Autocorrelation 3: Phylogenetic Autocorrelation

Michael Noonan

Biol 520C: Statistical modelling for biological data

Table of contents

Autocorrelation and ecology

The IID Assumption

Over the course of the last lecture two we saw how collecting data over time and space can result in autocorrelated data, breaking the IID assumption.

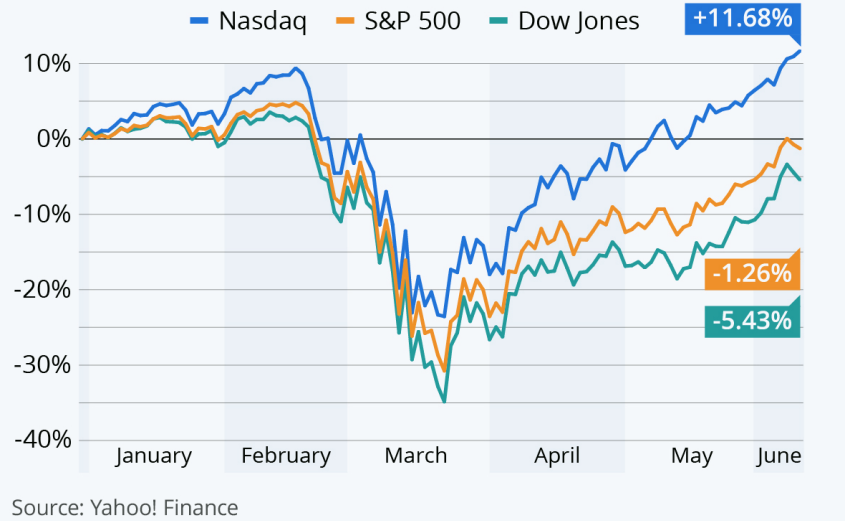
We also saw how modifying the off-diagonals of the correlation matrix can correct for autocorrelation.

Ecologists often find themselves collecting data repeatedly over space and time, but we're not the only field doing this.

Most of the methods used to correct for temporal or spatial autocorrelation come from other scientific fields.

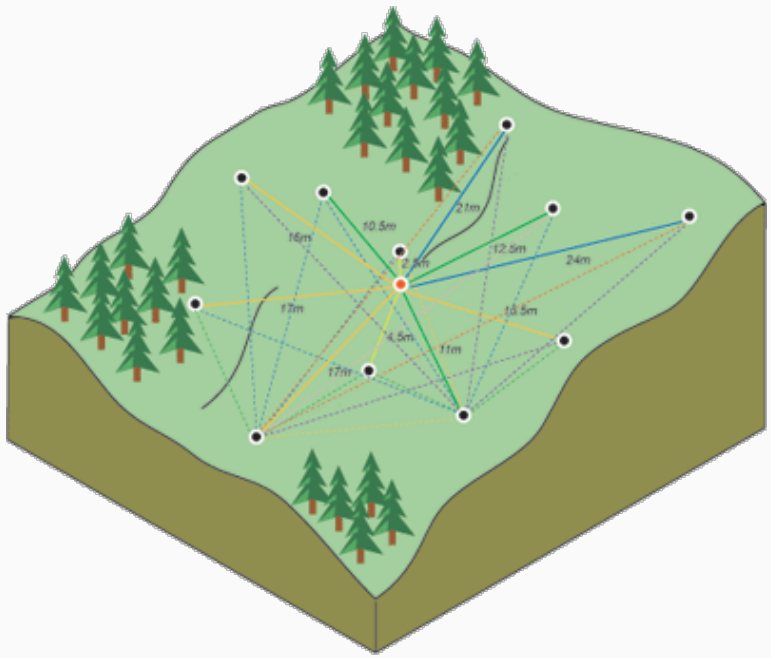
Time series analysis

Can you think of important data that are collected over time that would drive statistical developments? Data where the ability to predict the future might be profitable?



The models for correcting for temporal autocorrelation come from time series analysis and econometrics.

Can you think of important data that are collected over space? Data where the ability to predict where things occur might be profitable?



Source: ArcMap

The models for correcting for spatial autocorrelation come from geostats. with the goal of mapping mineral deposits from only a few boreholes.

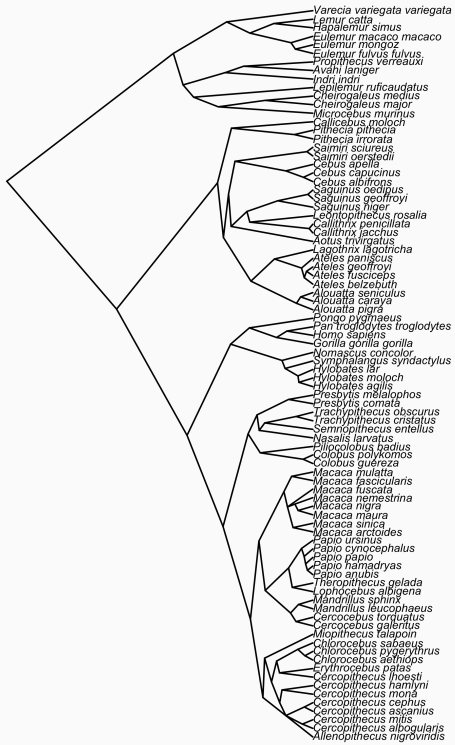
Phylogenetic Inertia

... but as ecologists, we don't only measure things over space or time and there are other potential sources of autocorrelation.

Darwin (1859) recognised that species do not start over from scratch after speciation and that characteristics are built on existing ones that were inherited from their ancestors.

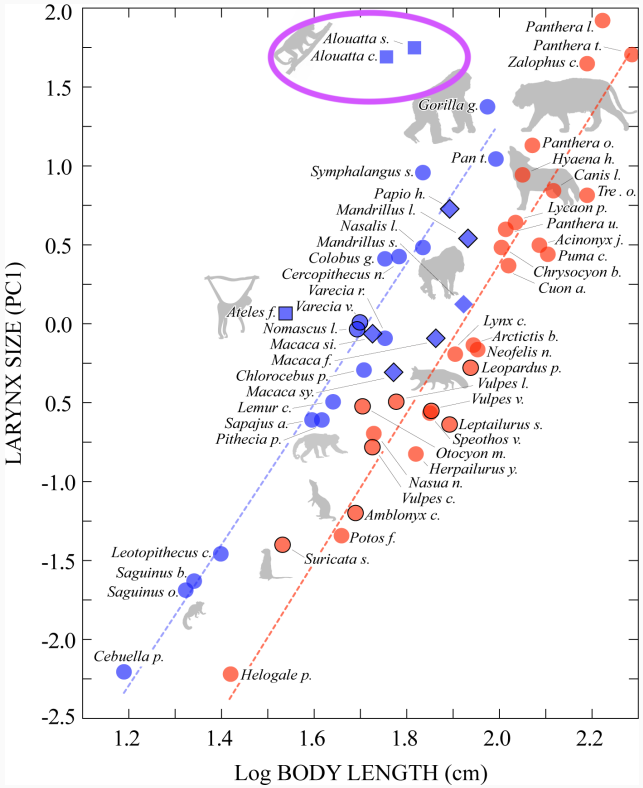
These inherited traits will limit the amount of evolution seen in the new taxa, especially if only a short amount of time has passed (evolutionarily speaking).

This phenomenon is referred to as **Phylogenetic inertia**.

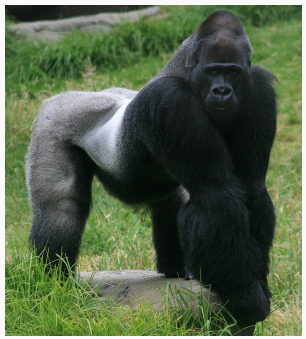


Phylogenetic Inertia cont.

Clusters of data from closely related species can have disproportionate effects and pull regressions lines in their direction if not accounted for.



(Bowling *et al.*, 2020)



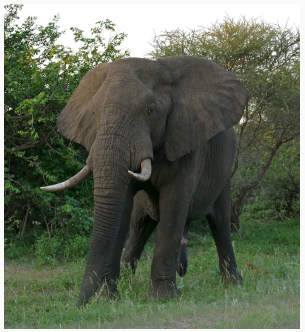
Gorilla gorilla, source: Wikipedia

Elephas maximus indicus, source: Pixabay

We probably have new information.

Elephas maximus sumatrensis, source: Wikimedia

Do we really have completely new information?

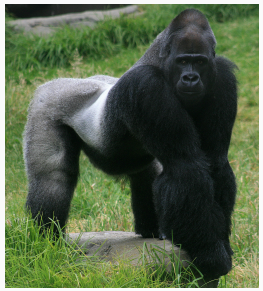


Loxodonta africana, source: Wikipedia

What about now?

Impact of phylogenetic inertia

The effects of phylogenetic inertia can be even more extreme if we have multiple observations per species.



