

Lab 4: Models of Continuous Trait Evolution

Comparative Biology and Macroevolution

April 26, 2019

This week, we will continue going over models of continuous trait evolution. In lecture, you have learned about Brownian Motion. Now, we will be introducing you to different models of continuous trait evolution and will be implementing them in lab.

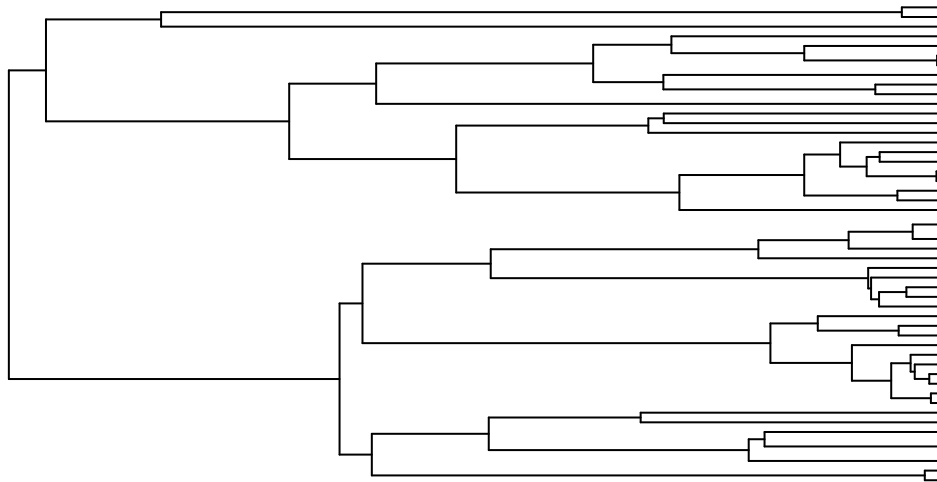
Classwork

Load libraries

```
library(geiger)
library(phyloilm)
library(phytools)
```

Import tree

```
tree = read.tree("lab4_part1_tree.tre")
plot(tree, show.tip.label = F)
```



Simulate data under several different models

```
# Simulate traits under a Brownian Motion Model
data1 <- rTrait(phy = tree, model = "BM",
               parameters = list(ancestral.state = 0, sigma2 = 0.1))
# Simulate traits under an Ornstein-Uhlenbeck Model
data2 <- rTrait(phy = tree, model = "OU",
               parameters = list(ancestral.state = 0,
                                sigma2 = 0.1, optimal.value = 10, alpha = 0.001))
# Simulate traits under an Early Burst Model
data3 <- rTrait(phy = tree, model = "EB",
               parameters = list(ancestral.state = 0, sigma2 = 0.1, rate = -0.0001))
# Simulate traits under a White Noise Model
data4 <- rnorm(n = length(tree$tip.label), mean = 0, sd = exp(1))
names(data4) <- tree$tip.label # Need to give names to the data4 vector
```

```
# rTrait() already does this for us
```

```
# To visualize the data
```

```
cbind(data1, data2, data3, data4)
```

