

# Exercise 14: Diversification analysis with RPANDA

June/2016

- RPANDA
  - Model selection
  - Estimation of model parameters
  - Plotting diversity through time
- Challenge

## RPANDA

RPANDA (<http://cran.r-project.org/web/packages/RPANDA/index.html>) is another recently developed R package that allows us to fit different models of rate variation through time and select the best fitting model using maximum likelihood analysis.

RPANDA shows basically two main differences from BAMM:

1. In RPANDA, the user must inform which models are going to be tested, whereas in BAMM the program itself will average among the rates for each clade;
2. In RPANDA, the user must know *a priori* which are the clades that might show a particular diversification regime, while BAMM will estimate where rate shifts are positioned using a rjMCMC algorithm.

The package also contains some simulation functions as well as datasets. You can find more details on CRAN (<http://cran.r-project.org/web/packages/RPANDA/RPANDA.pdf>).

## Model selection

For this exercise, we will test four different scenarios with all combinations of constant and variable speciation and extinction rates as follows (based on Morlon *et al.* 2011, PNAS (<http://www.pnas.org/content/108/39/16327.abstract?tab=ds>))

Read in the whale tree

```
library(ape)
library(RPANDA)
```

```
## Loading required package: picante
```

```
## Loading required package: vegan
```

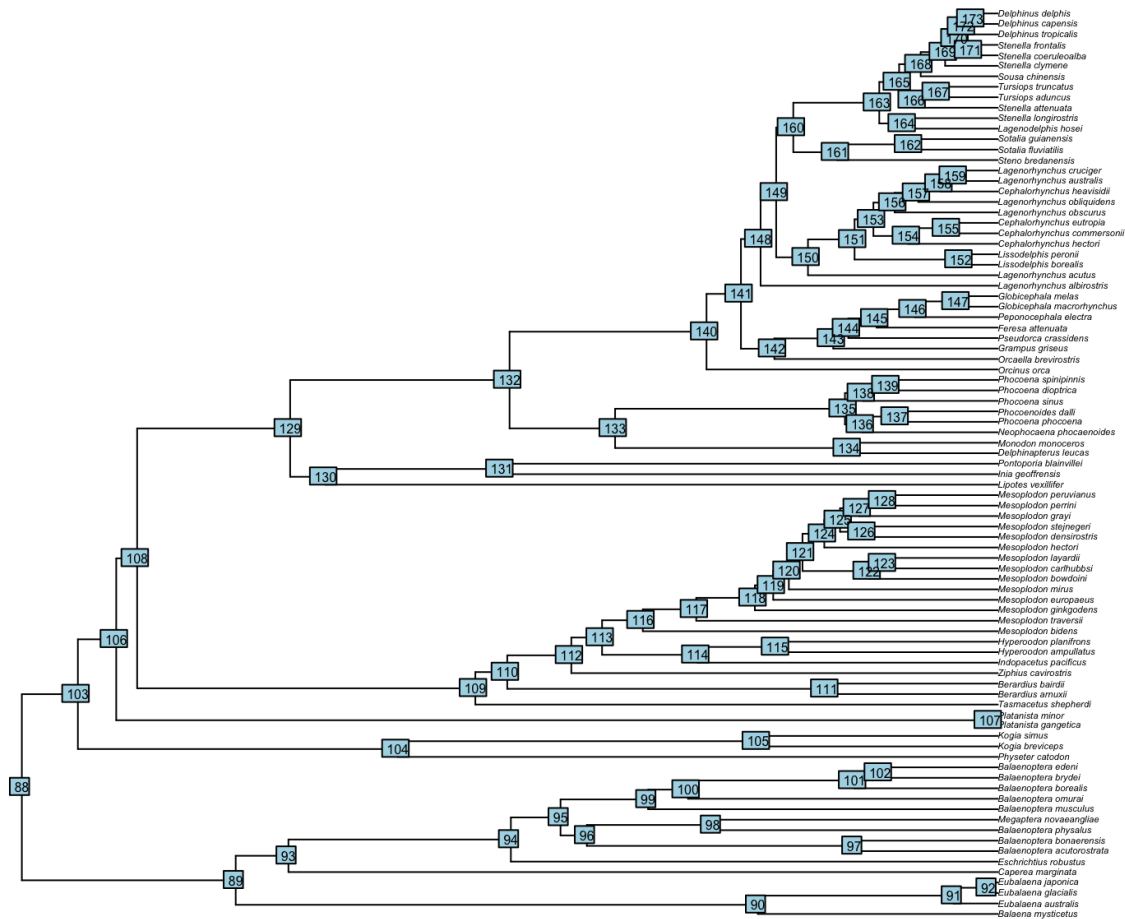
```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.4-0
```