N = 2

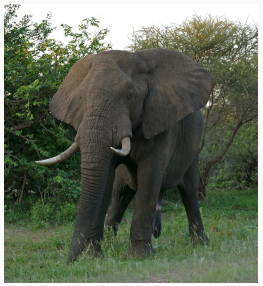
Gorilla gorilla, source: Wikipedia

N= 5

Elephas maximus indicus, source: Pixabay

N= 4

Elephas maximus sumatrensis, source: Wikimedia



N= 3

Loxodonta africana, source: Wikipedia

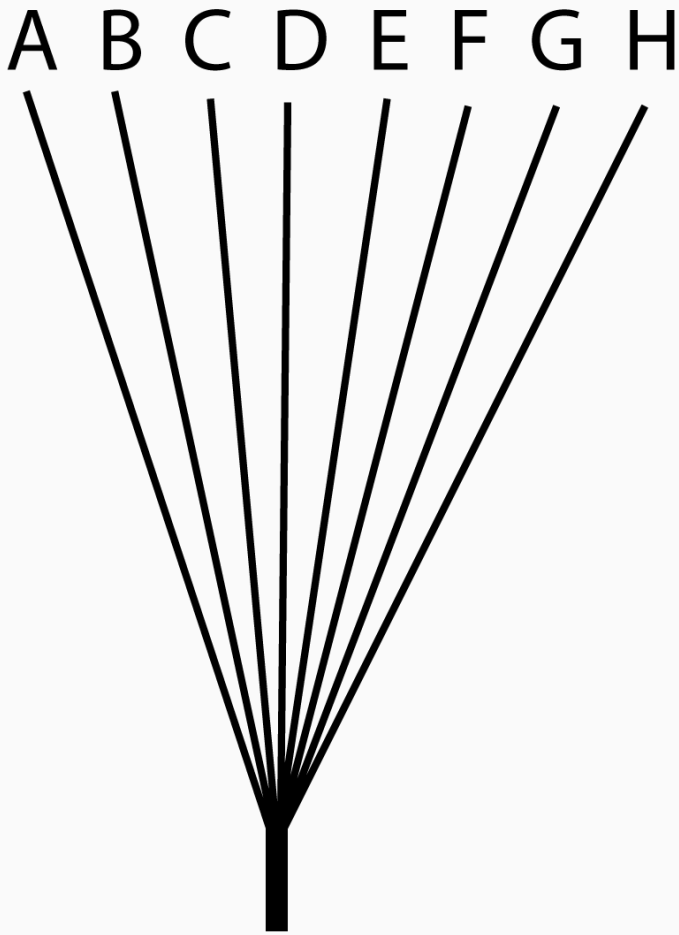
All the elephant data are likely to be similar because elephant species diverged from one another more recently than elephants did from gorillas.

The 2 gorilla data points cary useful information for making inter-specific comparisons, but would get out outweighed in a traditional regression.

Impact of phylogen. inertia cont.

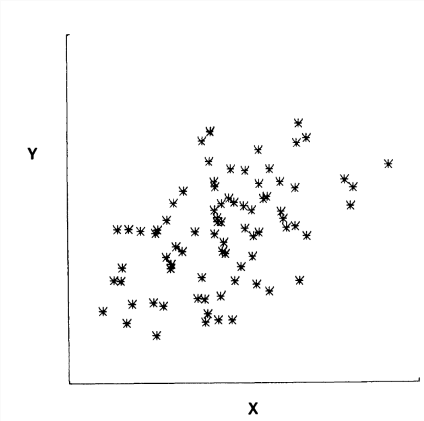
If we feed species data into a traditional, IID regression framework, we assuming evolution looks something like this.

Are we prepared to make that assumption?

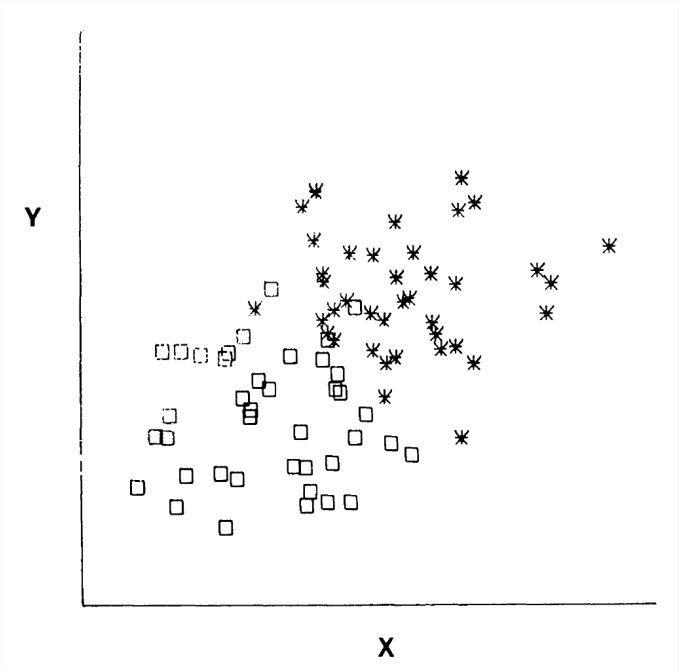


Impact of phylogen. inertia cont.[2]

As an extreme example, we could have data that looks something like:

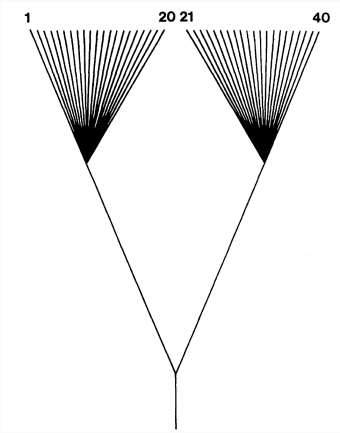


A regression could be highly misinformative.



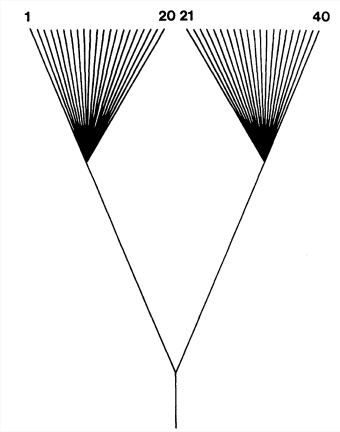
(Felsenstein, 1985)

But if we collected it from a tree that looks like this:

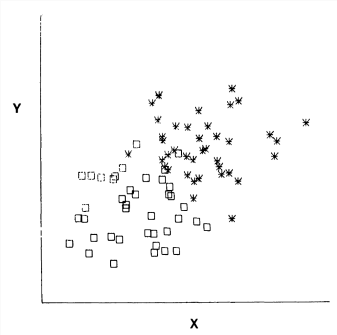


Phylogenetic correlation

If we knew we were only sampling from effectively 2 groups, we would probably not run a regression.



(Felsenstein, 1985)



(Felsenstein, 1985)

The challenge is that real phylogenies aren't this simple, and it's difficult to tailor our sampling around evolutionary histories.

Phylogenetic correlation

Problem:

"*... species are part of a hierarchically structured phylogeny, and thus cannot be regarded for statistical purposes as if drawn independently from the same distribution*" — Felsenstein (1985)

With spatially/temporally autocorrelated data, we could rely on advances in econometrics and geostatistics, but with phylogenetic correlation we were on our own.

Solution:

"*... if the techniques do not exist, then we must invent them.*"

— Hilborn & Mangel (1997)

Detecting Phylogenetic Autocorrelation

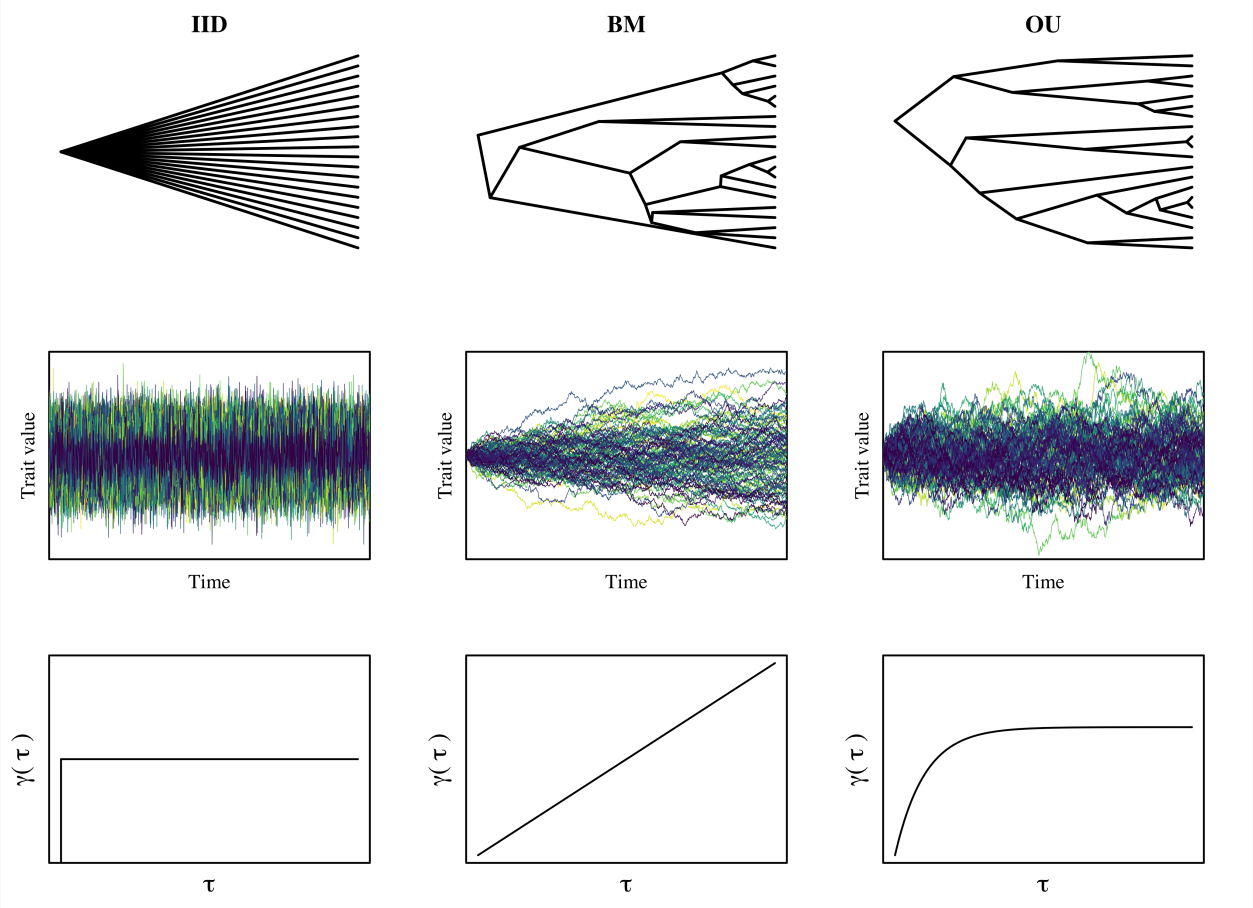
Detecting phylogenetic autocorrelation

For time series we had... the ACF.

