# Bioinformatics Lab3

*Milda Poceviciute, Fanny Karelius, Rab Nawaz Jan Sher and Saman Zahid*

*5 December 2018*

In step one we make a vector of accession number to get data of relevant species to be get from Genbank, then we fetch data.

```
library(msa)
library(ape)

x <- paste("AJ5345", 26:49, sep = "")
x <- c("Z73494", x)

#read a data from GenBank
sylvia.seq <- read.GenBank(x)
sylvia.seq
```

```
## 25 DNA sequences in binary format stored in a list.
##
## Mean sequence length: 1134.84
##    Shortest sequence: 1041
##     Longest sequence: 1143
##
## Labels:
## Z73494
## AJ534526
## AJ534527
## AJ534528
## AJ534529
## AJ534530
## ...
##
## Base composition:
##     a     c     g     t
## 0.270 0.354 0.131 0.245
## (Total: 28.37 kb)
```

It can be seen that sequence length is not same for all species. The shortest sequence is of length 1041 that is some data is missing.

In next step we make a taxon of species containing name of each specie. We have updated the names data with accession number "Z73494" and "AJ534548" with respect to its name in genbank.

```
taxa.sylvia <- attr(sylvia.seq, "species")
names(taxa.sylvia) <- names(sylvia.seq)

taxa.sylvia[1] <- "Sylvia_atricapilla"
taxa.sylvia[24] <- "Sylvia_abyssinica"
```

In this step trait data has been loaded which has mig.distance, mig.behaviour and geographic range of sylvia.

```
sylvia.eco <- read.table("sylvia_data.txt")
str(sylvia.eco)
```

```
## 'data.frame':    26 obs. of  3 variables:
##  $ mig.dist : int  0 5000 7500 5900 5500 3400 2600 0 0 0 ...
##  $ mig.behav: Factor w/ 3 levels "long","resid",..: 2 3 1 1 1 1 1 2 2 2 ...
##  $ geo.range: Factor w/ 3 levels "temp","temptrop",..: 3 2 2 2 2 2 2 3 3 3 ...
```

```
rownames(sylvia.eco)
```

```
##  [1] "Sylvia_abyssinica"    "Sylvia_atricapilla"   "Sylvia_borin"
##  [4] "Sylvia_nisoria"       "Sylvia_curruca"       "Sylvia_hortensis"
##  [7] "Sylvia_crassirostris" "Sylvia_leucomelaena"  "Sylvia_buryi"
## [10] "Sylvia_lugens"        "Sylvia_layardi"       "Sylvia_subcaeruleum"
## [13] "Sylvia_boehmi"        "Sylvia_nana"          "Sylvia_deserti"
## [16] "Sylvia_communis"      "Sylvia_conspicillata" "Sylvia_deserticola"
## [19] "Sylvia_undata"        "Sylvia_sarda"         "Sylvia_balearica"
## [22] "Sylvia_cantillans"    "Sylvia_mystacea"      "Sylvia_melanocephala"
## [25] "Sylvia_rueppelli"     "Sylvia_melanothorax"
```

We have determined the multiple sequence alignment and converted it into DNAbin object. Then pairwise distance from DNA sequence has been determined by using the multiple sequence alignment created. Kimura(K80), Galtier and Gouy(GG95),Jukes and Cantor(JC69) and raw distance (that is proportion or the number of sites that differ between each pair of sequences) are calculated.

We have used pairwise.deletion = TRUE because we have observed that there is missing data in our sequence (i-e of length 1041). Pairwise.deletion is a logical indicating whether to delete the sites with missing data in a pairwise way.

```
write.dna(sylvia.seq, file ="sylvia_seqs.fasta", format = "fasta")

sylvia_aligment <- msa("sylvia_seqs.fasta", type="dna")
```

```
## use default substitution matrix
```

```
sylvia.seq.ali <- as.DNAbin(sylvia_aligment)
rownames(sylvia.seq.ali) <- names(sylvia.seq)


syl.K80 <- dist.dna(sylvia.seq.ali, pairwise.deletion = TRUE)
syl.GG95 <- dist.dna(sylvia.seq.ali, model = "GG95", p = TRUE)
syl.JC69 <- dist.dna(sylvia.seq.ali, model = "JC69", p = TRUE)
syl.raw <- dist.dna(sylvia.seq.ali, model = "raw", p = TRUE)
```

Then we make the phylogenetic tree using neighbour joining tree and the distance matrix calculated above.

```
nj.sylvia.K80 <- nj(syl.K80)
nj.sylvia.GG95 <- nj(syl.GG95)

dist.topo(nj.sylvia.K80, nj.sylvia.GG95)
```

```
##       tree1
## tree2    12
```

then we determine the topological distance (that is counts the number of bipartitions shared between two trees). The distance matrix used here are K80 and GG95.

Since nj tree returns an unrooted tree, we are converting into rooted tree by setting the outgroup(the group wrt which roots are formed) as “AJ534526” specie in function f. This function is used in boot.phylo to sample the phylogenetic tree.