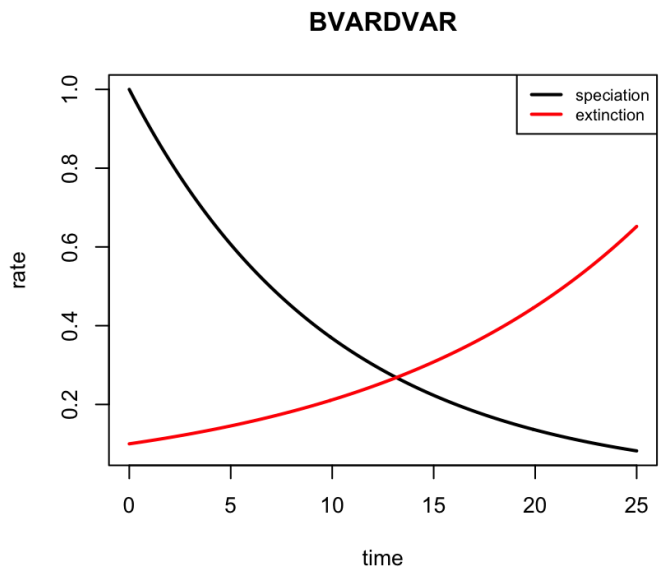
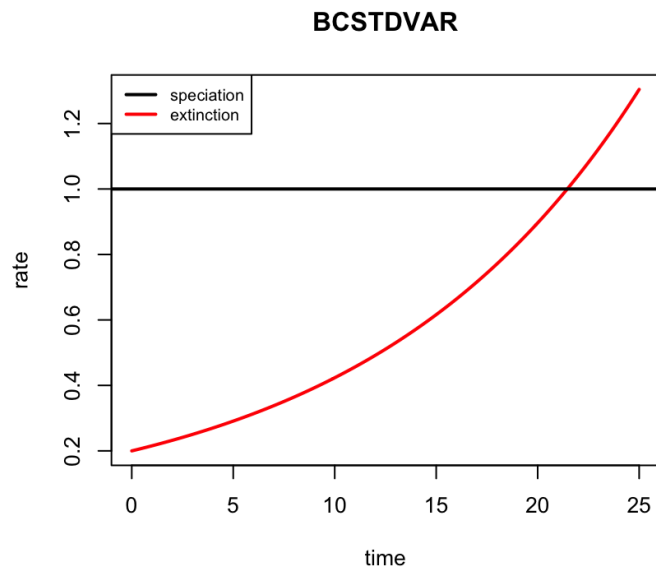
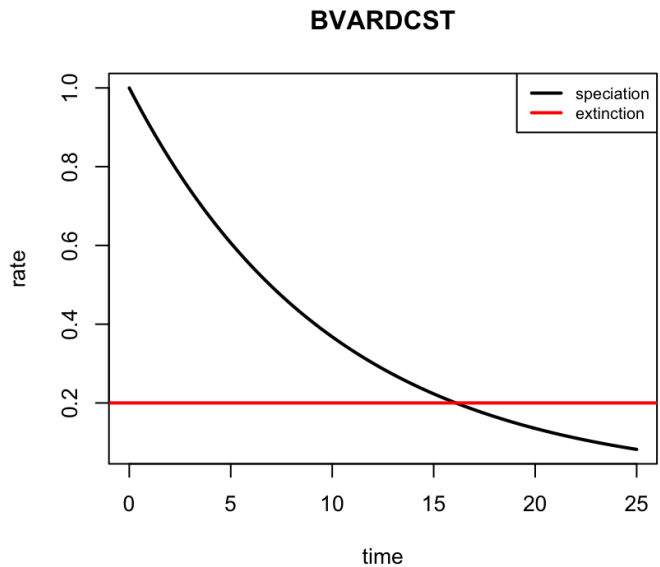
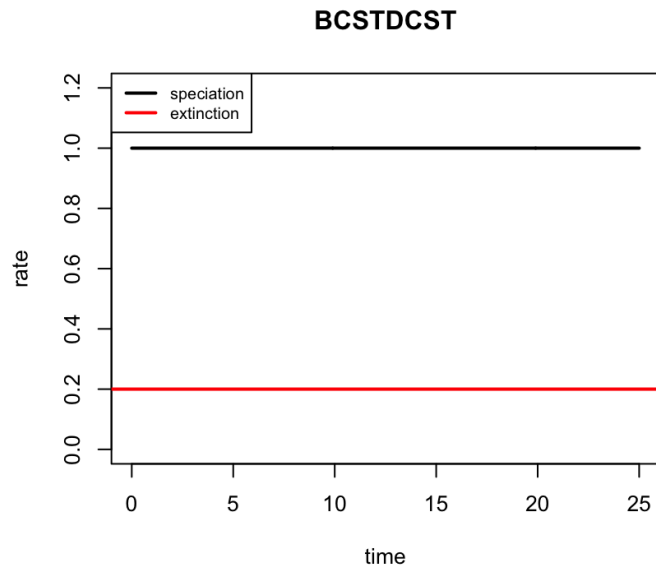
```
## Loading required package: nlme
```

```
whales <- read.tree("whaletree.tre")
plot(whales, cex = 0.35)
nodelabels(cex = 0.5)
```



```
lambda.cst <- function(x,y){y}
lambda.var <- function(x,y){y[1]*exp(y[2]*x)}
mu.cst <- function(x,y){y}
mu.var <- function(x,y){y[1]*exp(y[2]*x)}
```