For spatial data we had... variograms.

For phylogenetic data we have... nothing... yet.

Detecting phylogenetic autocorr. cont.



Source: Noonan et al. *under review*

The Comparative Method

The Comparative Methods

When we're carrying out these types of analyses, we're more interested in *between* species comparisons than *within* species.

Felsenstein's paper spurred a major ecological sub-discipline to form around finding a solution to this problem, termed 'Comparative Methods'.

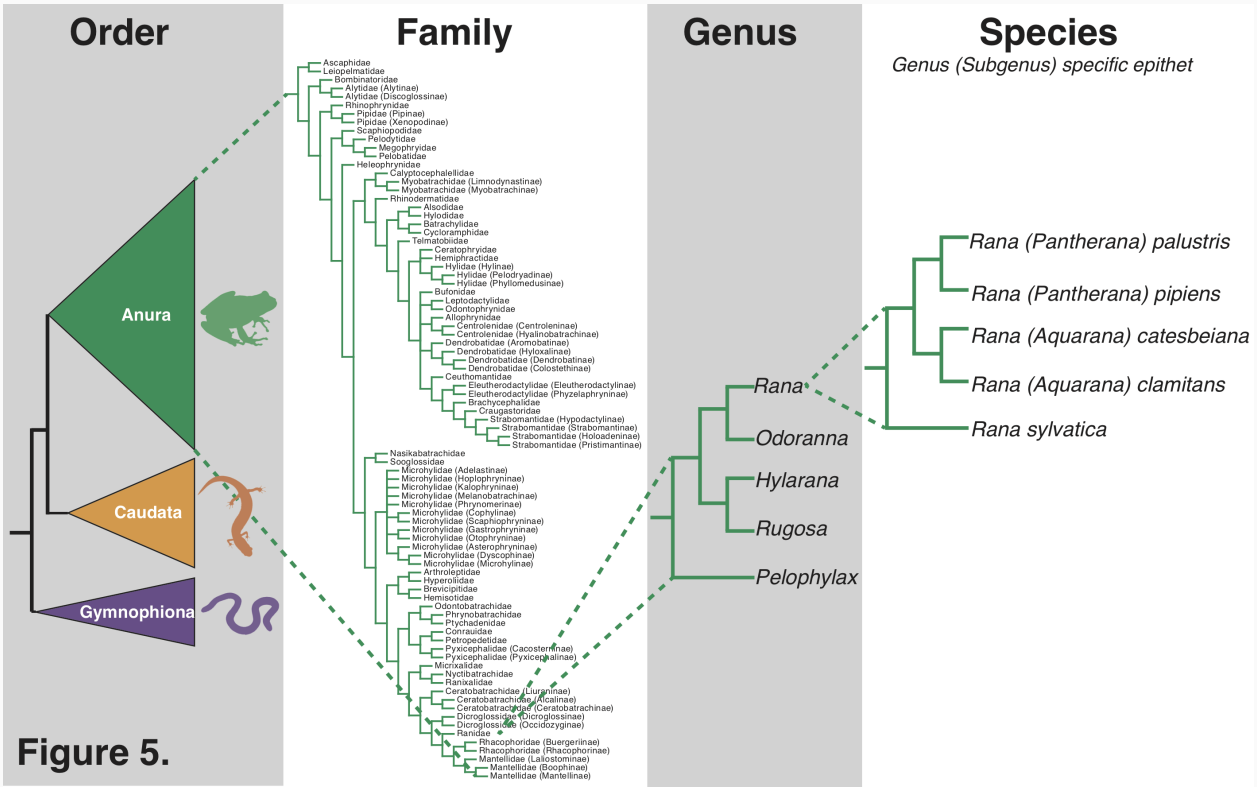
"*The non-independence can be circumvented in principle if adequate information on the phylogeny is available.*" — Felsenstein (1985)

In other words, if we can get information on how species are related, we can use this information to improve our models.

But how?

Phylogenetic Trees

The tree of life has a clear, nested ordering (Spp. in Genus, Genus in Family, etc.)



Source: amphibiaweb.org

One easy solution is to treat these data the same way you would any other nested data using random effects (and people do this).

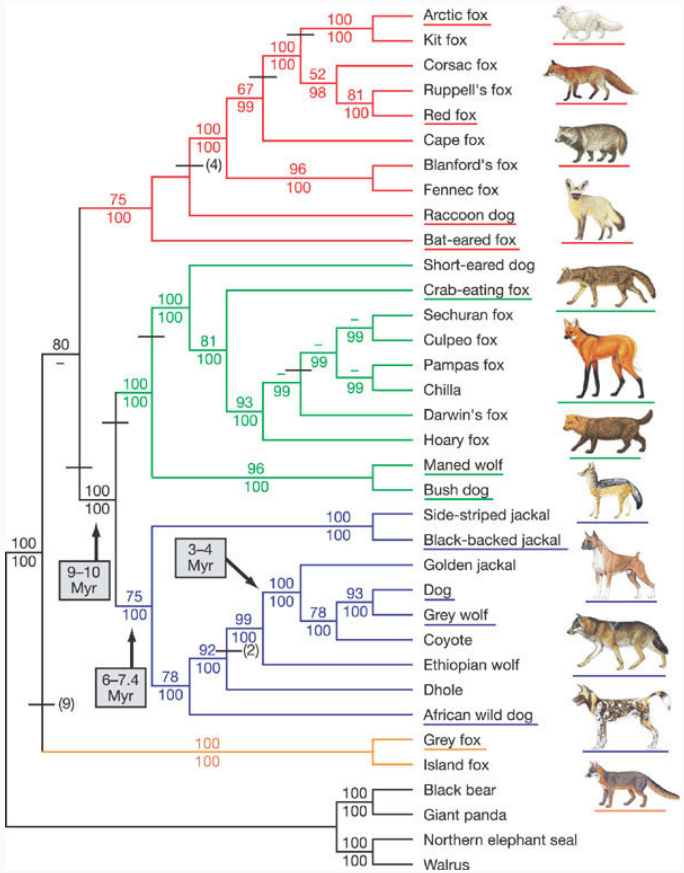
Phylogenetic Random Effects

This treats all species within a genus as independent.

But species within a genus will have diverged at different times... and genera within families, and so on...

In other words, this is only a half measure that largely ignores modern advances in phylogenetics.

Unsurprisingly, it's not particularly efficient (see Grafen, 1989, if you're interested in an in depth exploration of this).



Source: Lindblad-Toh *et al.* (2005)

Correcting Phylogenetic Autocorr.

Can you think of another solution?

In the 1980s, ecologists realised that if you use phylogenetic distances to modify the model's variance-covariance matrix you can get corrections that function much like those we used for spatial/temporal autocorrelation.

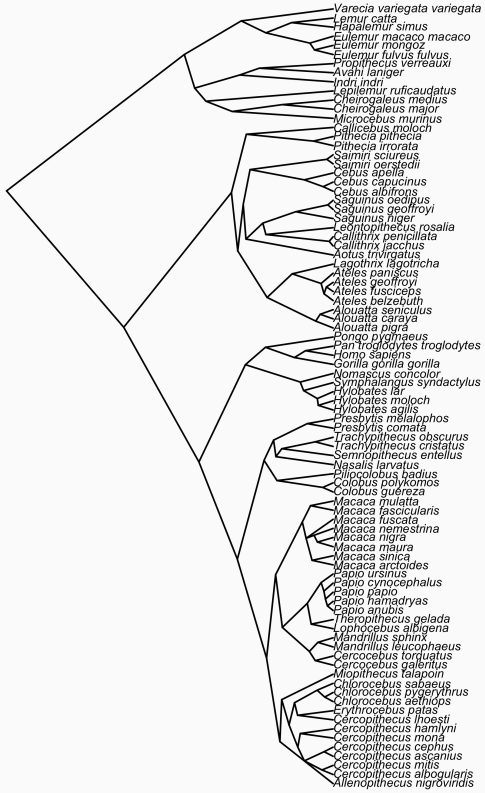
$$V = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

This is the basis of what we now call phylogenetic regression and it's what we're going to focus on for the rest of this lecture.

