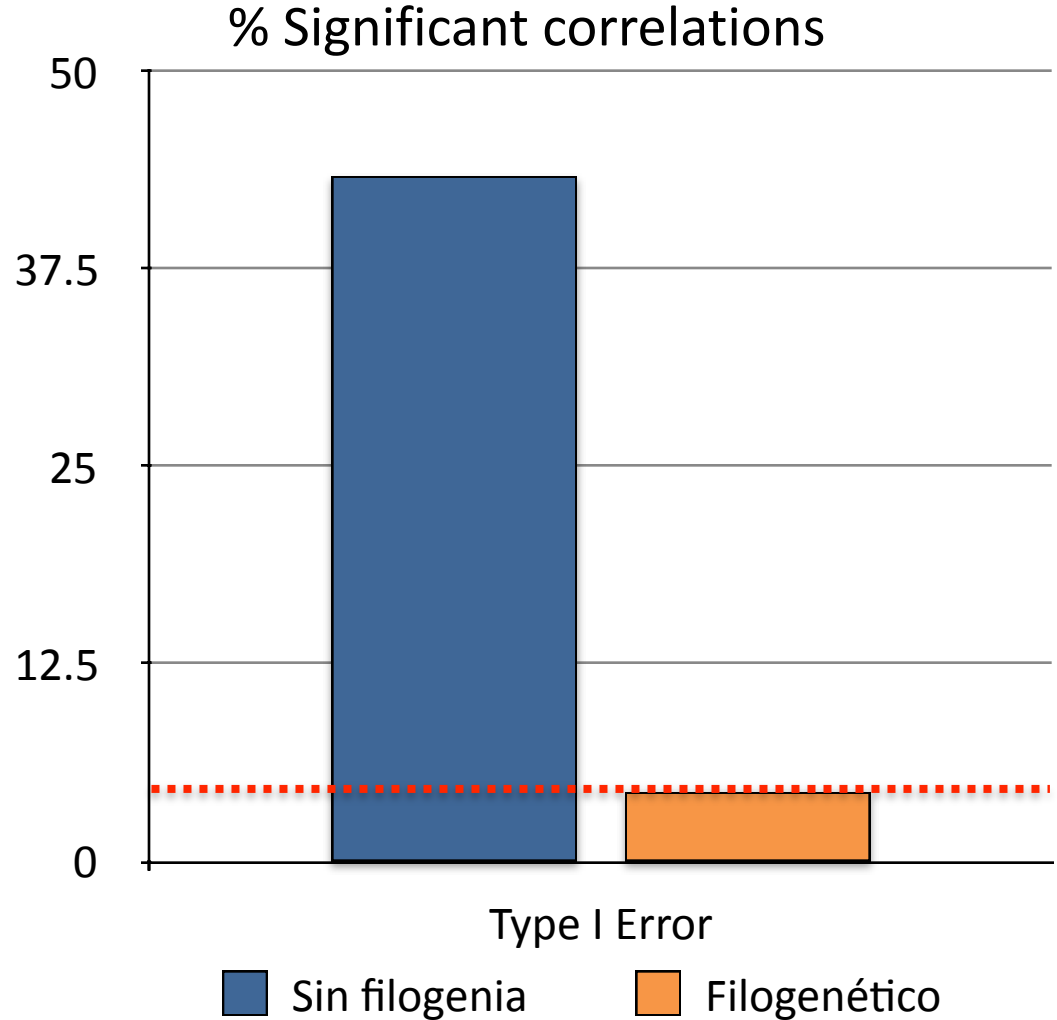
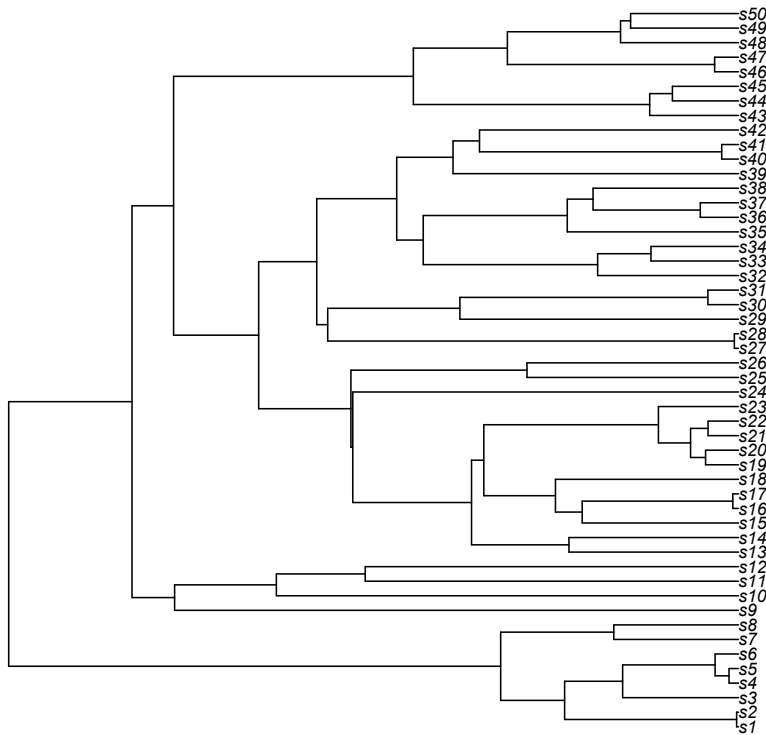
Alejandro Gonzalez-Voyer
Instituto de Ecología
UNAM