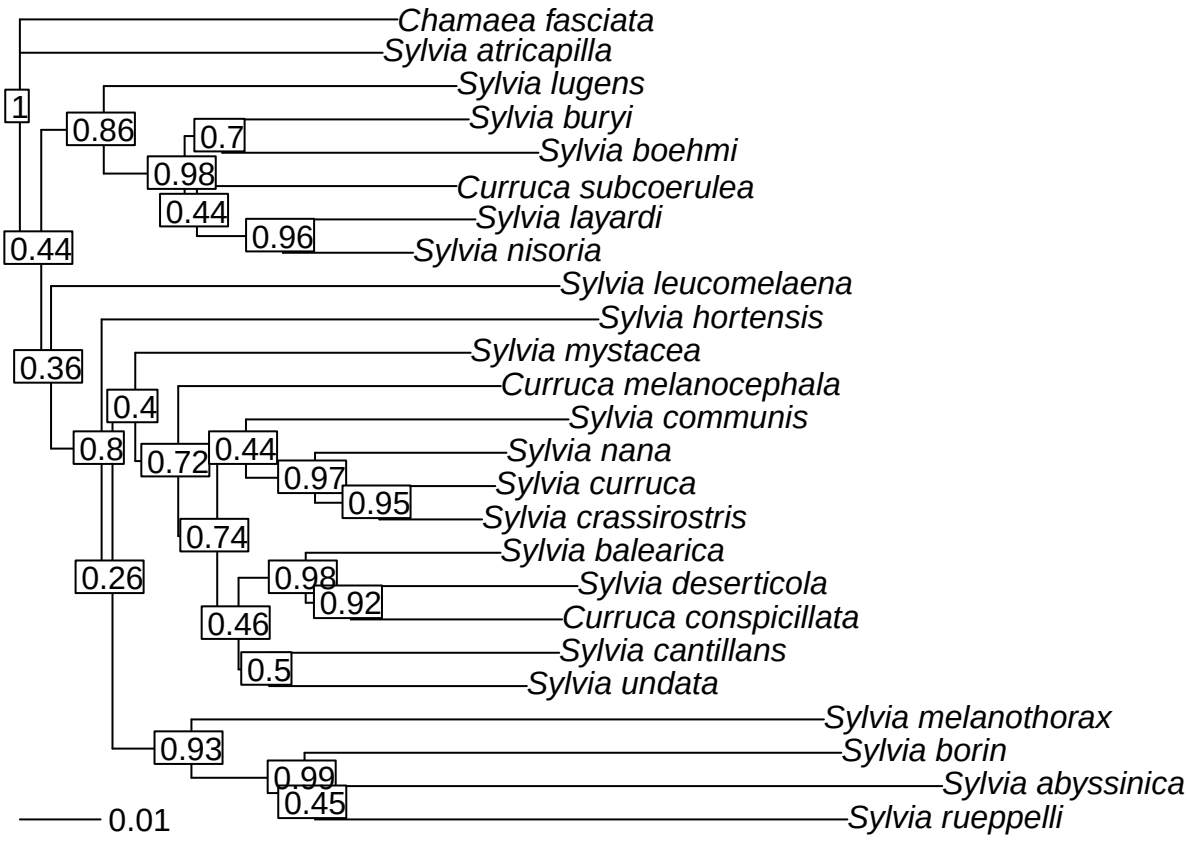
```
f <- function(xx) root(nj(dist.dna(xx, p=TRUE)), "AJ534526")
tr <- f(sylvia.seq.ali)
## same than: tr <- root(nj.sylvia.K80, "AJ534526")

nj.boot.sylvia <- boot.phylo(tr, sylvia.seq.ali, f, 200,
                             rooted = TRUE)
```

```
##
Running bootstraps:       100 / 200
Running bootstraps:       200 / 200
## Calculating bootstrap values... done.
```

```
nj.est <- tr
```

updating tip label of nj estimated to taxa.sylvia tips labels and then plotting the estimated phylogenetic tree.

```
nj.est$tip.label <- taxa.sylvia[tr$tip.label]
plot(nj.est, no.margin = TRUE)
nodelabels(round(nj.boot.sylvia / 200, 2), bg = "white")
add.scale.bar(length = 0.01)
```




```
write.tree(nj.est, "sylvia_nj_k80.tre")
```

The estimated proportion of each node has been shown in the tree. The tree is then saved in file **sylvia_nj_k80.tre**

## Question 1.2

Creating different models by changing the model parameter.

```
load("sylvia.RData")
nj.est <- read.tree("sylvia_nj_k80.tre")
nj.est <- drop.tip(nj.est, "Chamaea_fasciata")
DF <- sylvia.eco[nj.est$tip.label, ]
table(DF$geo.range, DF$mig.behav)
```

```
##
##            long resid short
##   temp        0     4     0
##   temptrop    9     0     2
##   trop        0     6     0
```

```
syl.er <- ace(DF$geo.range, nj.est, type = "d") # default is "ER"
syl.sym <- ace(DF$geo.range, nj.est, type="d", model="SYM")
syl.ard <- ace(DF$geo.range, nj.est, type="d", model="ARD")
```

```
sapply(list(syl.er, syl.sym, syl.ard), AIC)
```

```
## [1] 42.21419 44.15340 48.88874
```

We know that smaller the AIC better will be the fit, thus model with Equal rates is the best from the above result.

Model with Equal rates:

```
er_df <- data.frame(estimated_rate = syl.er$rates, standard_error = syl.er$se)

knitr::kable(er_df)
```

| estimated_rate | standard_error |
|---:|---:|
| 6.481183 | 2.442814 |

Symmetrical model:

```
sym_df <- data.frame(estimated_rate = syl.sym$rates, standard_error = syl.sym$se)

knitr::kable(sym_df)
```

| estimated_rate | standard_error |
|---:|---:|
| 6.387960 | 3.281622 |
| 0.000000 | NaN |
| 8.182045 | 2.146266 |

Model with different rates (no rate is same):

```
ard_df <- data.frame(estimated_rate = syl.ard$rates, standard_error = syl.ard$se)

knitr::kable(ard_df)
```

| estimated_rate | standard_error |
|---|---|
| 12.86408 | NaN |
| 0.00000 | NaN |
| 16.34903 | 6.041339 |
| 13.31018 | 4.896857 |
| 11.70748 | NaN |
| 0.00000 | NaN |

From the estimated rates and standard error, it can be seen that there is more uncertainty in the result of sym and ard models.

```
Q <- syl.er$index.matrix
diag(Q) <- 0
Q[1, 2] <- Q[2, 1] <- syl.er$rates[1]
Q[2, 3] <- Q[3, 2] <- syl.er$rates[1]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)

P
```

```
##                temp  temptrop       trop
## temp     0.72356901 0.2072472 0.06918383
## temptrop 0.20724716 0.5855057 0.20724716
## trop     0.06918383 0.2072472 0.72356901
```