Phylogenetic Distances and Phylogenetic Regression

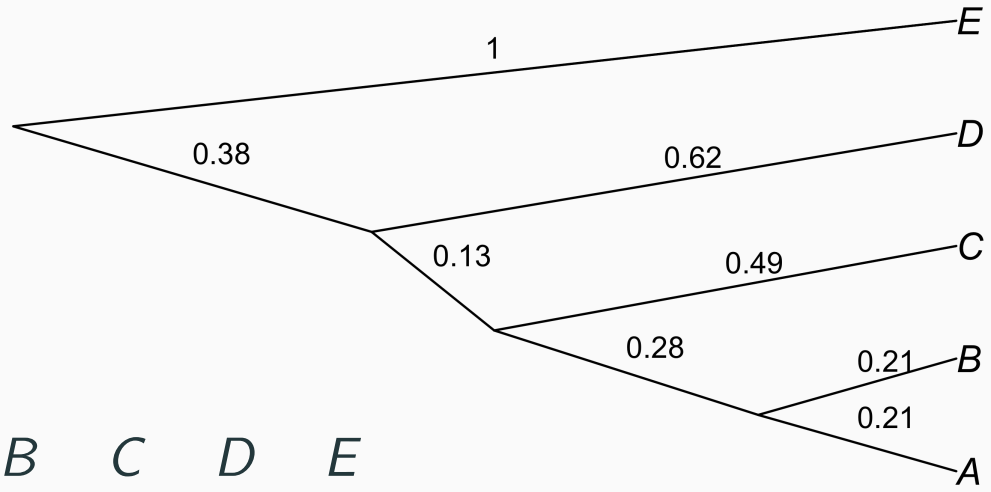
Phylogenetic regression

Our goal is to get from here, to here



$$V = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

Phylogenetic Correlation Matrix

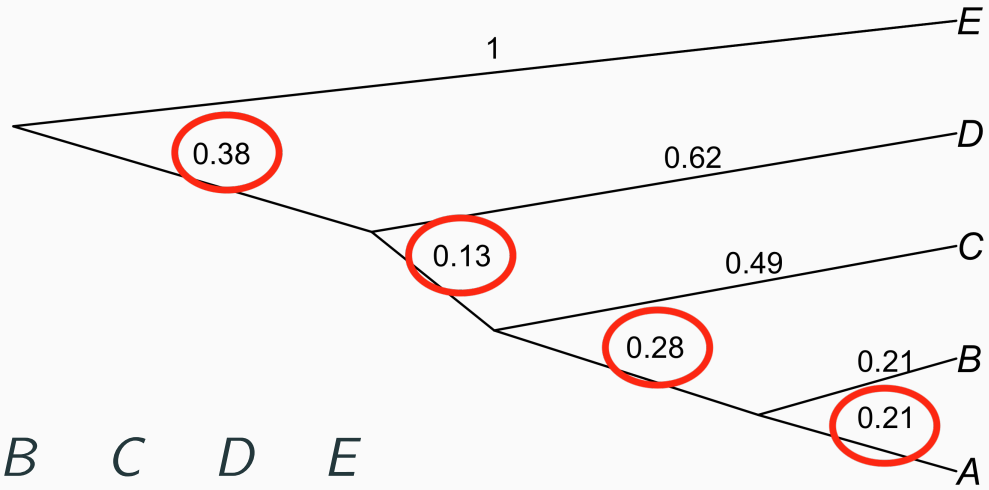


$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

First step are the diagonals.

How much of the tree does A share with A?

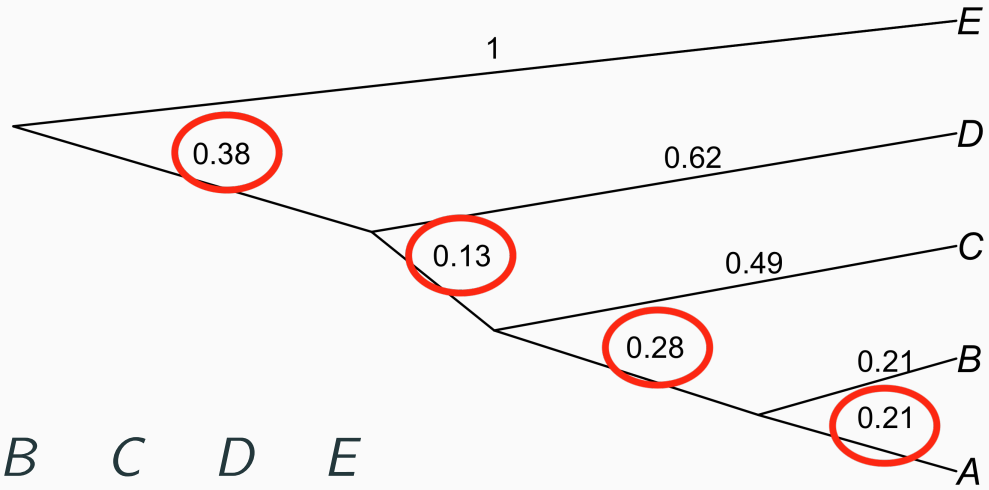
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

$$0.38 + 0.13 + 0.28 + 0.21 = 1$$

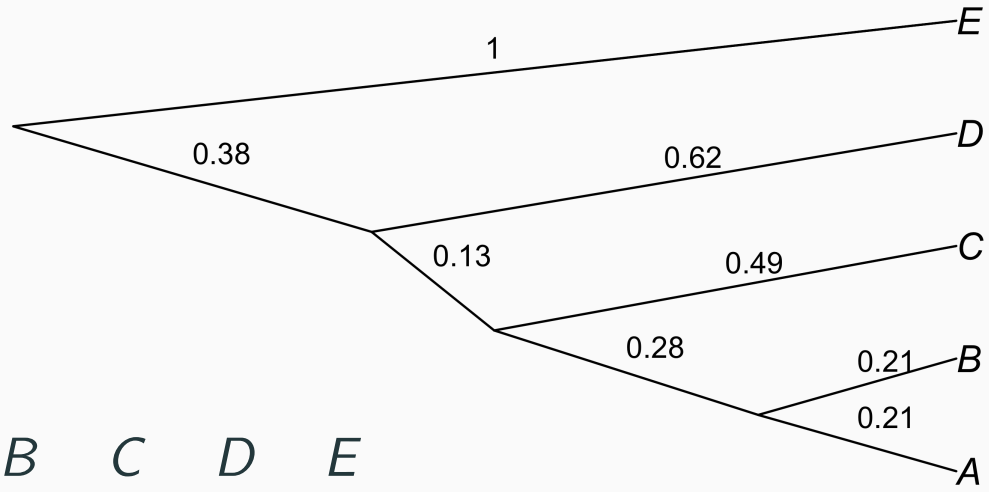
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

$$0.38 + 0.13 + 0.28 + 0.21 = 1$$

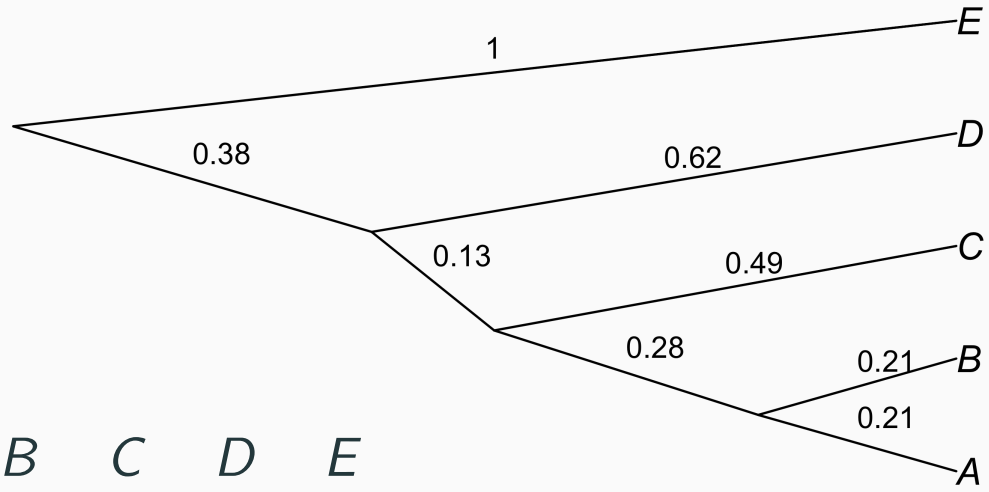
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

Repeat for all intra-specific comparisons.

Phylogenetic Correlation Matrix

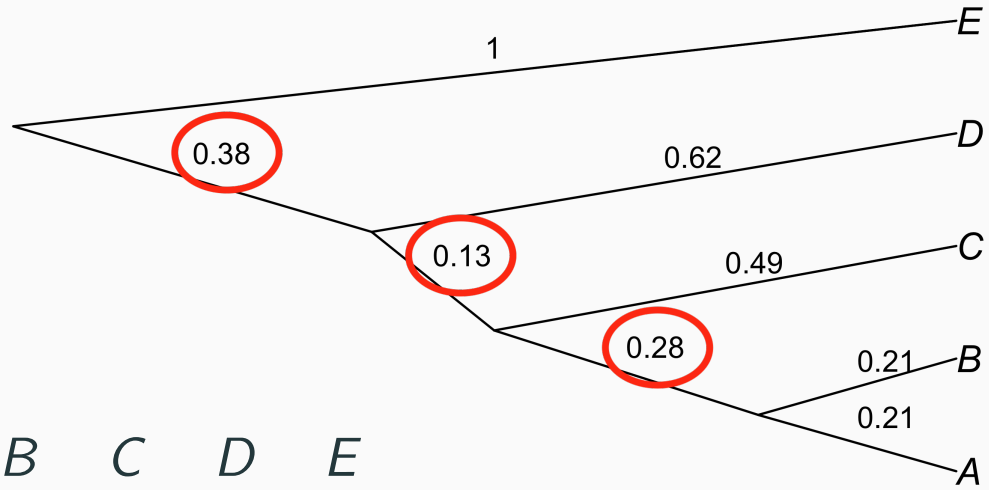


$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

Next are the off diagonals.

How much of the tree does A share with B?