##	data1	data2	data3	data4
## s1	0.6133503	0.620629825	0.65047289	-1.7522602
## s2	1.1448251	0.160470241	1.18424259	0.2437767
## s3	-3.0194143	-0.034905598	0.75141966	-1.5072039
## s4	-2.1942749	0.724540631	0.62976812	-0.8550049
## s5	-1.9981378	0.682709534	0.56109232	0.4565813
## s6	-2.5474880	-1.098587447	1.57297336	1.1391579
## s7	1.2416960	-0.888292921	-0.30915078	-3.1766362
## s8	-0.5062489	0.294593884	-1.46983939	2.5253283
## s9	-2.4585399	-1.849254816	-0.17473783	-2.7409100
## s10	-2.4257860	-1.830945767	-0.52287048	3.5590622
## s11	-2.5646569	-0.954107900	-0.23189489	1.3352931
## s12	-2.1074976	-0.683951589	-0.51704508	4.2977677
## s13	-2.1391872	-1.570448339	-0.08557540	1.5216900
## s14	-1.8917806	-1.388974505	0.08161937	-0.9546403
## s15	-2.0575664	-1.703449232	-1.33145710	1.1871513
## s16	-2.1599200	-2.062282786	-2.41116347	-0.8755641
## s17	-2.0969096	-2.756342182	-2.19063312	0.4679118
## s18	-4.2396267	-1.433067421	-1.50220587	0.9417776
## s19	-2.5589600	-1.507942325	-0.71839508	3.7887092
## s20	-2.7071435	-1.183790580	-0.28264966	2.5907586
## s21	-2.8573939	-0.685975449	-0.27435716	1.4483117
## s22	-2.5000493	-1.009934903	-0.31781927	1.1144189
## s23	-2.0784886	-0.838644089	0.17346099	-0.4015427
## s24	0.8275290	-0.992113464	-2.20964617	-3.1334819
## s25	0.1012660	-0.227332704	-1.29064925	-2.1182504
## s26	1.0984814	0.538294867	-1.56761450	-0.9588537
## s27	1.3489715	0.019174036	-1.87975648	0.1732785
## s28	1.4090826	0.001164438	-1.80194670	-2.3234087
## s29	-0.6211684	0.099946887	-0.59272191	-0.2671609
## s30	0.2175421	0.378376019	0.94693873	0.5566756
## s31	0.1281412	0.715288003	0.76095618	-0.9053524
## s32	1.1999366	0.566762496	-0.54736847	0.6590633
## s33	1.2327950	0.689015729	-0.70116780	-2.3159206
## s34	0.4087199	0.268464114	-0.55379858	-3.4611231
## s35	0.7727631	0.360080785	-0.42857113	-1.2142284
## s36	0.5340954	-0.408501422	0.04947743	2.9606752
## s37	-2.2805158	1.768048491	0.41539792	3.1874690
## s38	-1.9340246	0.658115090	0.90176148	-2.9882666
## s39	-2.7326752	-0.053162276	-0.48523331	5.3039745
## s40	-1.8426755	2.246001096	-3.00091577	-2.9835945
## s41	-1.4047809	0.049523852	0.06038423	-3.0605724
## s42	-0.9578309	-0.757488732	-1.16122339	4.1285989
## s43	-1.9638090	1.552113254	-0.81050878	1.1777321
## s44	-1.4685246	0.780019804	-0.72829931	-4.3107319
## s45	-1.3234507	0.639034718	-0.92915933	0.1388280
## s46	-1.6688103	0.595224262	-0.44144118	5.8562723
## s47	-0.1107365	0.815649076	0.55114392	-3.8928408
## s48	-1.7853285	-0.117528925	-0.02147996	0.3537489

```
## s49  1.5013819  1.468353938 -2.08428088  1.0327908
## s50  1.4065997  0.871361478 -2.74787996  1.0710665
```

Model Fitting

Now we are going to fit models to the simulated datasets.

The format of the code

- `data1.bm` is the fit of the Brownian Motion model to `data1`
- `fitContinuous` is a function in *geiger* that fits various models of continuous character evolution in a likelihood framework.

The `fitContinuous` function takes multiple arguments. To get a full list of them, running `?fitContinuous` will bring up the help file. In this lab we will be passing the following arguments to the function:

- `phy` a phylogenetic tree
- `dat` a data vector for a single trait, where the names of the vector elements need to match the tips in the phylogeny or `phy`.
- `model` the model we want to fit, e.g. "BM", "OU", "EB", "white". Remember it needs to be a vector.

```
# Fit models to data1 (BM)
data1.bm <- fitContinuous(phy = tree, dat = data1, model = "BM")
data1.ou <- fitContinuous(phy = tree, dat = data1, model = "OU")
data1.eb <- fitContinuous(phy = tree, dat = data1, model = "EB")
data1.wn <- fitContinuous(phy = tree, dat = data1, model = "white")

# Fit models to data2 (OU)
data2.bm <- fitContinuous(phy = tree, dat = data2, model = "BM")
data2.ou <- fitContinuous(phy = tree, dat = data2, model = "OU")
data2.eb <- fitContinuous(phy = tree, dat = data2, model = "EB")
data2.wn <- fitContinuous(phy = tree, dat = data2, model = "white")

# Fit models to data3 (EB)
data3.bm <- fitContinuous(phy = tree, dat = data3, model = "BM")
data3.ou <- fitContinuous(phy = tree, dat = data3, model = "OU")
data3.eb <- fitContinuous(phy = tree, dat = data3, model = "EB")
data3.wn <- fitContinuous(phy = tree, dat = data3, model = "white")

# Fit models to data4 (WN)
data4.bm <- fitContinuous(phy = tree, dat = data4, model = "BM")
data4.ou <- fitContinuous(phy = tree, dat = data4, model = "OU")
data4.eb <- fitContinuous(phy = tree, dat = data4, model = "EB")
data4.wn <- fitContinuous(phy = tree, dat = data4, model = "white")
```

Cool! Now lets look inside one of these model fits

```
data1.bm

## GEIGER-fitted comparative model of continuous data
## fitted 'BM' model parameters:
## sigsq = 0.117726
## z0 = -0.866101
##
## model summary:
## log-likelihood = -53.693569
## AIC = 111.387139
```

```
## AICc = 111.642458
## free parameters = 2
##
## Convergence diagnostics:
## optimization iterations = 100
## failed iterations = 0
## frequency of best fit = 1.00
##
## object summary:
## 'lik' -- likelihood function
## 'bnd' -- bounds for likelihood search
## 'res' -- optimization iteration summary
## 'opt' -- maximum likelihood parameter estimates
```

Question: How many parameters are in the ‘BM’ model?

Compare results using AIC

Recall:

Let k be the number of estimated parameters in the model. Let \hat{L} be the maximum value of the likelihood function for the model. The AIC value of that model is

$$AIC = 2k - 2\ln(\hat{L})$$

```
BM.Data <- c(data1.bm$opt$aic , data1.ou$opt$aic, data1.eb$opt$aic, data1.wn$opt$aic)
OU.Data <- c(data2.bm$opt$aic , data2.ou$opt$aic, data2.eb$opt$aic, data2.wn$opt$aic)
EB.Data <- c(data3.bm$opt$aic , data3.ou$opt$aic, data3.eb$opt$aic, data3.wn$opt$aic)
WN.Data <- c(data4.bm$opt$aic , data4.ou$opt$aic, data4.eb$opt$aic, data4.wn$opt$aic)
AICresults <- rbind(BM.Data, OU.Data, EB.Data, WN.Data)
colnames(AICresults) <- c("Brownian Motion", "Ornstein-Uhlenbeck", "Early Burst", "White noise")
AICresults
```

```
##           Brownian Motion Ornstein-Uhlenbeck Early Burst White noise
## BM.Data      111.38714          113.3871    113.02406    189.5057
## OU.Data       96.78119          96.2807     98.78124    152.4305
## EB.Data      103.70755         105.0186    105.70757    149.9463
## WN.Data      309.50862         238.9292    311.50878    235.2841
```

Introducing Disparity Through Time

One way to understand the time course of morphological diversification is through a *disparity-through-time* (DTT) plot.

There are multiple ways to calculate disparity. The most popular approach is to use the average pairwise Euclidean distances between species. This measure relates to the variance and estimates the dispersion of the points in space (Harmon et al. 2003)