In ER model the transition occurs on the basis of equal probability. It can be seen that temp to temp and trop to trop transition has high probability as compared to temptrop to temptrop thus it can be assumed that temptrop is least stable of the three.

```
Q <- syl.sym$index.matrix
diag(Q) <- 0
Q[1, 2] <- Q[2, 1] <- syl.sym$rates[1]
Q[2, 3] <- Q[3, 2] <- syl.sym$rates[3]
Q[1, 3] <- Q[3, 1] <- syl.sym$rates[2]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)
P
```

```
##                temp  temptrop      trop
## temp     0.7597780 0.1989812 0.0412408
## temptrop 0.1989812 0.5577356 0.2432832
## trop     0.0412408 0.2432832 0.7154760
```

In symmetric model, the results are almost same as the ER model that is temp to temp and trop to trop transition has high probability as compared to temptrop to temptrop thus it can be assumed that temptrop is least stable of the three in this case as well.

```
Q <- syl.ard$index.matrix
diag(Q) <- 0

Q[1, 2]  <- syl.ard$rates[1]
Q[2, 1]  <- syl.ard$rates[3]
Q[2, 3] <- syl.ard$rates[4]
Q[3, 2] <- syl.ard$rates[6]
Q[1, 3] <- syl.ard$rates[2]
Q[3, 1] <- syl.ard$rates[5]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)
P
```

```
##               temp  temptrop       trop
## temp     0.6557337 0.2517611 0.09250518
## temptrop 0.4041527 0.3270385 0.26880879
## trop     0.3426719 0.0813665 0.57596161
```

In ARD model all rates are different that is it is not symmetric therefore different rates are assigned while making Q matrix.

From the result, it can be seen that temp to temp probability is highest of all so temp to temp is the most stable one, trop to trop also have considerably high probability but the transition of temptrop to temptrop has even lower probability than temptop to temp. Thus temptrop is the least stable in this case as well.

Therefore we can say that temp and trop has higher chances of surviving through evolution

```
#install.packages("ade4")
#install.packages("mvMORPH")
#install.packages("mvSLOUCH")
#install.packages("adephylo")
library(ade4)
```

```
##
## Attaching package: 'ade4'

## The following object is masked from 'package:Biostrings':
##
##     score

## The following object is masked from 'package:IRanges':
##
##     score

## The following object is masked from 'package:BiocGenerics':
##
##     score
```

```
library(mvMORPH)
```

```
## Loading required package: phytools

## Loading required package: maps

## Loading required package: corpcor
```

```
## Loading required package: subplex
```

```
## ##
## ## mvMORPH package (1.1.0)
```

```
## ## Multivariate evolutionary models
```

```
## ##
## ## See the tutorials: browseVignettes("mvMORPH")
```

```
## ##
## ## To cite package 'mvMORPH': citation("mvMORPH")
## ##
```

```
library(mvSLOUCH)
```

```
## Loading required package: ouch
```

```
## Loading required package: numDeriv
```

```
## Loading required package: mvtnorm
```

```
library(ape)
#devtools::install_github("kopperud/slouch")
# install.packages("rlang")
# install.packages("tibble")
library(tibble)
library(rlang)
```

# Questions 2

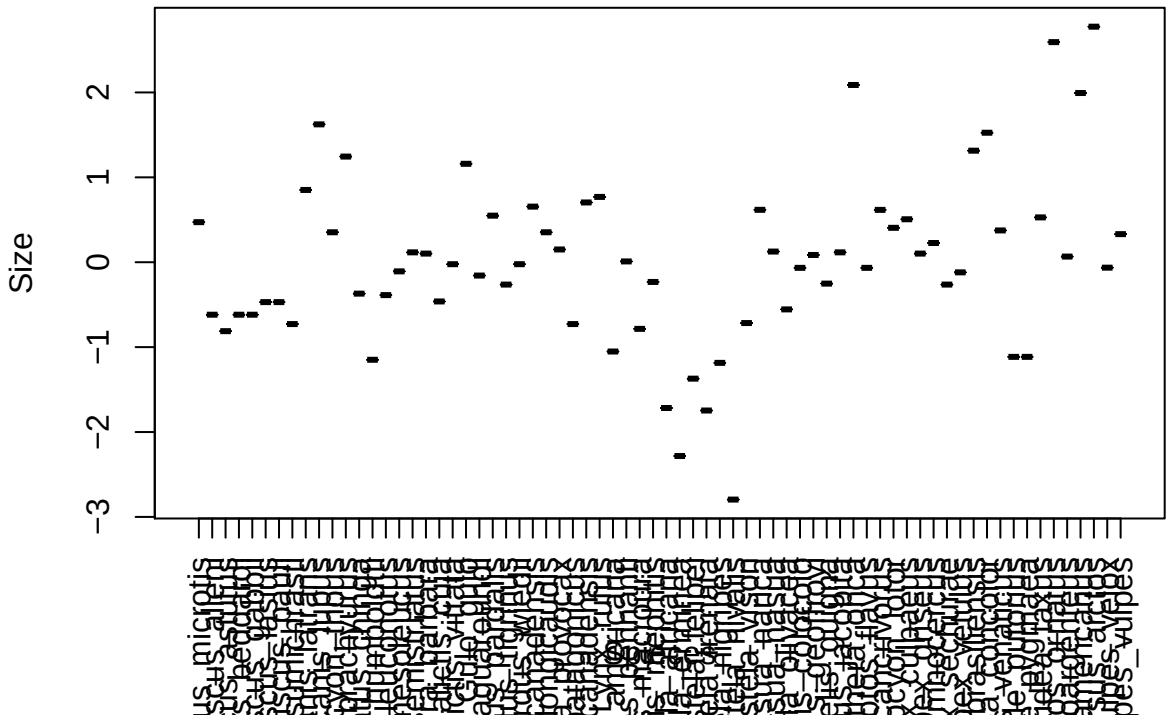
## Q2.1

In this question we analyse the data from *ade4* package. It contains the phylogeny of 70 carnivora as reported by Diniz-Filho and Torres (2002), and provides the geographic range size and body size corresponding to them.