The rate variations for the four possible pairwise combinations look like this:



Having decided which models one would like to test, the next step is to extract the clades that most likely share similar diversification regimes. As said previously, RPANDA needs the user to *a priori* indicate these clades. In this example, we will use the four main cetacean families plus the rest of cetaceans as a fifth clade.

```
##create a pdf of the tree with the internal nodes labelled
pdf("whale_tree.pdf", width = 4, height =16)
plot(whales, cex = 0.5)
nodelabels(cex = 0.5)
dev.off()
```

```
## quartz_off_screen
##                2
```

*#Use these node numbers to extract clades.*

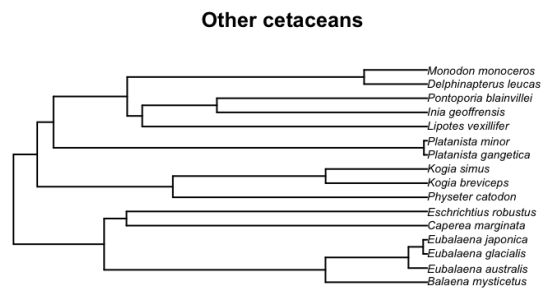
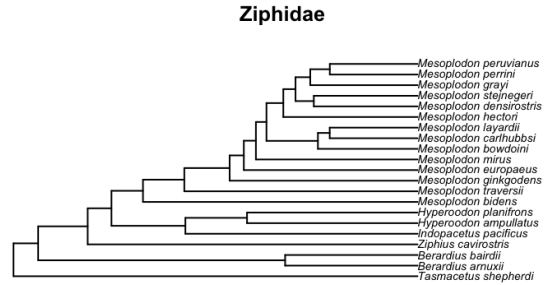
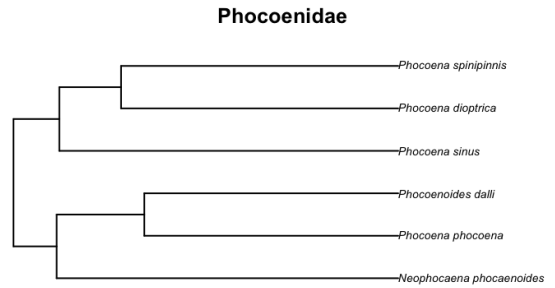
```
balaenidae.tree <- extract.clade(whales, 90)
balaenopteridae.tree <- extract.clade(whales, 95)
delphinidae.tree <- extract.clade(whales, 140)
phocoenidae.tree <- extract.clade(whales, 135)
ziphidae.tree <- extract.clade(whales, 109)
```

```
par(mfcol = c(3, 2))
plot(balaenidae.tree, main = "Balaenidae")
plot(balaenopteridae.tree, main = "Balaenopteridae")
plot(delphinidae.tree, main = "Delphinidae")
plot(phocoenidae.tree, main = "Phocoenidae")
plot(ziphidae.tree, main = "Ziphidae")
```