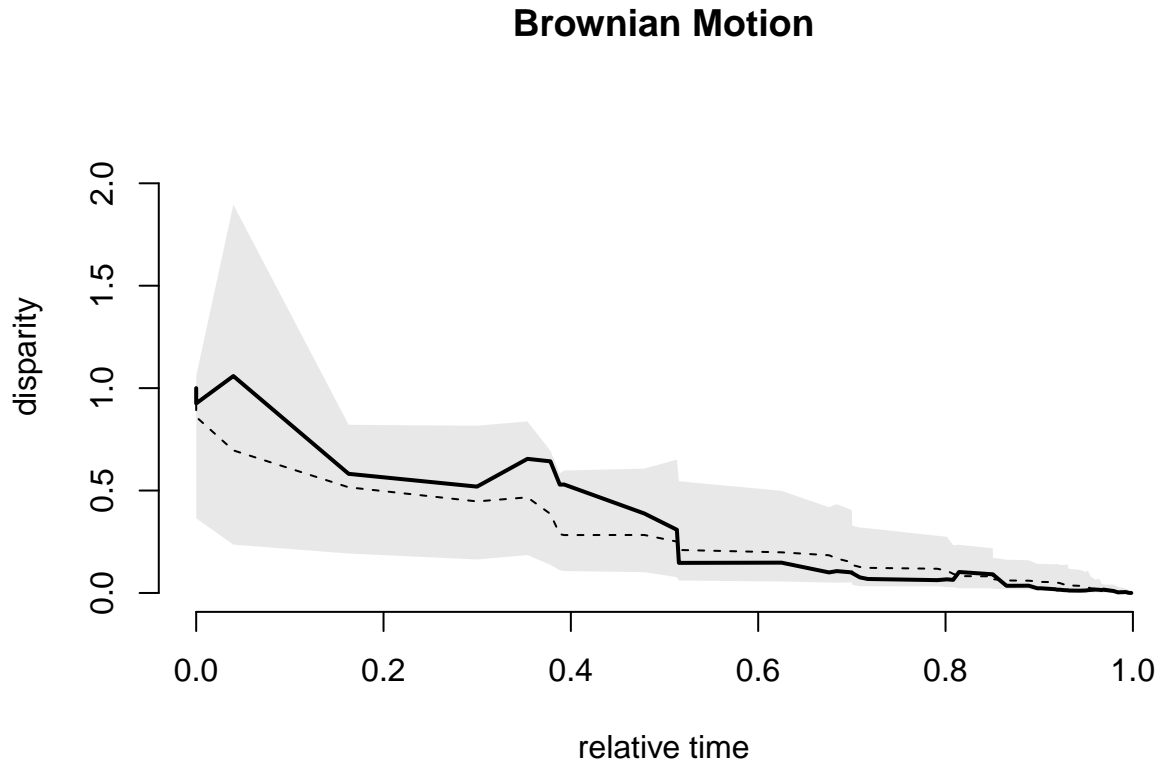
To generate a dtt plot, we first calculate disparity for the entire clade. Then we can calculate disparity for each subclade in the phylogeny. Finally, we can calculate relative disparities for each subclade by dividing the subclade’s disparity by the disparity of the entire clade.

Finally, we can calculate the mean relative disparity for each node as the average of all the relative disparities of all the subclades whose ancestral lineages were present at the time.

Small values (near 0) indicate there is little variation within the subclades compared to the the group as a whole. This can indicate that the subclades have undergone specialization.

Large values (near 1) indicate there is a large proportion of the total variation is within the subclades. This can indicate that subclades have undergone convergence.

```
ddt1 = dtt(phy = tree, data = data1, nsim = 100, index = c("avg.sq"), plot = TRUE)
title("Brownian Motion")
```