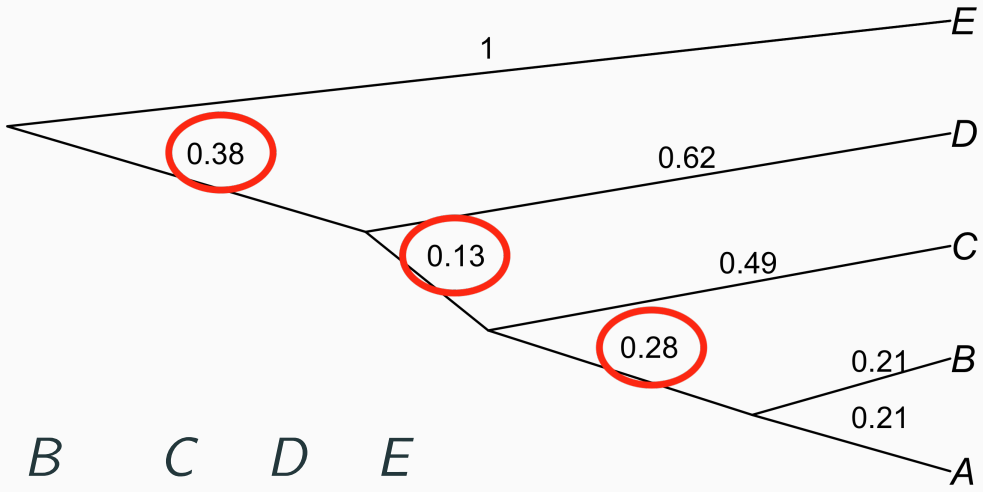
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

$$0.38 + 0.13 + 0.28 = 0.79$$

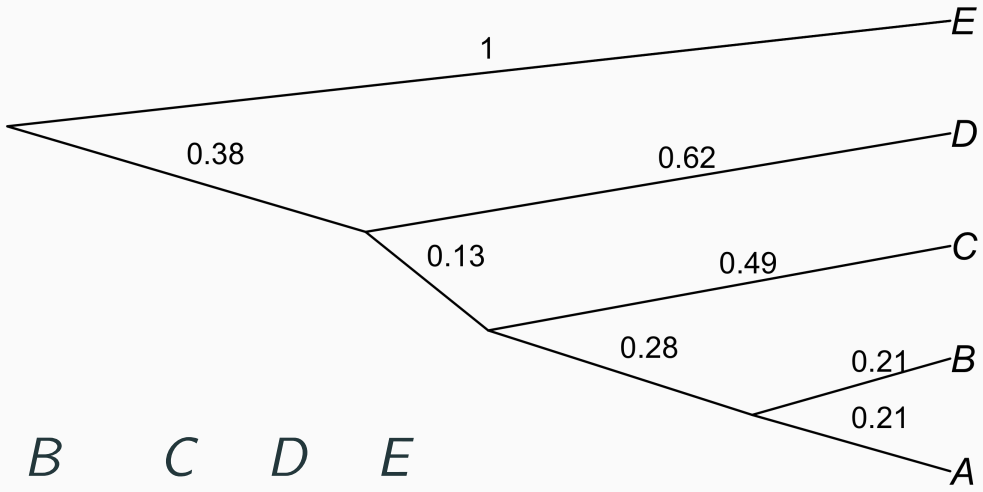
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0.79 & 0 & 0 & 0 \\ 0.79 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

$$0.38 + 0.13 + 0.28 = 0.79$$

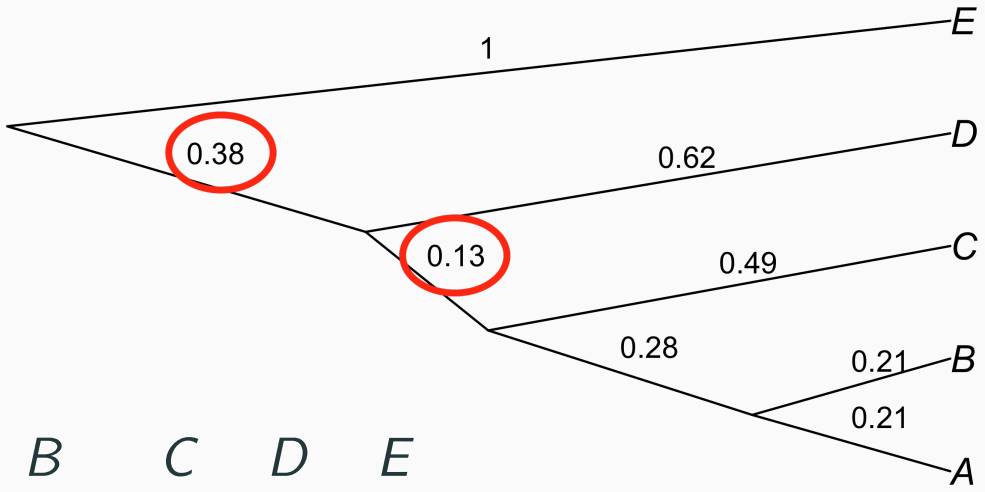
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0.79 & 0 & 0 & 0 \\ 0.79 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

Repeat for A & C

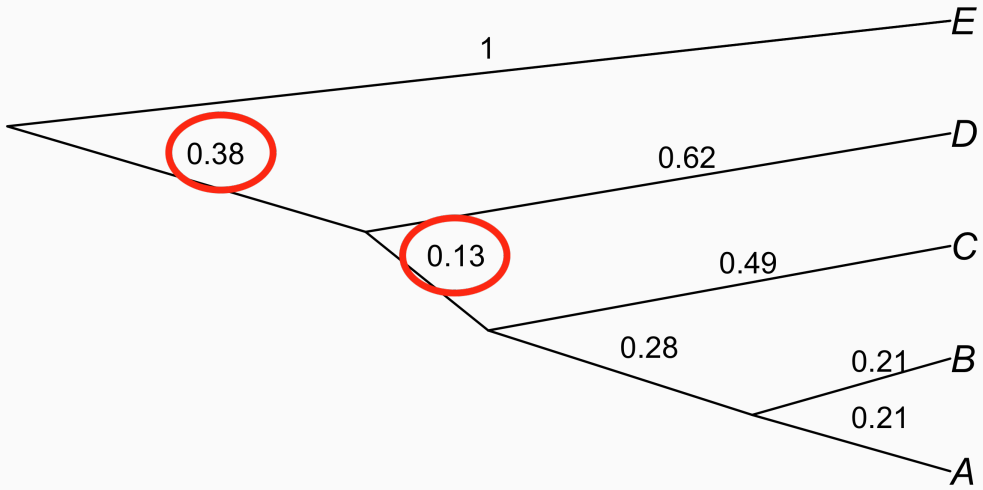
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0.79 & 0 & 0 & 0 \\ 0.79 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

$$0.38 + 0.13 = 0.51$$

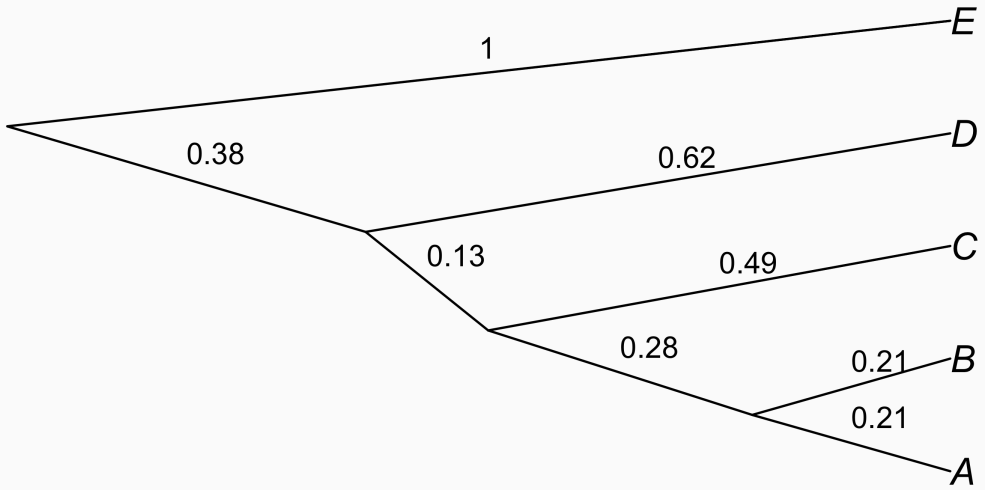
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\left[\begin{array}{ccccc}
1 & 0.79 & 0.51 & 0 & 0 \\
0.79 & 1 & 0 & 0 & 0 \\
0.51 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
$$

$$0.38 + 0.13 = 0.51$$

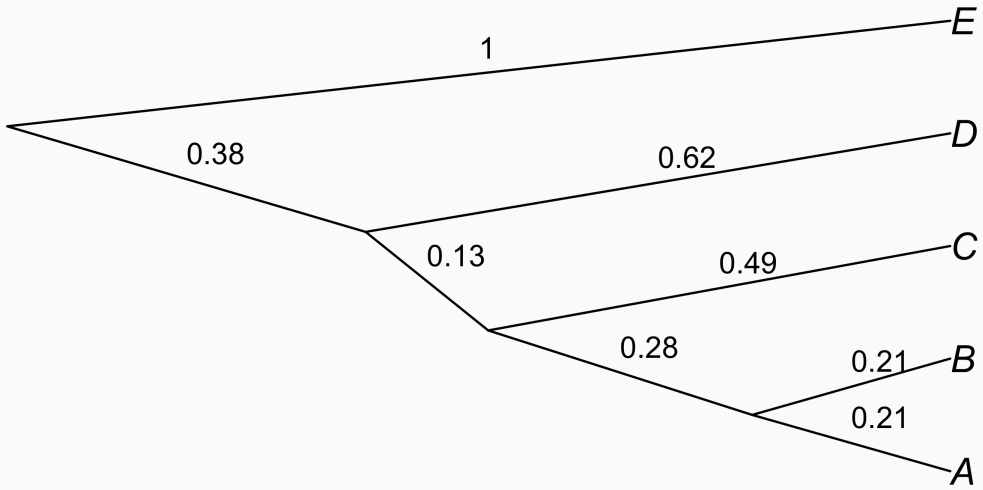
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\begin{bmatrix} 1 & 0.79 & 0.51 & 0.38 & 0 \\ 0.79 & 1 & 0 & 0 & 0 \\ 0.51 & 0 & 1 & 0 & 0 \\ 0.38 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

... and so on.