*#We also need to capture the rest of the cetaceans that do not fall into these clades*

```
othercetaceans.tree <- drop.tip(whales, c(balaenopteridae.tree$tip.label, delphinidae.tree$tip.label,
phocoenidae.tree$tip.label, ziphidae.tree$tip.label))

plot(othercetaceans.tree, main = "Other cetaceans")
```



5/12

```

balaenidae <- c(whales$tip.label[grepl("Balaena",whales$tip.label)],whales$tip.label[grepl("Eubalaena",whales$tip.label)])
balaenopteridae <- c(whales$tip.label[grepl("Balaenoptera",whales$tip.label)],whales$tip.label[grepl("Megaptera",whales$tip.label)])
delphinidae <- c(whales$tip.label[grepl("Delphinus",whales$tip.label)],whales$tip.label[grepl("Cephalorhynchus",whales$tip.label)],whales$tip.label[grepl("Feresa",whales$tip.label)],whales$tip.label[grepl("Globicephala",whales$tip.label)],whales$tip.label[grepl("Lagenodelphis",whales$tip.label)],whales$tip.label[grepl("Lagenorhynchus",whales$tip.label)],whales$tip.label[grepl("Lissodelphis",whales$tip.label)],whales$tip.label[grepl("Orcaella",whales$tip.label)],whales$tip.label[grepl("Orcinus",whales$tip.label)],whales$tip.label[grepl("Peponocephala",whales$tip.label)],whales$tip.label[grepl("Pseudorca",whales$tip.label)],whales$tip.label[grepl("Sotalia",whales$tip.label)],whales$tip.label[grepl("Sousa",whales$tip.label)],whales$tip.label[grepl("Stenella",whales$tip.label)],whales$tip.label[grepl("Steno",whales$tip.label)],whales$tip.label[grepl("Tursiops",whales$tip.label)],whales$tip.label[grepl("Grampus",whales$tip.label)])
phocoenidae <- c(whales$tip.label[grepl("Neophocaena",whales$tip.label)],whales$tip.label[grepl("Phocoenoides",whales$tip.label)])
ziphidae <- c(whales$tip.label[grepl("Berardius",whales$tip.label)],whales$tip.label[grepl("Hyperoodon",whales$tip.label)],whales$tip.label[grepl("Indopacetus",whales$tip.label)],whales$tip.label[grepl("Mesoplodon",whales$tip.label)],whales$tip.label[grepl("Tasmacetus",whales$tip.label)],whales$tip.label[grepl("Ziphius",whales$tip.label)])
balaenidae.tree <- drop.tip(whales,whales$tip.label[-match(balaenidae,whales$tip.label)])

```

```

balaenopteridae.tree <- drop.tip(whales,whales$tip.label[-match(balaenopteridae,whales$tip.label)])
delphinidae.tree <- drop.tip(whales,whales$tip.label[-match(delphinidae,whales$tip.label)])
phocoenidae.tree <- drop.tip(whales,whales$tip.label[-match(phocoenidae,whales$tip.label)])
ziphidae.tree <- drop.tip(whales,whales$tip.label[-match(ziphidae,whales$tip.label)])
othercetaceans.tree <- drop.tip(whales,c(balaenopteridae,delphinidae, phocoenidae, ziphidae))

```

Lets fit our four models to the tree for the phocoenidae.

```

#fit a constant rate bd model, bvarcst
fit_bd(phocoenidae.tree, max(branching.times(phocoenidae.tree)), f.lamb=lambda.cst, f.mu=mu.cst,
  lamb_par=0.4, mu_par=0, cst.lamb=TRUE, cst.mu=TRUE, cond="crown", f=87/89, dt=1e-3)

```

```

## model :
## [1] "birth death"
##
## LH :
## [1] -11.89603
##
## aicc :
## [1] 31.79207
##
## lamb_par :
## [1] 0.1410234
##
## mu_par :
## [1] -6.012454e-08