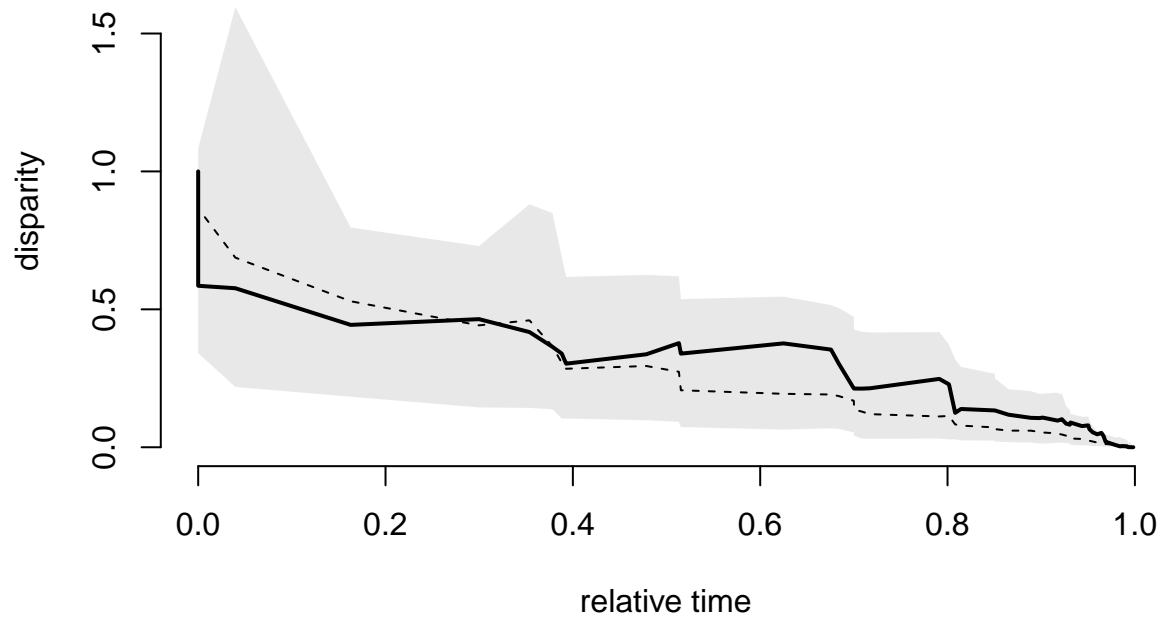
```
ddt2 = dtt(phy = tree, data = data2, nsim = 100, index = c("avg.sq"), plot = TRUE)
title("Ornstein-Uhlenbeck")
```

```
## Warning in title("Ornstein-Uhlenbeck"): conversion failure on 'Ornstein-
## Uhlenbeck' in 'mbsToSbcs': dot substituted for <e2>
```

```
## Warning in title("Ornstein-Uhlenbeck"): conversion failure on 'Ornstein-
## Uhlenbeck' in 'mbsToSbcs': dot substituted for <80>
```

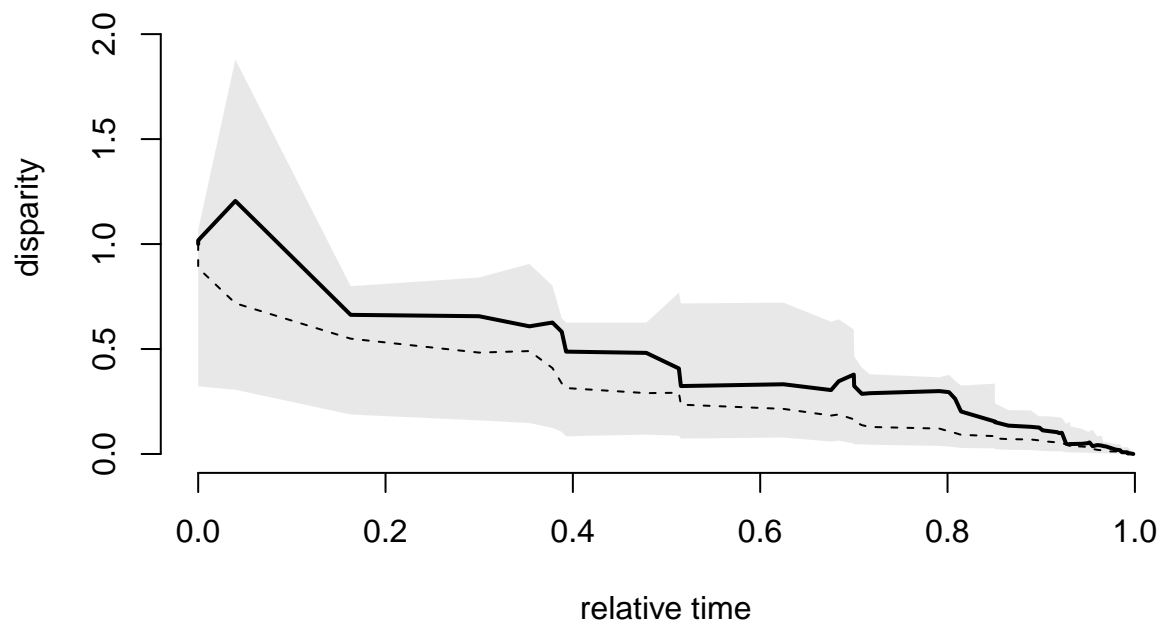
```
## Warning in title("Ornstein-Uhlenbeck"): conversion failure on 'Ornstein-
## Uhlenbeck' in 'mbsToSbcs': dot substituted for <93>
```