Risks of ignoring phylogenetic non-independence

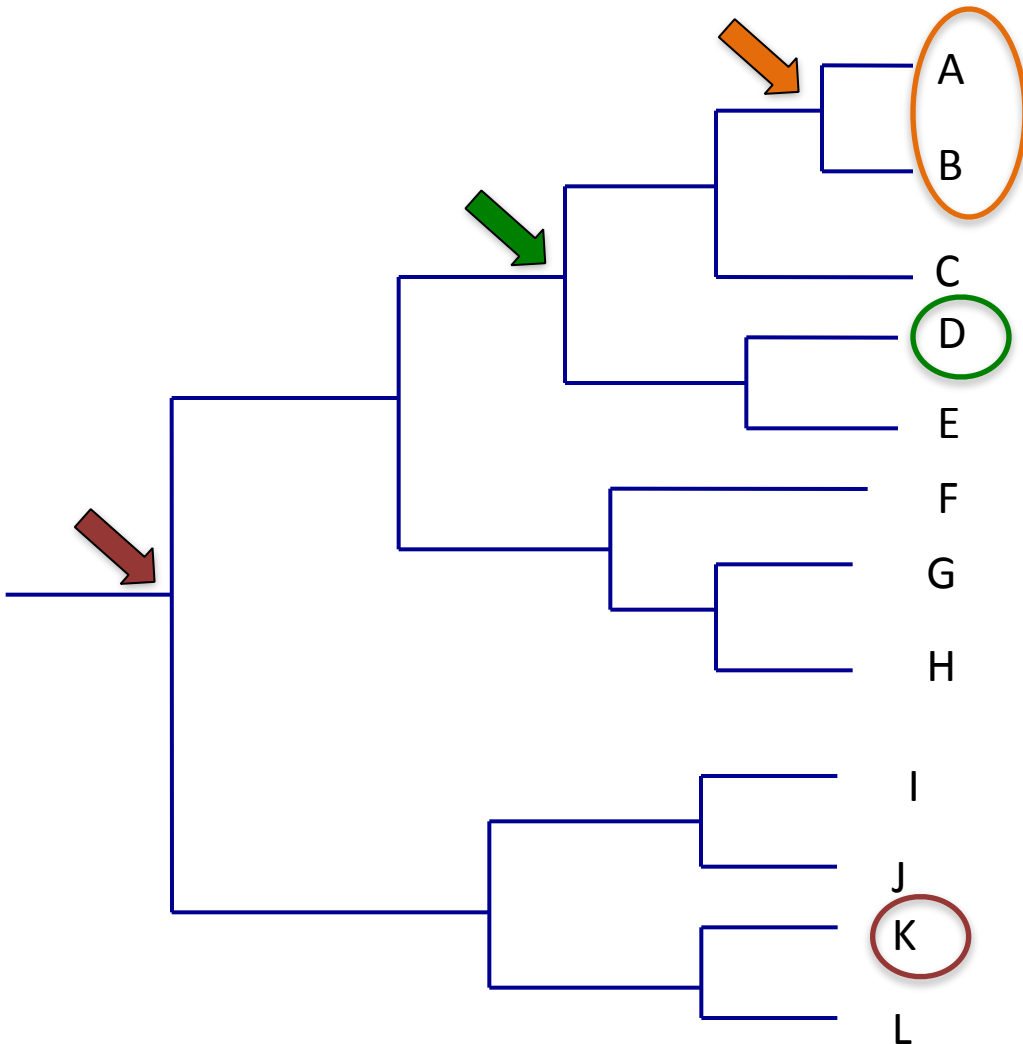
500 simulations of phenotypic evolution for 2 independent traits along the phylogeny