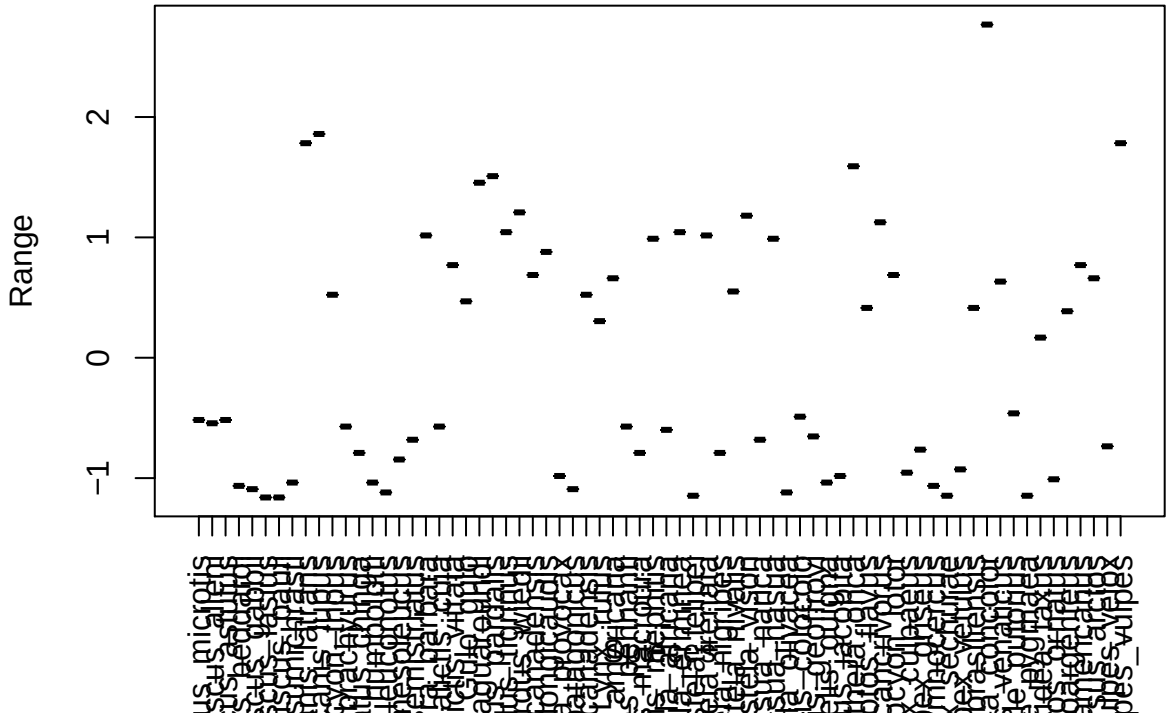
```
data(carni70)
carni70_tree <- newick2phylog(carni70$tre)


size <- scalewt(log(carni70$tab))[,1] # scaled log body size in kg
names(size) <- row.names(carni70$tab)
yrange <- scalewt(carni70$tab[,2])# scaled geographic range in km
names(yrange) <- row.names(carni70$tab)
plot_data <- data.frame(spieces = row.names(carni70$tab), size = size, range =yrange )
```

Plots below show the body sizes (scaled log), the ranges (scaled), as well as size and range dependencies that correspond to the different species. There seem to be no obvious linear relation between these two traits.

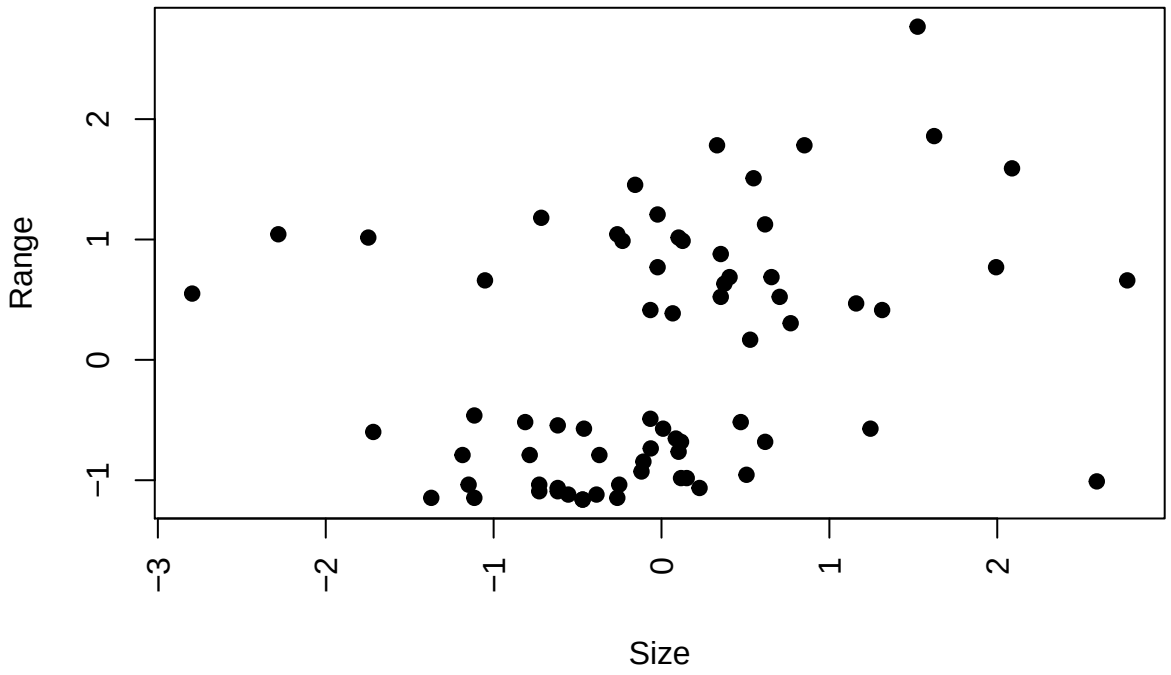
```
par(mfrow=c(1,1),las=3)

plot(plot_data$spieces, plot_data$size, xlab = "Spieces", ylab = "Size", pch = 19,
     col = rgb(0, 0, 0, 0.1))
```




```
plot(plot_data$spieces, plot_data$range, xlab = "Spieces", ylab = "Range", pch = 19,
     col = rgb(0, 0, 0, 0.1))
```