Ornstein...Uhlenbeck



```
ddt3 = dtt(phy = tree, data = data3, nsim = 100, index = c("avg.sq"), plot = TRUE)
title("Early Burst")
```

Early Burst



Inferring Mode of Trait Evolution

Now try to infer the mode of trait evolution on an unknown dataset

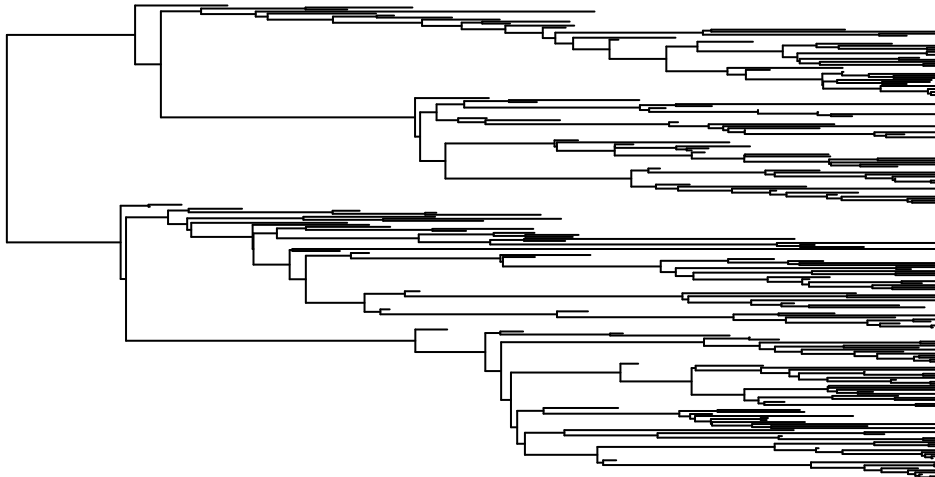
```
lab4_tree <- read.tree("lab4_tree.tre")
lab4_data <- read.csv("lab4_trait_data.csv", stringsAsFactors = F)
head(lab4_data)
```

```
##      X trait_value
## 1 s127   -2.346941
## 2 s128   -2.386864
## 3 s126   -2.326470
## 4 s129   -2.388348
## 5 s130   -2.551628
## 6 s131   -1.645700
```

```
lab4_tree
```