Phylogenetic Generalized Least Squares (PGLS)

- Generalized least squares model where the phylogenetic relationships among species are incorporated in the error term
- Variance covariance matrix describes the expected similarity among species based on the degree of shared ancestry
- More precisely: expected covariance among residuals due to shared ancestry (phylogenetic relationships) among species

Covariance and phylogenetic relationships



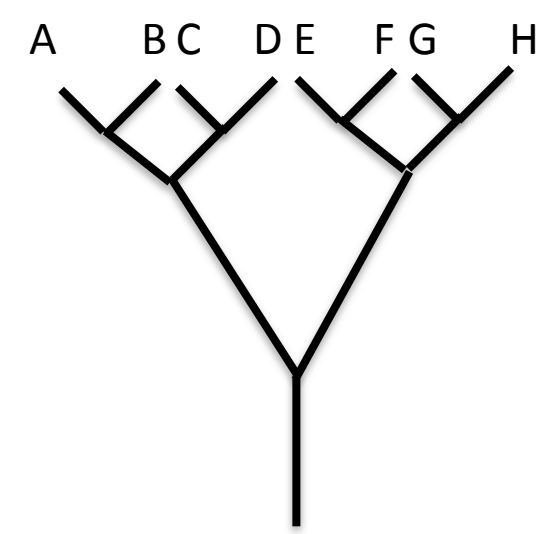
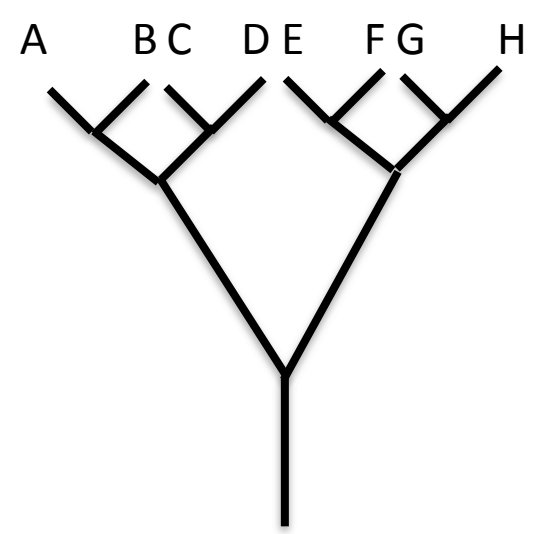
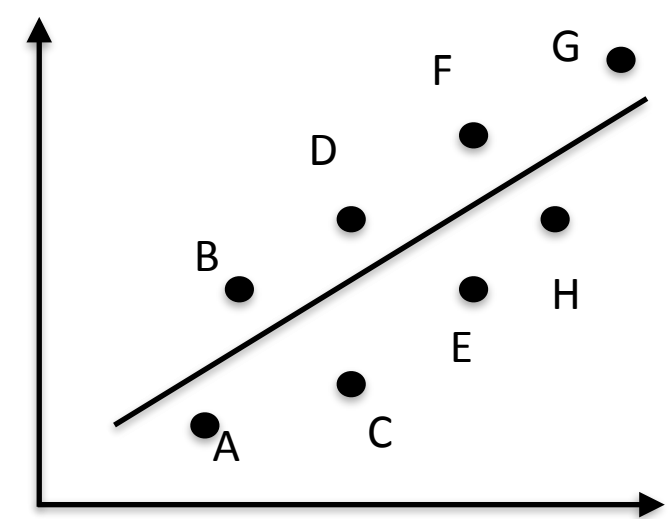
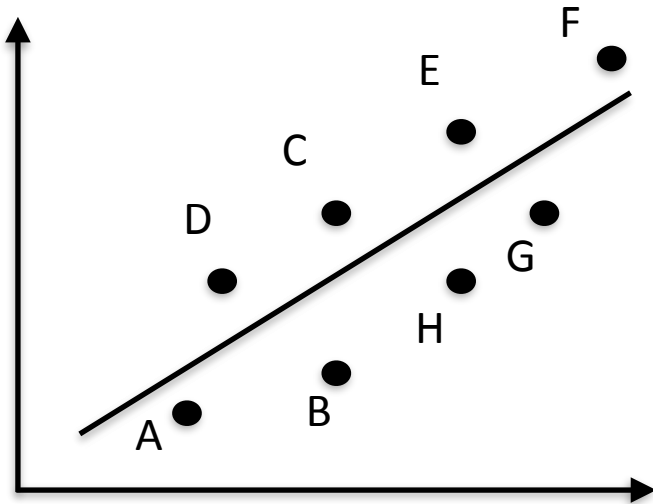
A & B: Expected covariance = **high**