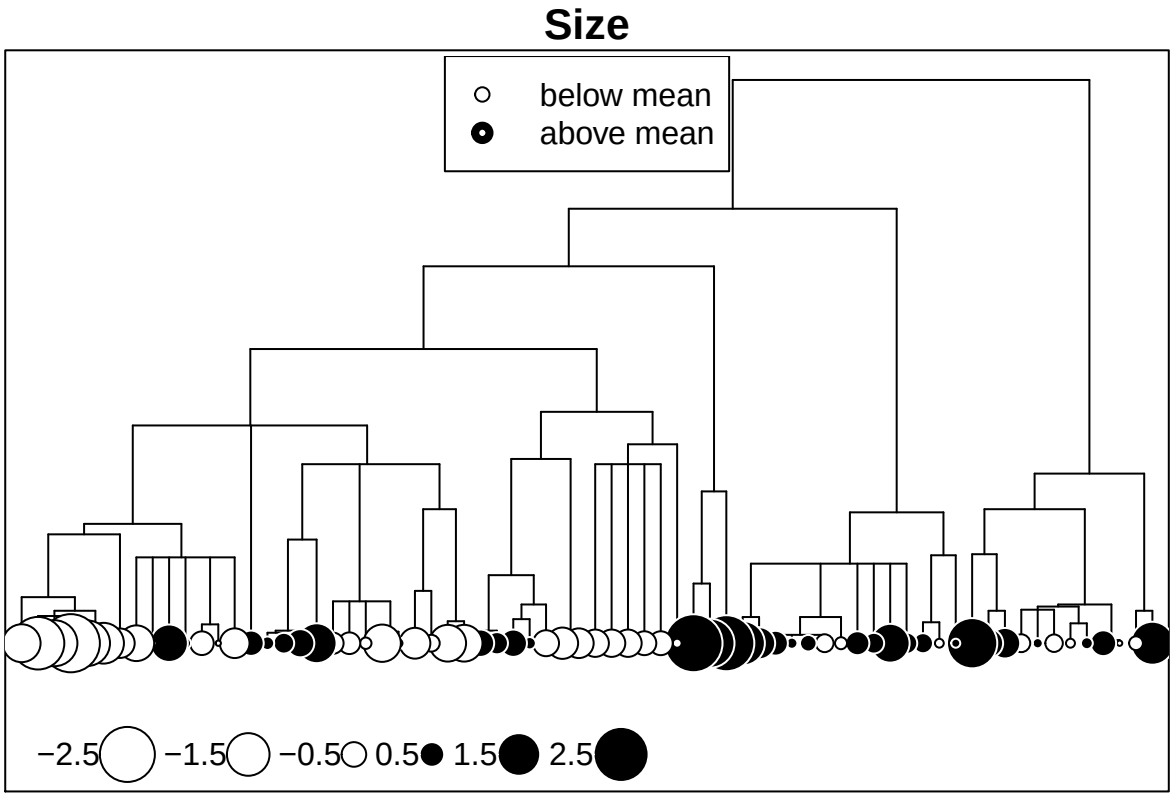



```
plot(plot_data$size, plot_data$range, xlab = "Size", ylab = "Range", pch = 19)
```