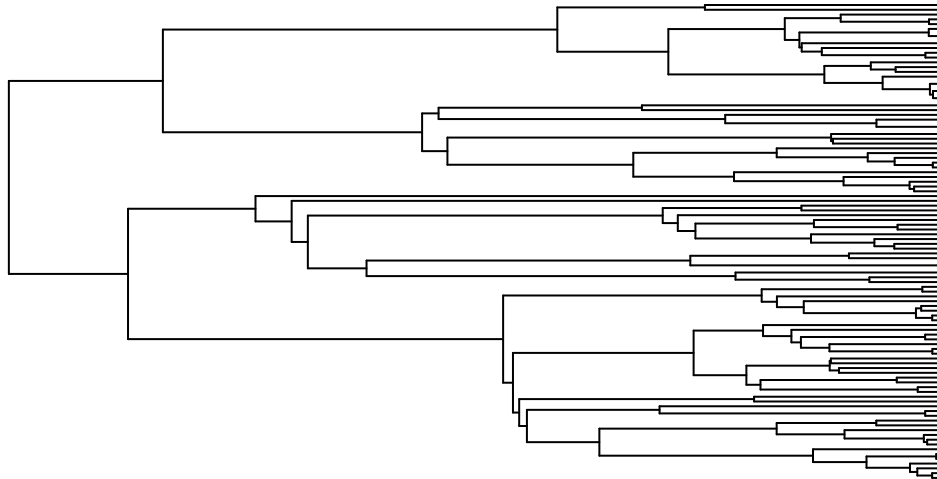
```
##
## Phylogenetic tree with 236 tips and 235 internal nodes.
##
## Tip labels:
## s127, s128, s126, s129, s130, s131, ...
##
## Rooted; includes branch lengths.
```

```
plot(lab4_tree, show.tip.label = F)
```



Question: What is interesting about this tree?

```
# Remove the extinct tips
extant_only_lab4_tree <- drop.extinct(lab4_tree)
plot(extant_only_lab4_tree, show.tip.label = F)
```



```
# Pull out the trait value column and name it
trait_values <- lab4_data$trait_value
names(trait_values) <- lab4_data$X
head(trait_values)
```

```
##      s127      s128      s126      s129      s130      s131
## -2.346941 -2.386864 -2.326470 -2.388348 -2.551628 -1.645700
```

```
# Extract the trait data for only the living species
extant_only_trait_values <- trait_values[intersect(
  names(trait_values), extant_only_lab4_tree$tip.label)]
```

```
# Fit models to extant only tree
extant.bm <- fitContinuous(phy = extant_only_lab4_tree,
                           dat = extant_only_trait_values, model = "BM")
extant.ou <- fitContinuous(phy = extant_only_lab4_tree,
                           dat = extant_only_trait_values, model = "OU")
extant.eb <- fitContinuous(phy = extant_only_lab4_tree,
                           dat = extant_only_trait_values, model = "EB")
extant.wn <- fitContinuous(phy = extant_only_lab4_tree,
                           dat = extant_only_trait_values, model = "white")
```

```
# Fit models to total tree
total.bm <- fitContinuous(phy = lab4_tree, dat = trait_values, model = "BM")
total.ou <- fitContinuous(phy = lab4_tree, dat = trait_values, model = "OU")
total.eb <- fitContinuous(phy = lab4_tree, dat = trait_values, model = "EB")
total.wn <- fitContinuous(phy = lab4_tree, dat = trait_values, model = "white")
```

```
Extant.Data <- c(extant.bm$opt$aic, extant.ou$opt$aic, extant.eb$opt$aic, extant.wn$opt$aic)
Total.Data <- c(total.bm$opt$aic, total.ou$opt$aic, total.eb$opt$aic, total.wn$opt$aic)
AIC_results <- rbind(Extant.Data, Total.Data)
colnames(AIC_results) <- c("Brownian Motion", "Ornstein Uhlenbeck", "Early Burst", "White noise")
AIC_results
```