```

```
#how does this compare to the bd fit under ape from this AM?
bd<-function(x){
  if(class(x)!="birthdeath") stop("x should be an object of class 'birthdeath'")
  b<-x$para[2]/(1-x$para[1])
  d<-b-x$para[2]
  setNames(c(b,d),c("b","d"))
}

fit.bd<-birthdeath(phocoenidae.tree)
bd(fit.bd)
```

```
##           b           d
## 0.1383232 0.0000000
```

```
# diversitree fit
# library(diversitree)
# pbModel<-make.bd(phocoenidae.tree,sampling.f=87/89)
# bdMLFit<-find.mle(bdModel,c(0.1,0.05),method = "optim",lower = 0)
# bdMLFit

##fit a model where the speciation rate varies through time and the extinction rate is constant,
bvardcst
fit_bd(phocoenidae.tree, max(branching.times(phocoenidae.tree)), f.lamb=lambda.var, f.mu=mu.cst,
  lamb_par= c(0.4,-0.05),mu_par=0,expo.lamb=TRUE,cst.mu=TRUE,cond="crown",f=87/89,dt=1e-3)
```

```
## model :
## [1] "birth death"
##
## LH :
## [1] -8.066742
##
## aicc :
## [1] 34.13348
##
## lamb_par :
## [1] 0.002377951 1.140276500
##
## mu_par :
## [1] -1.58053e-07
```

This function will help automate the calculation of the likelihood for the RPANDA models.

```

library(RPANDA)
fit.multi.rpanda <- function(tree,par)
{
  bcstdcst <- fit_bd(tree, max(branching.times(tree)), f.lamb=lambda.cst, f.mu=mu.cst, lam
b_par=par[[1]][1],mu_par=par[[1]][2],cst.lamb=TRUE,cst.mu=TRUE,cond="crown",f=87/89,dt=1e-3)
  bvardcst <- fit_bd(tree, max(branching.times(tree)), f.lamb=lambda.var, f.mu=mu.cst, lam
b_par=par[[2]][c(1,2)],mu_par=par[[2]][3],expo.lamb=TRUE,cst.mu=TRUE,cond="crown",f=87/89,dt=1e-
3)
  bcstdvar <- fit_bd(tree, max(branching.times(tree)), f.lamb=lambda.cst, f.mu=mu.var, lam
b_par=par[[3]][1],mu_par=par[[3]][c(2,3)],cst.lamb=TRUE,expo.mu=TRUE,cond="crown",f=87/89,dt=1e-
3)
  bvardvar <- fit_bd(tree, max(branching.times(tree)), f.lamb=lambda.var, f.mu=mu.var, lam
b_par=par[[4]][c(1,2)],mu_par=par[[4]]
[c(3,4)],expo.lamb=TRUE,expo.mu=TRUE,cond="crown",f=87/89,dt=1e-3)
  return(list("bcstdcst"=bcstdcst,"bvardcst"=bvardcst,"bcstdvar"=bcstdvar,"bvardvar"=bvard
var))
}
whales.par <- list(c(0.4,0),c(0.4,-0.05,0),c(0.4,0.1,0.05),c(0.4,-0.05,0.1,0.05)) #we need to su
pply starting parameter values for optimization to RPANDA

```

## Estimation of model parameters

In the possession of all models and clades, we can finally estimate the parameters of all the four models to each of the five clades and create an AICc table in order to select which one is the best model that describes the changes in both diversification rates (speciation and extinction) rates. (This part of the code take a while to complete)

```

results <- list("balaenopteridae.res"=fit.multi.rpanda(balaenopteridae.tree,whales.par),
  "delphinidae.res" = fit.multi.rpanda(delphinidae.tree,whales.par),
  "phocoenidae.res" = fit.multi.rpanda(phocoenidae.tree,whales.par),
  "ziphidae.res" = fit.multi.rpanda(ziphidae.tree,whales.par),
  "othercetaceans.res"= fit.multi.rpanda(othercetaceans.tree,whales.par))

```

```

aic.table <- matrix(nrow=4,ncol=5,NA)
for(i in 1:5)
{
  for(j in 1:4)
  {
    aic.table[j,i] <- results[[i]][[j]]$aicc
  }
}
colnames(aic.table) <- c("Balaenopteridae","Delphinidae","Phocoenidae","Ziphidae","Other Cetacea
ns")
rownames(aic.table) <- c("bcstdcst","bvardcst","bcstdvar","bvardvar")
aic.table