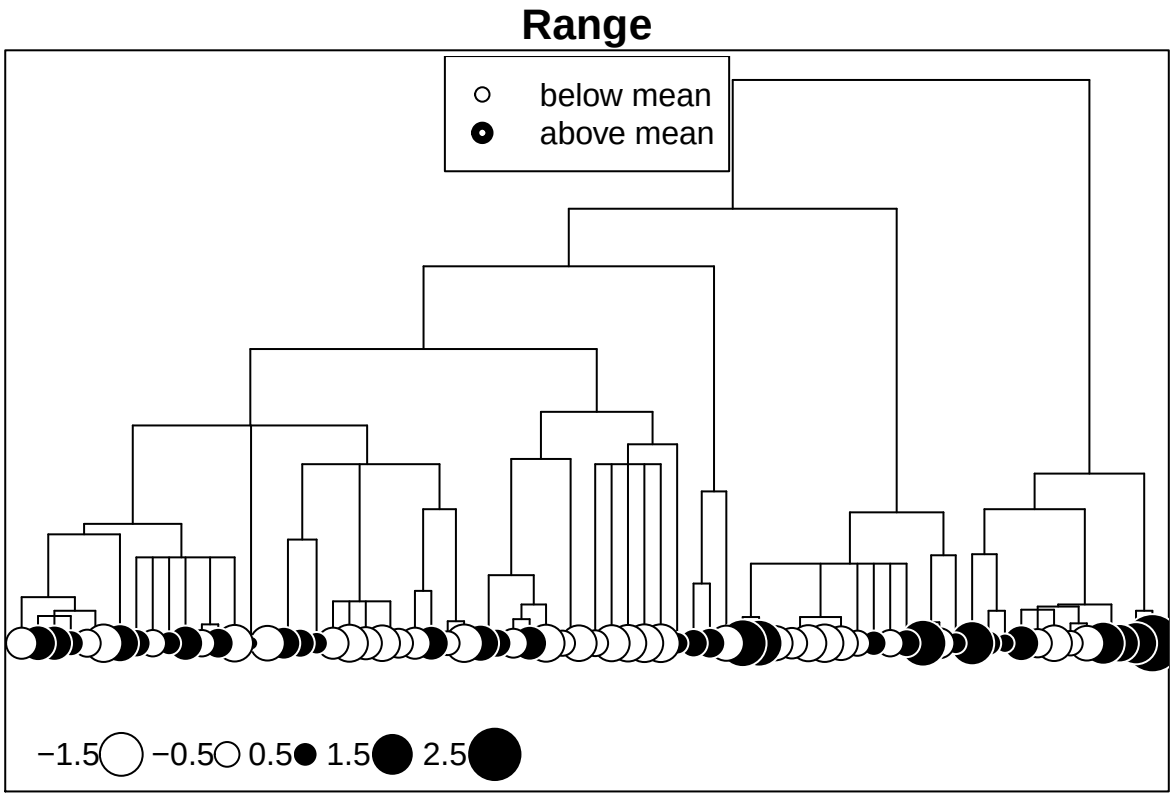



The plots below shows the phylogenetic tree. The first plot shows how the size trait evolved for different species. The white nodes represent the size below the mean while the black ones represent the body sizes above the mean. Similarly, the second tree plot shows the phylogenetic tree with respect to range.

```
tre <- as.phylo(carni70_tree)
carni70_tree <- newick2phylog(carni70$tre)

par(mfrow = c(1, 1), oma = c(1, 1, 2, 1))
symbols.phylog(carni70_tree,size)
mtext("Size", side = 3, line = 0, outer = TRUE, font = 2, cex = 1.3)
par(fig = c(0, 1, .1, .9), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("top", c("below mean", "above mean"), col = c("black", "black"), lty=c(NA, NA),pch = c(1, 1), lwd
```




```
par(mfrow = c(1, 1), oma = c(1, 1, 2, 1))
symbols.phylog(carni70_tree,yrange)
mtext("Range", side = 3, line = 0, outer = TRUE, font = 2, cex = 1.3)
par(fig = c(0, 1, .1, .9), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("top", c("below mean", "above mean"), col = c("black", "black"), lty=c(NA, NA),pch = c(1, 1), lwd
```




### Q2.2

**Both traits evolve as independent Brownian motions**

```
fit1_size <- mvBM(tre, data = carni70$tab[,1], model="BM1")
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results for multiple rate BM1 model --
## LogLikelihood:    -324.2639
## AIC:      652.5278
## AICc:     652.7069
## 2 parameters
##
## Estimated rate matrix
## ---------------------
##
##  73.70335
##
## Estimated root state
## ---------------------
##
## theta: 38.43947
```

```
fit1_range <- mvBM(tre, data = carni70$tab[,2], model="BM1")
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results for multiple rate BM1 model --
## LogLikelihood:    -264.9734
## AIC:      533.9469
## AICc:     534.126
## 2 parameters
##
## Estimated rate matrix
## ---------------------
##
##  13.54524
##
## Estimated root state
## ---------------------
##
## theta: 13.78439
```

The models above are independent BMs for the size and range traits. The models estimated two parameters each: * Evolutionary rate matrix: is related to the variance. Hence, the trait size has much higher variance than the range trait:

```
fit1_size$sigma
```

```
##
##  73.70335
```

```
fit1_range$sigma
```

```
##
##  13.54524
```

- Estimated ancestral states: this shows that the expected value of the body size is higher than the expected value of the range.

```
fit1_size$theta
```

```
##
## theta: 38.43947
```

```
fit1_range$theta
```

```
##
## theta: 13.78439
```

**The traits evolve as a correlated Brownian motion.**

```
fit2 <- mvBM(tre, data = carni70$tab, model="BM1")
```

```
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results for multiple rate BM1 model --
```

```
## LogLikelihood:    -588.7853
## AIC:      1187.571
## AICc:     1188.018
## 5 parameters
##
## Estimated rate matrix
## ---------------------
##            size     range
## size  73.703348  3.579341
## range  3.579341 13.545240
##
## Estimated root state
## ---------------------
##            size    range
## theta: 38.43947 13.78439
```

The same parameters are estimated when both traits are assumed to have correlation between them:

```
cat(" Evolutionary rate matrix:")
```

```
##  Evolutionary rate matrix:
```

```
cat("\n")
```

```
fit2$sigma
```

```
##            size     range
## size  73.703348  3.579341
## range  3.579341 13.545240
```

```
cat(" Estimated ancestral states:")
```

```
##  Estimated ancestral states:
```

```
cat("\n")
```

```
fit2$theta
```

```
##            size    range
## theta: 38.43947 13.78439
```

We see that range and size has a positive covariance under this model. The variances and the ancestral states are very close to the ones estimated by the univariate models.

### Both traits evolve as independent Ornstein-Uhlenbeck processes

```
fit3_size <- mvOU(tre, data = carni70$tab[,1], model="OU1")
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results --
## LogLikelihood:    -324.2639
## AIC:      654.5278
## AICc:     654.8914
## 3 parameters
##
```

```
## Estimated theta values
## ----------------------
##
## OU1 38.43947
##
## ML alpha values
## ----------------------
##
##  8.737191e-10
##
## ML sigma values
## ----------------------
##
##  73.70246
```

```
fit3_range <- mvOU(tre, data = carni70$tab[,2], model="OU1")
```

```
## species in the matrix are assumed to be in the same order as in the phylogeny, otherwise specify row
## successful convergence of the optimizer
## a reliable solution has been reached
##
## -- Summary results --
## LogLikelihood:    -248.9895
## AIC:      503.979
## AICc:     504.3427
## 3 parameters
##
## Estimated theta values
## ----------------------
##
## OU1 11.08609
##
## ML alpha values
## ----------------------
##
##  0.2295954
##
## ML sigma values
## ----------------------
##
##  38.05329
```

The two models estimated independent Ornstein-Uhlenbeck processes for each trait. The parameters estimated are:

```
cat("Alpha: strength of selection - low alpha favours a random change during the time step \n")
```

```
## Alpha: strength of selection - low alpha favours a random change during the time step
```

```
fit3_size$alpha
```

```
##
##  8.737191e-10
```

```
fit3_range$alpha
```

```
##
```

```
##  0.2295954
```

```
cat("Sigma: evolutionary rate matrix \n")
```

```
## Sigma: evolutionary rate matrix
```

```
fit3_size$sigma
```

```
##
##  73.70246
```

```
fit3_range$sigma
```

```
##
##  38.05329
```

```
cat("Theta: estimated ancestral states \n")
```

```
## Theta: estimated ancestral states
```

```
fit3_size$theta
```

```
##
## OU1 38.43947
```

```
fit3_range$theta
```

```
##
## OU1 11.08609
```

It seems that the size trait is more likely to change over time than the range trait due to much lower alpha value of size. The sigma and theta values are very similar to the ones from the independent Brownian motion models.