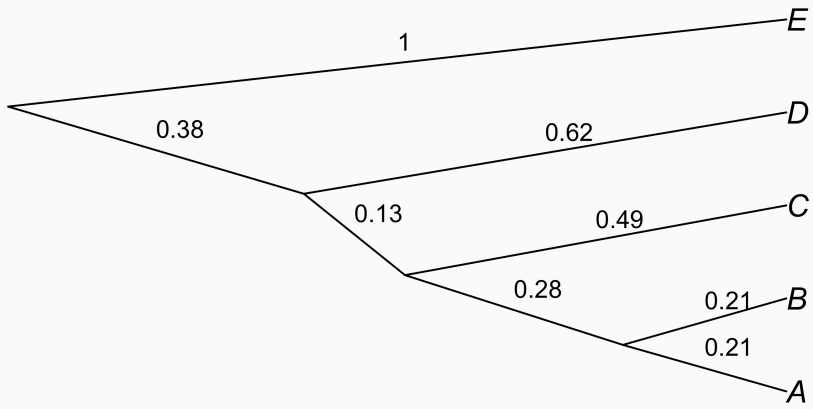
Phylogenetic Correlation Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & D & E \end{array} \\
\left[\begin{array}{ccccc}
1 & 0.79 & 0.51 & 0.38 & 0 \\
0.79 & 1 & 0.51 & 0.38 & 0 \\
0.51 & 0.51 & 1 & 0.38 & 0 \\
0.38 & 0.38 & 0.38 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
$$

Repeat for all species in the tree.

Phylogenetic Correlation Matrix in R



We'll use the R package ape

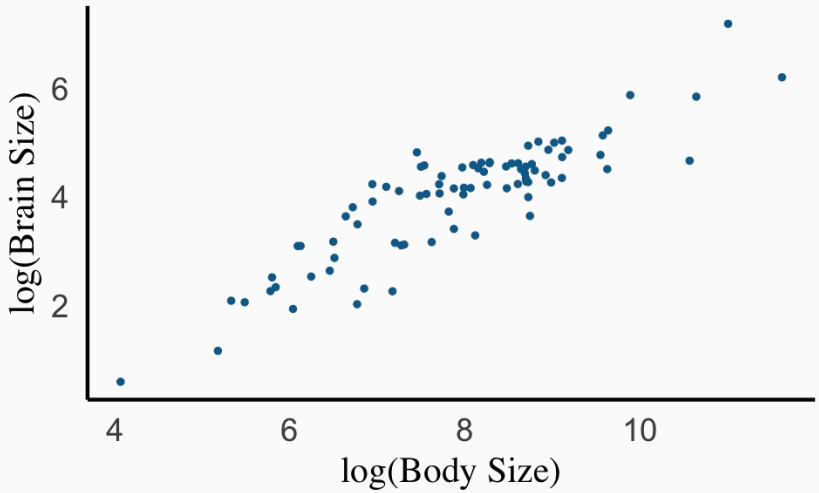
```
library(ape)

#Create our example tree
txt <- "((((A:0.21,B:0.21):0.28,C:0.49):0.13,D:0.62):0.38,E:1.00);"

tree.examp <- read.tree(text = txt)


vcv(tree.examp)

     A    B    C    D E
A 1.00 0.79 0.51 0.38 0
B 0.79 1.00 0.51 0.38 0
C 0.51 0.51 1.00 0.38 0
D 0.38 0.38 0.38 1.00 0
E 0.00 0.00 0.00 0.00 1
```

With a phylogenetic correlation matrix in hand, all that's left to do is feed this into the model just as we did with temporal or spatial autocorrelation.

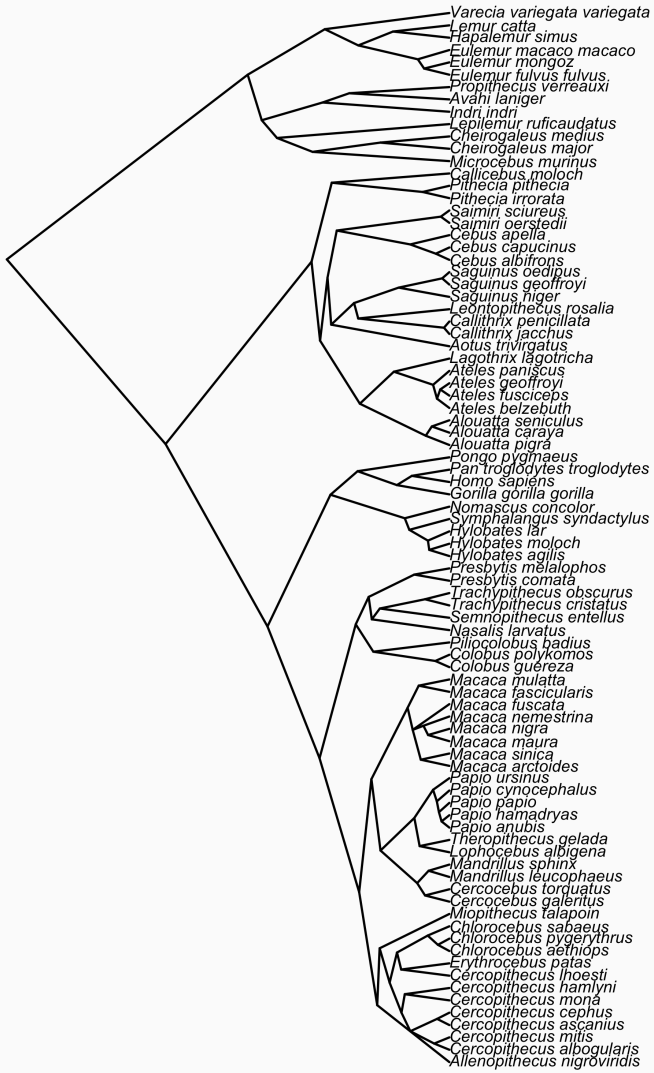
Phylogenetic Regression in R

Primate trait data

We'll work with a dataset of primate brain sizes and body sizes obtained from AnthroTree Workshop and Duke University.



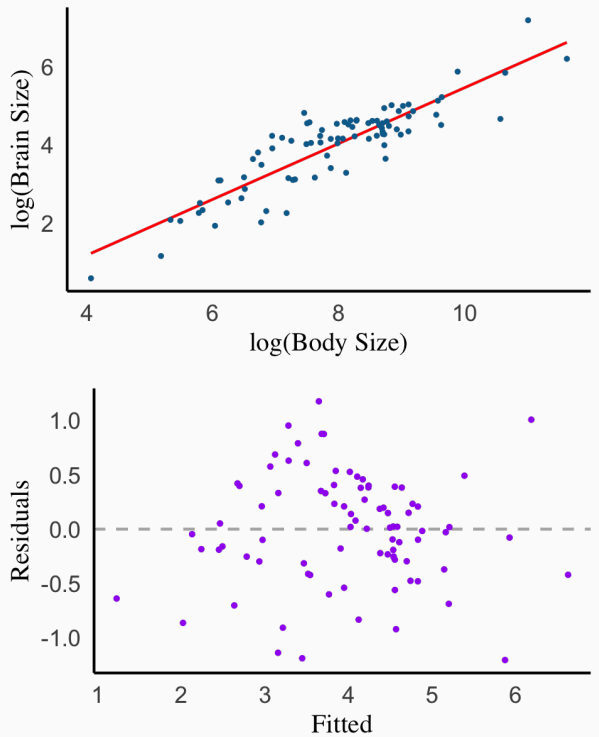
And a phylogenetic tree derived from Arnold *et al.* (2010).



Primate regression

We're interested in knowing what the relationship between brain size and body size is.

```
library(nlme)
library(ape)
data <- read.csv("brain_body_spec.csv")

FIT <- gls(lg_brain ~ lg_body, data = data,
           method = "ML")

summary(FIT)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC      BIC    logLik
  132.6719 140.035 -63.33597

Coefficients:
                 Value Std.Error   t-value p-value
(Intercept) -1.6860136 0.3310534 -5.092874       0
lg_body      0.7154667 0.0411488 17.387323       0
```

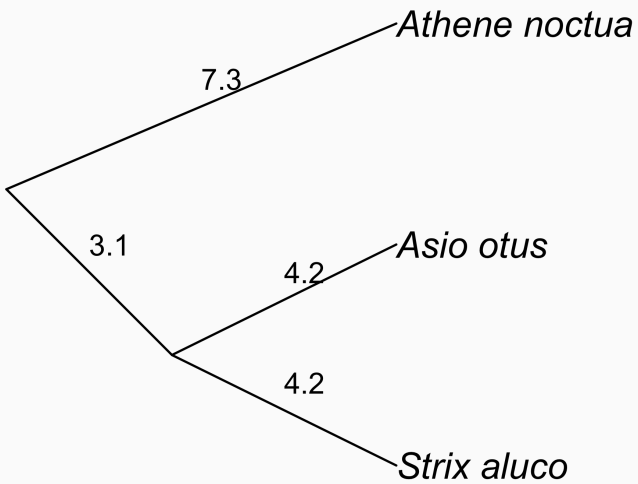


Fit and residuals look perfectly fine on the surface.

Primate phylogeny

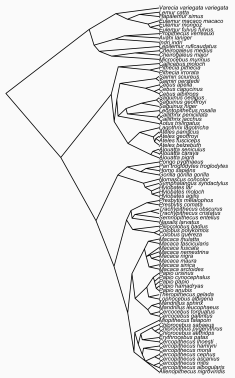
If you're unlucky, there will not be an existing phylogeny to work with and you will have to build it by hand.

```
TREE <- "((Strix_aluco:4.2,Asio_otus:4.2):3.1,Athene_
      noctua:7.3);"


tree.owls <- read.tree(text = TREE)


plot(tree.owls, type = "c)
```



If you're lucky, you can import an existing phylogeny file.