A & D: Expected covariance = **low**

A & K: Expected covariance = **null**

Variance covariance matrix

	A	B	C	D	E	F	G	H	I	J	K	L
A	δ^2	+	+	+	+	-	-	-	0	0	0	0
B		δ^2	+	+	+	-	-	-	0	0	0	0
C			δ^2	+	+	-	-	-	0	0	0	0
D				δ^2	+	-	-	-	0	0	0	0
E					δ^2	-	-	-	0	0	0	0
F						δ^2	+	+	0	0	0	0
G							δ^2	+	0	0	0	0
H								δ^2	0	0	0	0
I									δ^2	+	+	+
J										δ^2	+	+
K											δ^2	+
L												δ^2



- Advantages of PGLS:
 - Model of trait evolution is determined by the variance covariance matrix, hence can accommodate different evolutionary models
 - Can incorporate an estimate of intraspecific measurement error in the model
 - Can combine continuous and discrete traits simply in a single model
 - Can obtain the intercept value

Phylogenetic generalized least squares

- Use of PGLS does **not** “remove” part of the correlation due to the phylogeny
- The effect of taking phylogeny into account is to reduce the variance around the estimated slope
- Their means are not because since estimates of regression coefficients are unbiased whether or not the correct phylogeny is taken into account.

Models of Evolution

- PGLS method estimates an evolutionary parameter simultaneously with model fit
- Evolutionary parameter (λ , α) describes the model of evolution & adjusts the variance covariance matrix to the model
 - λ : adjusts the variance covariance matrix to a Brownian evolution model
 - 1: evolution follows a Brownian model
 - 0: evolution does not follow a Brownian model, there is no phylogenetic signal in the residuals

PIC vs PGLS

- If covariance in the residuals follows a strict Brownian model PIC & PGLS should converge in the same result
- Regressions through the origin of PICs are identical to those obtained using PGLS with an intercept included
- **However** when evolution is not Brownian PGLS outperforms PICs
- Maximum likelihood estimate of lambda allows for necessary correction for estimated covariance of residuals

The 7 deadly sins of comparative analyses

1. Putting undue faith in models with low R^2
 - Only a single package in R estimates R^2 for phylogenetic GLS models
 - Some question the accuracy of R^2 estimates in phylogenetic GLS

2. Reporting the results of phylogenetically independent and phylogenetic analyses:
 - the two models make different assumptions about the distribution of data
 - at best regarded as alternative models of the same data
 - as such they should NOT be treated equally
3. Not testing distributional assumptions:
 - Phylogenetic analyses make assumptions about the distribution of residuals that are the same as those made in nonphylogenetic analyses
 - R offers different ways of testing residual distributions (in caper simply use plot() function)

4. Data dredging:

- because data are not generated experimentally, the relationships found are only correlative
- possibility of hidden variables mean that significant correlations may be mistakenly taken as implying a causal relationship when no such relationship exists
- should avoid the temptation of including a large number of predictors to “see what comes out”

5. Treating residuals as data:
 - Use of residuals to control for confounding variables is wrong
 - Can result in biases
 - Use ANCOVA to control for allometric effects or confounding variables
6. Ignoring alternative models:
 - Use model comparison with AIC or AIC_c
7. Ignoring quality control of data:
 - Data from disparate sources may be of different quality
 - Low quality data compromises statistical power
 - Missing data can lead to biases in the outcome of analyses

Correlates of species richness

- Intrinsic or extrinsic species characteristics may influence speciation rates
- Can use PGLS models to analyze the correlation between species characteristics and species richness
- Dependent variable in such models is rate of diversification or species richness