### The traits evolve as a bivariate Ornstein-Uhlenbeck process

```
phyltree = ape2ouch(tre)
fit5 <- ouchModel(phyltree = phyltree, data = carni70$tab)
```

The bivariate OU model estimated parameters:

```
cat("A: strength of selection")
cat("\n")
fit5$FinalFound$ParamsInModel$A
cat("vY0: estimated ancestral state for range (OU part)")
cat("\n")
fit5$FinalFound$ParamsInModel$vY0
cat("Syy is the sigma value for range")
cat("\n")
fit5$FinalFound$ParamsInModel$Syy
```

### *size* evolves as a Brownian motion and *range* as an Ornstein-Uhlenbeck process adapting to it

```
traits <-  data.frame(range=carni70$tab[,2],size=carni70$tab[,1])
row.names(traits) <- row.names(carni70$tab)
```

```
fit6 <-mvslouchModel(phyltree, traits,kY=1)
#
```

The bivariate OUBM model estimated these parameters:

```
cat("A: strength of selection")
cat("\n")
fit6$FinalFound$ParamsInModel$A
cat("vY0: estimated ancestral state for range (OU part)")
cat("\n")
fit6$FinalFound$ParamsInModel$vY0
cat("vX0: estimated ancestral state for size (BM part)")
cat("\n")
fit6$FinalFound$ParamsInModel$vX0
cat("Syy is the sigma value for range")
cat("\n")
fit6$FinalFound$ParamsInModel$Syy
cat("Sxx is the sigma value for size")
cat("\n")
fit6$FinalFound$ParamsInModel$Sxx
```

We can see that sigma values changes considerably in comparison to the first three models, but the estimated ancestral states remained almost the same.

### Comparison

All 5 models above tried to fit the parameters based on the given phylogenetic tree, and the observed trait values for the 70 species. We compare the models based on the AICc criterion:

```
cat("Comparison of the AICc values:\n")
cat("Both traits evolve as independent Brownian motions \n")
fit1_size$AICc + fit1_range$AICc
cat("The traits evolve as a correlated Brownian motion \n")
fit2$AICc
cat("Both traits evolve as independent Ornstein-Uhlenbeck processes \n")
fit3_size$AICc+fit3_range$AICc
cat("The traits evolve as a bivariate Ornstein-Uhlenbeck process \n")
fit5$`FinalFound`$ParamSummary$aic.c
cat("size evolves as a Brownian motion and range as an Ornstein-Uhlenbeck process \n")
fit6$`FinalFound`$ParamSummary$aic.c
```