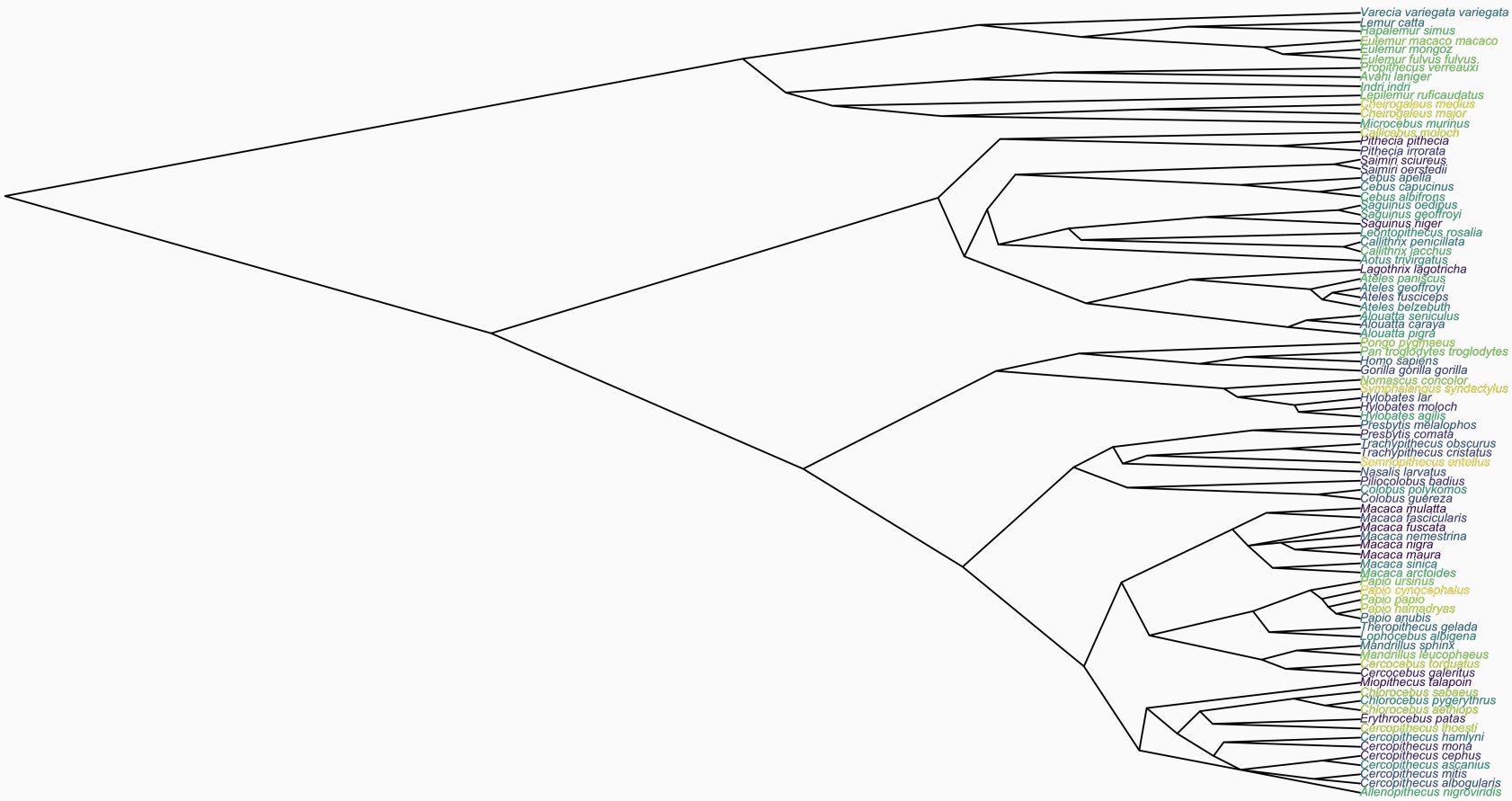
```
tree<-read.tree("primate_tree.phy")

summary(tree)

Phylogenetic tree: tree

  Number of tips: 86
  Number of nodes: 84
  Branch lengths:
    mean: 5.991844
    variance: 37.40793
...


plot(tree, type = "c")
```



Phylogenetic Correlation

The residuals looked good, but closely related species have similar brain sizes.

```
COLS <- viridis::viridis(nrow(data))
plot(tree,  type = "c", tip.color=COLS[order(data$lg_brain)])
```



Phylogenetic Corrections

Phylogenetic correlation structures can be added via the `R` package ape.

Just like spatial and temporal autocorrelation, there are a number of alternatives to chose from:

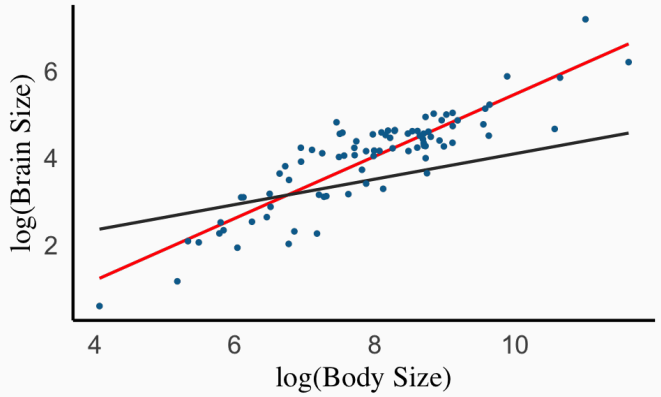
- `corBrownian` Brownian motion model (Felsenstein 1985)
- `corPagel` The cov. matrix defined in Freckelton et al. (2002)
- `corMartins` The cov. matrix defined in Martins and Hansen (1997)
- `corGrafen` The cov. matrix defined in Grafen (1989)
- `corBlomberg` The cov. matrix defined in Blomberg et al. (2003)

Brownian motion correlation in R

The Brownian model calculates covariance matrices exactly like we did earlier.

```
FIT.bm <- gls(lg_brain ~ lg_body,
              correlation = corBrownian(phy = tree,
                                        form = ~species),
              data = data, method = "ML")
```

```
summary(FIT.bm)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC      BIC    logLik
  55.26968 62.63273 -24.63484

Correlation Structure: corBrownian
 Formula: ~species
 Parameter estimate(s):
numeric(0)

Coefficients:
               Value Std.Error  t-value p-value
(Intercept) 1.164443 0.5342900 2.179421  0.0321
lg_body     0.293175 0.0393847 7.443880  0.0000
```



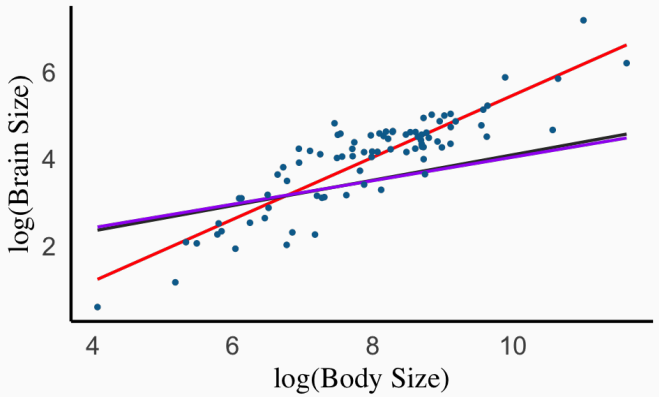

```
AIC(FIT, FIT.bm)
       df       AIC
FIT     3 132.67193
FIT.bm  3  55.26968
```

Pagel's λ Correlation Structure

We can also adjust the covariance matrix according to the strength of the phylogenetic signal in the data. This correlation structure is obtained by multiplying the off-diagonal elements derived from Brownian motion by λ.

```
FIT.pgl <- gls(lg_brain ~ lg_body,
               correlation = corPagel(value = 0.5,
                                      phy = tree,
                                      form = ~ species),
               data = data, method = "ML")
```

```
summary(FIT.pgl)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC      BIC    logLik
  53.24331 63.0607 -22.62166

Correlation Structure: corPagel
 Formula: ~species
 Parameter estimate(s):
  lambda
1.010269

Coefficients:
                Value Std.Error  t-value p-value
(Intercept) 1.3252923 0.5471048 2.422374  0.0176
lg_body     0.2715456 0.0362619 7.488460  0.0000
```




```
AIC(FIT, FIT.bm, FIT.pgl)
        df       AIC
FIT      3 132.67193
FIT.bm   3  55.26968
FIT.pgl  4  53.24331
```

The value of λ is $\approx$ the extent to which species differences are predicted by phylogeny.

The Ornstein-Uhlenbeck process

Under Brownian motion, the expected difference between species is proportional to the amount of time since they split from a common ancestor.

This assumes that species can evolve ever greater differences without any constraints (i.e., evolution is infinitely diffusive). Think of a flying whale with sparrow sized wings.

Real evolution does have some bounds, and the OU process constrains evolutionary change by including an 'attractor', α.

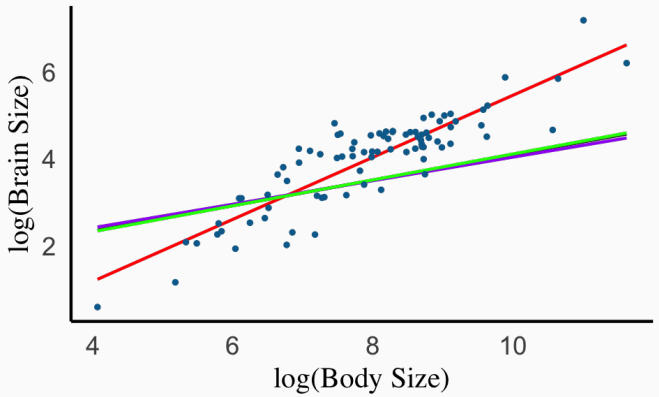
Under an OU process, the farther a species trait evolves away from the attractor, the stronger the tendency for the next step in its evolution to be toward the attractor rather than away from it.