```

##	Balaenopteridae	Delphinidae	Phocoenidae	Ziphidae	Other Cetaceans
## bcstdcst	56.73336	171.4150	31.79207	132.6440	119.4690
## bvardcst	58.44456	170.0338	34.13348	130.1843	117.5562
## bcstdvar	58.32515	170.6453	41.79207	127.6094	115.3723
## bvardvar	65.67244	171.3258	64.13348	133.2251	118.9789



```
par.table <- data.frame("Balaenopteridae"=c(results[[1]]$bcstdcst$lamb_par[1:2],results[[1]]$bcstdcst$mu_par[1:2]),"Delphinidae"=c(results[[2]]$bvardcst$lamb_par[1:2],results[[2]]$bvardcst$mu_par[1:2]),"Phocoenidae"=c(results[[3]]$bcstdcst$lamb_par[1:2],results[[3]]$bcstdcst$mu_par[1:2]),"Ziphiidae"=c(results[[4]]$bcstdcst$lamb_par[1:2],results[[4]]$bcstdcst$mu_par[1:2]),"Other Cetaceans"=c(results[[5]]$bcstdvar$lamb_par[1:2],results[[5]]$bcstdvar$mu_par[1:2]))
par.table
```

##	Balaenopteridae	Delphinidae	Phocoenidae	Ziphiidae	Other.Cetaceans
## 1	7.343080e-02	1.405102e-01	1.410234e-01	9.485149e-02	0.1857764
## 2	NA	1.257642e-01	NA	NA	NA
## 3	-6.391602e-10	-1.521115e-07	-6.012454e-08	7.545982e-08	0.8313274
## 4	NA	NA	NA	NA	-0.1742352