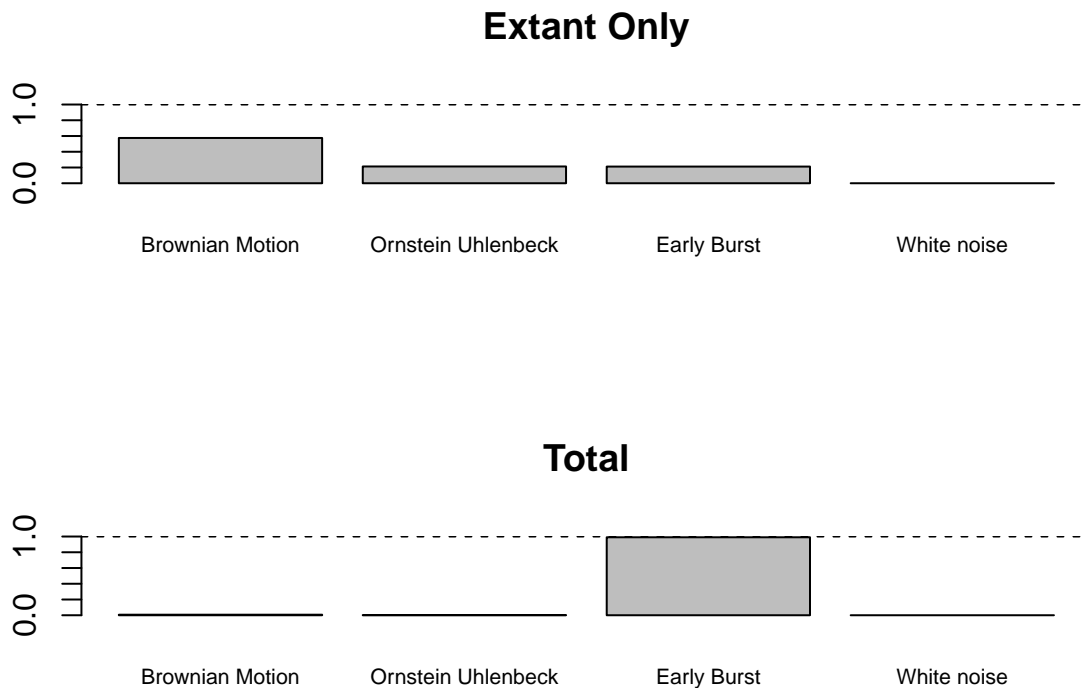
```
##      Brownian Motion Ornstein Uhlenbeck Early Burst White noise
## Extant.Data      147.2491      149.2349      149.2492      313.6844
## Total.Data       338.3724      340.3716      328.4289      722.1610
```

Transform AIC values into Akaike weights, which can be interpreted as conditional probabilities for each model


```
Extant_aicw <- aicw(Extant.Data)
Total_aicw <- aicw(Total.Data)

par(mfrow = c(2,1))
barplot(Extant_aicw[,3],
        names.arg = c("Brownian Motion", "Ornstein Uhlenbeck", "Early Burst", "White noise"),
        ylim = c(0, 1), main = "Extant Only", cex.names = 0.7)
abline(h = 1, lty = "dashed") # Makes a dashed line at height 1 (y = 1)

barplot(Total_aicw[,3],
        names.arg = c("Brownian Motion", "Ornstein Uhlenbeck", "Early Burst", "White noise"),
        ylim = c(0, 1), main = "Total", cex.names = 0.7)
abline(h = 1, lty = "dashed") # Makes a dashed line at height 1 (y = 1)
```



Question: How does incorporating fossil data impact our inference of trait evolution?

Exercises Due Next Week

Did differences in body size and dietary indices observed today evolve when terrestrial Carnivorans adaptively radiated into different niches?

- Fit Brownian Motion, Ornstein-Uhlenbeck, Early Burst, and White Noise models to C1 (Cross sectional shape of the upper canine), RLGA (Relative upper grinding area), and the log cuberoot of body mass.

Hint: Van Valkenburgh and Koepfli 1993, Friscia et al. 2007 and Slater and Friscia 2019 have additional information about the functional significance of the dietary indices.

- Make sure to match data with tree!
- Include the standard error (SE) when fitting macroevolutionary models

Hint:

```
myfit <- fitContinuous(phy = your_tree, dat = your_trait_data, SE = your_trait_SE, models = c("BM"))
```

- Choose best model out of 4 candidate models using AIC for each trait
- Report the AIC or AICw for all of the models in a table
- Provide a biological interpretation of your results.

*How does the **ecology** of this clade help explain the **evolution** of body size and dietary indices. Additionally, what do the models tell you about how evolution in these traits have occurred through time?*

Hint: Is there a functional difference between the upper canines and the relative lower grinding area of the lower molars?

- Remember to use proper citations.