The OU correlation Structure

The OU cov. structure is def. as $_{ij} = \gamma e^{(-\alpha t_{ij})}$ (Martins & Hansen 1997)

where t_{ij} is the phylogenetic distance between species i and j and γ is a constant.

```
FIT.mrt <- gls(lg_brain ~ lg_body,
               correlation = corMartins(value = 0,
                                        phy = tree,
                                        form = ~ species),
               data = data, method = "ML")

summary(FIT.mrt)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC      BIC    logLik
  60.49565 70.31304 -26.24782

Correlation Structure: corMartins
 Formula: ~species
 Parameter estimate(s):
      alpha
0.001953125

Coefficients:
                Value Std.Error  t-value p-value
(Intercept) 1.1255154 1.4586385 0.771621  0.4425
lg_body     0.2987502 0.0396937 7.526379  0.0000
```



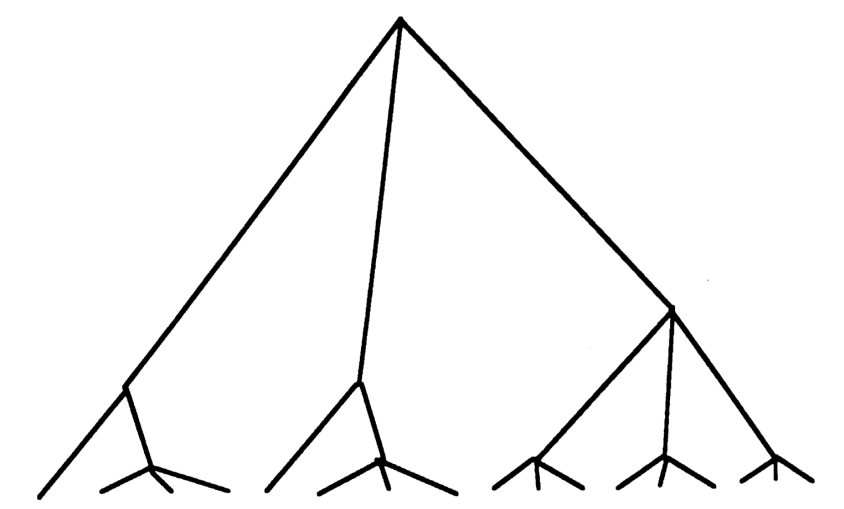

```
AIC(FIT, FIT.bm, FIT.pgl,
    FIT.mrt)
         df       AIC
FIT       3 132.67193
FIT.bm    3  55.26968
FIT.pgl   4  53.24331
FIT.mrt   4  60.49565
```

Grafen's Correlation Structure

Grafen's correlation structure has an additional parameter, $\rho > 0$.

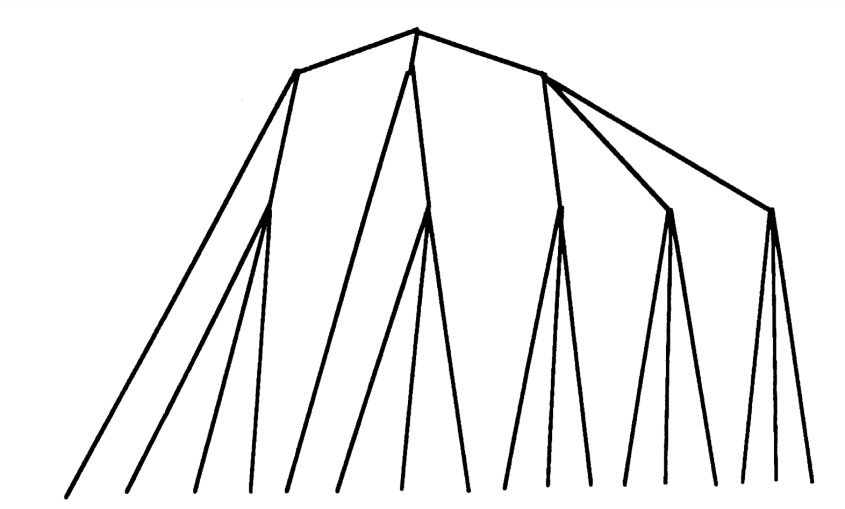
Branch lengths are raised to the power of ρ which allows the tree to be stretched (as a correction for uncertainty in branch lengths)

High value of ρ



(Grafen, 1989)

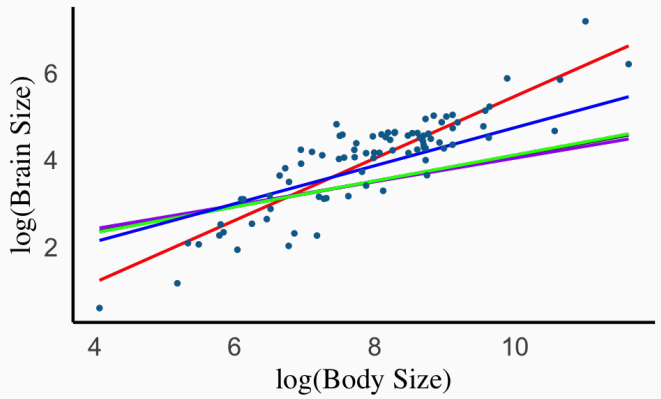
Low value of ρ



(Grafen, 1989)

Grafen's Correlation Structure


```
FIT.grfn <- gls(lg_brain ~ lg_body,
                correlation = corGrafen(value = 0.5,
                                        phy = tree,
                                        form= ~species),
                data = data, method = "ML")

summary(FIT.grfn)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC    BIC   logLik
  78.96361 88.781 -35.4818

Correlation Structure: corGrafen
 Formula: ~species
 Parameter estimate(s):
      rho
0.4234677

Coefficients:
                Value Std.Error  t-value p-value
(Intercept) 0.3586923 0.4425732 0.810470    0.42
lg_body     0.4385990 0.0457473 9.587422    0.00
```

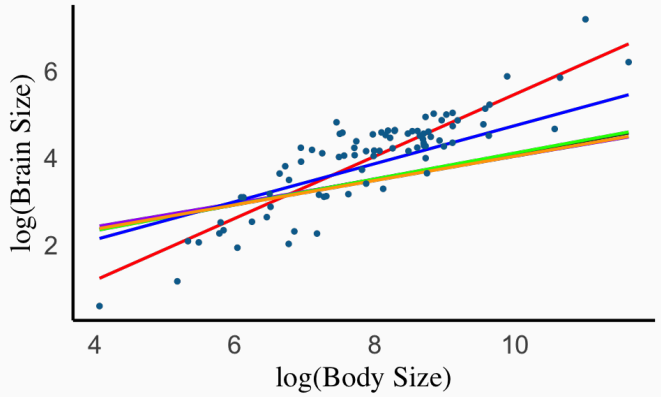



```
AIC(FIT, FIT.bm, FIT.pgl,
    FIT.mrt, FIT.grfn)
          df       AIC
FIT        3 132.67193
FIT.bm     3  55.26968
FIT.pgl    4  53.24331
FIT.mrt    4  60.49565
FIT.grfn   4  78.96361
```

Blomberg et al.'s Correlation Structure

This model assumes that continuous traits evolve under a BM model which rates accelerates (if $g < 1$) or decelerates (if $g > 1$) through time. If $g = 1$, then the model reduces to a Brownian motion model.

```
FIT.blm <- gls(lg_brain ~ lg_body,
               correlation = corBlomberg(value = 0.5,
                                         phy = tree,
                                         form= ~species),
               data = data, method = "ML")

summary(FIT.blm)

Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC      BIC    logLik
  56.29417 66.11156 -24.14708

Correlation Structure: corBlomberg
 Formula: ~species
 Parameter estimate(s):
       g
2.062626

Coefficients:
               Value Std.Error  t-value p-value
(Intercept) 1.230993 0.7602091 1.619282  0.1091
lg_body     0.280819 0.0385963 7.275809  0.0000
```




```
AIC(FIT, FIT.bm, FIT.pgl,
    FIT.mrt, FIT.grfn, FIT.blm)
          df       AIC
FIT        3 132.67193
FIT.bm     3  55.26968
FIT.pgl    4  53.24331
FIT.mrt    4  60.49565
FIT.grfn   4  78.96361
FIT.blm    4  56.29417
```

Corrected model

Original Model

```
Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC     BIC    logLik
  132.6719 140.035 -63.33597

Coefficients:
                 Value Std.Error   t-value p-value
(Intercept) -1.6860136 0.3310534 -5.092874       0
lg_body      0.7154667 0.0411488 17.387323       0

 Correlation:
        (Intr)
lg_body -0.986

Residual standard error: 0.5053653
Degrees of freedom: 86 total; 84 residual
```

Pagel's λ correlation model

```
Generalized least squares fit by maximum likelihood
  Model: lg_brain ~ lg_body
  Data: data
       AIC     BIC    logLik
  53.24331 63.0607 -22.62166

Coefficients:
                Value Std.Error  t-value p-value
(Intercept) 1.3252923 0.5471048 2.422374  0.0176
lg_body     0.2715456 0.0362619 7.488460  0.0000

 Correlation:
        (Intr)
lg_body -0.496

Residual standard error: 0.8674091
Degrees of freedom: 86 total; 84 residual

Correlation Structure: corPagel
 Formula: ~species
 Parameter estimate(s):
  lambda
1.010269
```

Technical considerations

This type of analysis is really sensitive to missing data. If some species have NAs for certain traits, but are listed in the tree, the model will fail (cov matrix isn't calculated correctly).

The workflow we went over today assumes 1 datapoint per species (we used species means). If you want to include everything in the model, it's possible, but gets messy.

References

Arnold, C., Matthews, L.J. & Nunn, C.L. (2010). The 10ktrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews*, 19, 114–118.

Bowling, D.L., Dunn, J.C., Smaers, J.B., Garcia, M., Sato, A., Hantke, G., Handschuh, S., Dengg, S., Kerney, M., Kitchener, A.C. *et al.* (2020). Rapid evolution of the primate larynx? *PLoS biology*, 18, e3000764.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the Preservation of Favoured Races in the Struggle for Life*. 1st edn. John Murray, London.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125, 1–15.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 326, 119–157.

Hilborn, R. & Mangel, M. (1997). *The ecological detective: confronting models with data*. vol. 28. Princeton University Press.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M.J., Sutter, N.B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A.C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. & Lander, E.S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438, 803–819.