To compare AICc criterion, we sum the values for the BM independent models. It seems the bivariate OUBM model is the best fit since it has the lowest AICc value.

```
knitr::opts_chunk$set(echo = TRUE)
library(msa)
library(ape)

x <- paste("AJ5345", 26:49, sep = "")
x <- c("Z73494", x)

#read a data from GenBank
sylvia.seq <- read.GenBank(x)
sylvia.seq
```

```
taxa.sylvia <- attr(sylvia.seq, "species")
names(taxa.sylvia) <- names(sylvia.seq)

taxa.sylvia[1] <- "Sylvia_atricapilla"
taxa.sylvia[24] <- "Sylvia_abyssinica"


sylvia.eco <- read.table("sylvia_data.txt")
str(sylvia.eco)
rownames(sylvia.eco)


write.dna(sylvia.seq, file ="sylvia_seqs.fasta", format = "fasta")

sylvia_aligment <- msa("sylvia_seqs.fasta", type="dna")
sylvia.seq.ali <- as.DNAbin(sylvia_aligment)
rownames(sylvia.seq.ali) <- names(sylvia.seq)


syl.K80 <- dist.dna(sylvia.seq.ali, pairwise.deletion = TRUE)
syl.GG95 <- dist.dna(sylvia.seq.ali, model = "GG95", p = TRUE)
syl.JC69 <- dist.dna(sylvia.seq.ali, model = "JC69", p = TRUE)
syl.raw <- dist.dna(sylvia.seq.ali, model = "raw", p = TRUE)

nj.sylvia.K80 <- nj(syl.K80)
nj.sylvia.GG95 <- nj(syl.GG95)

dist.topo(nj.sylvia.K80, nj.sylvia.GG95)


f <- function(xx) root(nj(dist.dna(xx, p=TRUE)), "AJ534526")
tr <- f(sylvia.seq.ali)
## same than: tr <- root(nj.sylvia.K80, "AJ534526")

nj.boot.sylvia <- boot.phylo(tr, sylvia.seq.ali, f, 200,
                             rooted = TRUE)
nj.est <- tr

nj.est$tip.label <- taxa.sylvia[tr$tip.label]
plot(nj.est, no.margin = TRUE)
nodelabels(round(nj.boot.sylvia / 200, 2), bg = "white")
add.scale.bar(length = 0.01)
write.tree(nj.est, "sylvia_nj_k80.tre")
load("sylvia.RData")
nj.est <- read.tree("sylvia_nj_k80.tre")
nj.est <- drop.tip(nj.est, "Chamaea_fasciata")
DF <- sylvia.eco[nj.est$tip.label, ]
table(DF$geo.range, DF$mig.behav)

syl.er <- ace(DF$geo.range, nj.est, type = "d") # default is "ER"
syl.sym <- ace(DF$geo.range, nj.est, type="d", model="SYM")
syl.ard <- ace(DF$geo.range, nj.est, type="d", model="ARD")
```

```r
sapply(list(syl.er, syl.sym, syl.ard), AIC)


er_df <- data.frame(estimated_rate = syl.er$rates, standard_error = syl.er$se)

knitr::kable(er_df)

sym_df <- data.frame(estimated_rate = syl.sym$rates, standard_error = syl.sym$se)

knitr::kable(sym_df)
ard_df <- data.frame(estimated_rate = syl.ard$rates, standard_error = syl.ard$se)

knitr::kable(ard_df)
Q <- syl.er$index.matrix
diag(Q) <- 0
Q[1, 2] <- Q[2, 1] <- syl.er$rates[1]
Q[2, 3] <- Q[3, 2] <- syl.er$rates[1]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)

P
Q <- syl.sym$index.matrix
diag(Q) <- 0
Q[1, 2] <- Q[2, 1] <- syl.sym$rates[1]
Q[2, 3] <- Q[3, 2] <- syl.sym$rates[3]
Q[1, 3] <- Q[3, 1] <- syl.sym$rates[2]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)
P
Q <- syl.ard$index.matrix
diag(Q) <- 0

Q[1, 2]  <- syl.ard$rates[1]
Q[2, 1]  <- syl.ard$rates[3]
Q[2, 3] <- syl.ard$rates[4]
Q[3, 2] <- syl.ard$rates[6]
Q[1, 3] <- syl.ard$rates[2]
Q[3, 1] <- syl.ard$rates[5]

# Q[] <- c(0, syl.mod$rates)[Q + 1]
diag(Q) <- -rowSums(Q)
P <- matexpo(0.05 * Q)
rownames(P) <- c("temp", "temptrop", "trop")
colnames(P) <- rownames(P)
P
```

```r
#install.packages("ade4")
#install.packages("mvMORPH")
#install.packages("mvSLOUCH")
#install.packages("adephylo")
library(ade4)
library(mvMORPH)
library(mvSLOUCH)
library(ape)
#devtools::install_github("kopperud/slouch")
# install.packages("rlang")
# install.packages("tibble")
library(tibble)
library(rlang)
data(carni70)
carni70_tree <- newick2phylog(carni70$tre)


size <- scalewt(log(carni70$tab))[,1] # scaled log body size in kg
names(size) <- row.names(carni70$tab)
yrange <- scalewt(carni70$tab[,2])# scaled geographic range in km
names(yrange) <- row.names(carni70$tab)
plot_data <- data.frame(spieces = row.names(carni70$tab), size = size, range =yrange )


par(mfrow=c(1,1),las=3)

plot(plot_data$spieces, plot_data$size, xlab = "Spieces", ylab = "Size", pch = 19,
     col = rgb(0, 0, 0, 0.1))

plot(plot_data$spieces, plot_data$range, xlab = "Spieces", ylab = "Range", pch = 19,
     col = rgb(0, 0, 0, 0.1))

plot(plot_data$size, plot_data$range, xlab = "Size", ylab = "Range", pch = 19)


tre <- as.phylo(carni70_tree)
carni70_tree <- newick2phylog(carni70$tre)

par(mfrow = c(1, 1), oma = c(1, 1, 2, 1))
symbols.phylog(carni70_tree,size)
mtext("Size", side = 3, line = 0, outer = TRUE, font = 2, cex = 1.3)
par(fig = c(0, 1, .1, .9), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("top", c("below mean", "above mean"), col = c("black", "black"), lty=c(NA, NA),pch = c(1, 1), lwd

par(mfrow = c(1, 1), oma = c(1, 1, 2, 1))
symbols.phylog(carni70_tree,yrange)
mtext("Range", side = 3, line = 0, outer = TRUE, font = 2, cex = 1.3)
par(fig = c(0, 1, .1, .9), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("top", c("below mean", "above mean"), col = c("black", "black"), lty=c(NA, NA),pch = c(1, 1), lwd
```

```
fit1_size <- mvBM(tre, data = carni70$tab[,1], model="BM1")
fit1_range <- mvBM(tre, data = carni70$tab[,2], model="BM1")
fit1_size$sigma
fit1_range$sigma
fit1_size$theta
fit1_range$theta

fit2 <- mvBM(tre, data = carni70$tab, model="BM1")
cat(" Evolutionary rate matrix:")
cat("\n")
fit2$sigma
cat(" Estimated ancestral states:")
cat("\n")
fit2$theta

fit3_size <- mvOU(tre, data = carni70$tab[,1], model="OU1")
fit3_range <- mvOU(tre, data = carni70$tab[,2], model="OU1")

cat("Alpha: strength of selection - low alpha favours a random change during the time step \n")
fit3_size$alpha
fit3_range$alpha
cat("Sigma: evolutionary rate matrix \n")
fit3_size$sigma
fit3_range$sigma
cat("Theta: estimated ancestral states \n")
fit3_size$theta
fit3_range$theta
phyltree = ape2ouch(tre)
fit5 <- ouchModel(phyltree = phyltree, data = carni70$tab)

cat("A: strength of selection")
cat("\n")
fit5$FinalFound$ParamsInModel$A
cat("vY0: estimated ancestral state for range (OU part)")
cat("\n")
fit5$FinalFound$ParamsInModel$vY0
cat("Syy is the sigma value for range")
cat("\n")
fit5$FinalFound$ParamsInModel$Syy


traits <-  data.frame(range=carni70$tab[,2],size=carni70$tab[,1])
row.names(traits) <- row.names(carni70$tab)


fit6 <-mvslouchModel(phyltree, traits,kY=1)
#
cat("A: strength of selection")
cat("\n")
fit6$FinalFound$ParamsInModel$A
cat("vY0: estimated ancestral state for range (OU part)")
cat("\n")
fit6$FinalFound$ParamsInModel$vY0
```

```
cat("vX0: estimated ancestral state for size (BM part)")
cat("\n")
fit6$FinalFound$ParamsInModel$vX0
cat("Syy is the sigma value for range")
cat("\n")
fit6$FinalFound$ParamsInModel$Syy
cat("Sxx is the sigma value for size")
cat("\n")
fit6$FinalFound$ParamsInModel$Sxx
cat("Comparison of the AICc values:\n")
cat("Both traits evolve as independent Brownian motions \n")
fit1_size$AICc + fit1_range$AICc
cat("The traits evolve as a correlated Brownian motion \n")
fit2$AICc
cat("Both traits evolve as independent Ornstein-Uhlenbeck processes \n")
fit3_size$AICc+fit3_range$AICc
cat("The traits evolve as a bivariate Ornstein-Uhlenbeck process \n")
fit5$`FinalFound`$ParamSummary$aic.c
cat("size evolves as a Brownian motion and range as an Ornstein-Uhlenbeck process \n")
fit6$`FinalFound`$ParamSummary$aic.c
```