## Plotting diversity through time

After selecting which model best fits the trees, we can estimate the diversity trajectory through time for each of the five clades.

```

# Function to calculate species richness in a given point in time
div.times <-
c(max(branching.times(balaenopteridae.tree)),max(branching.times(delphinidae.tree)),max(branchin
g.times(phocoenidae.tree)),max(branching.times(ziphidae.tree)),max(branching.times(othercetacean
s.tree)))

# Function modified from plot_dtt from RPANDA package
plotdtt <- function (fit.bd, tot_time, N0, col=1, add=FALSE, div.time, xlim, ylim)
{
  if (!inherits(fit.bd, "fit.bd"))
    stop("object \"fit.bd\" is not of class \"fit.bd\"")
  t <- seq(tot_time-div.time, tot_time, 0.01)
  if ("f.mu" %in% attributes(fit.bd)$names) {
    r <- function(t) {
      -fit.bd$f.lamb(t) + fit.bd$f.mu(t)
    }
    R <- function(s) {
      RPANDA:::.Integrate(Vectorize(r), 0, s)
    }
    N <- N0 * exp(Vectorize(R)(t))
    #dev.new()

    if(add==FALSE)
    {
      plot(-t, N, type = "l", xlab = "time", ylab = "Number of species",
        main = "Diversity Through Time", col=col, xlim=xlim, ylim=ylim)
    }
    else
    {
      lines(-t, N, type = "l", xlab = "time", ylab = "Number of species",
        main = "Diversity Through Time", col=col, xlim=xlim, ylim=ylim)
    }
  }
  else {
    r <- function(t) {
      -fit.bd$f.lamb(t)
    }
    R <- function(s) {
      RPANDA:::.Integrate(Vectorize(r), 0, s)
    }
    N <- N0 * exp(Vectorize(R)(t))
    #dev.new()

    if(add==FALSE)
    {
      plot(-t, N, type = "l", xlab = "time", ylab = "Number of species",
        main = "Diversity Through Time",col=col, xlim=xlim, ylim=ylim)
    }
    else
    {
      lines(-t, N, type = "l", xlab = "time", ylab = "Number of species",
        main = "Diversity Through Time",col=col, xlim=xlim, ylim=ylim)
    }
  }
}

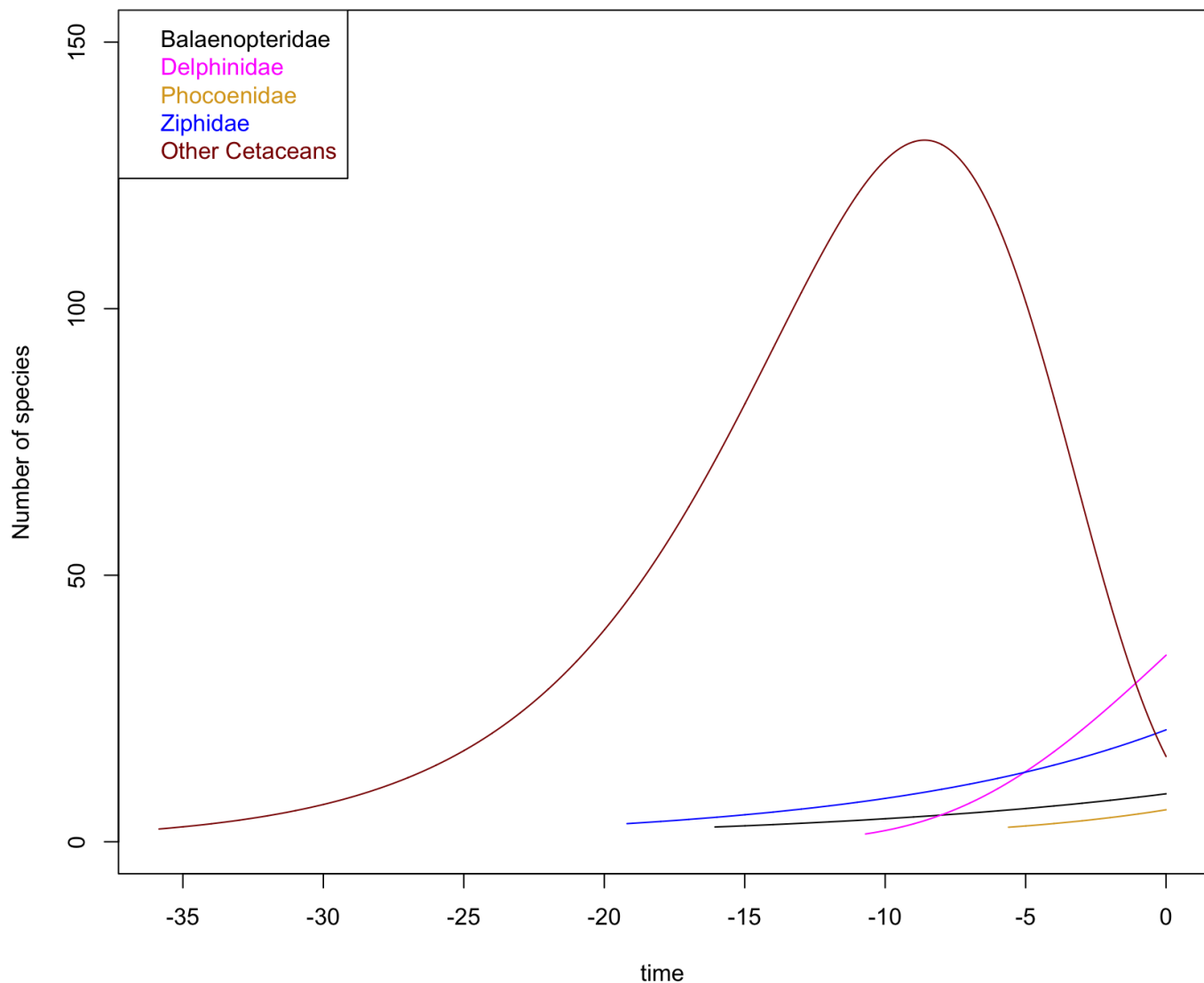
```

```

plotdtt(results$balaenopteridae$bcstdcst,div.times[1],N0=Ntip(balaenopteridae.tree),xlim=c(-
max(div.times),0),ylim=c(0,150),div.time=div.times[1])
plotdtt(results$delphinidae$bvardcst,div.times[2],N0=Ntip(delphinidae.tree),col=6,add=TRUE,xlim=c(
max(div.times),0),ylim=c(0,150),div.time=div.times[2])
plotdtt(results$phocoenidae$bcstdcst,div.times[3],N0=Ntip(phocoenidae.tree),col="goldenrod",add=T
RUE,xlim=c(-max(div.times),0),ylim=c(0,150),div.time=div.times[3])
plotdtt(results$ziphidae$bcstdcst,div.times[4],N0=Ntip(ziphidae.tree),col=4,add=TRUE,xlim=c(-
max(div.times),0),ylim=c(0,150),div.time=div.times[4])
plotdtt(results$othercetaceans$bcstdvar,div.times[5],N0=Ntip(othercetaceans.tree),col="darkred",a
dd=TRUE,xlim=c(-max(div.times),0),ylim=c(0,150),div.time=div.times[5])
legend("topleft",legend=c("Balaenopteridae","Delphinidae","Phocoenidae","Ziphipidae","Other Cetace
ans"),text.col=c(1,6,"goldenrod",4,"darkred"))

```

Diversity Through Time



## Challenge

1. Fit models using RPANDA for the **entire** whale phylogeny, and plot its diversity through time using parts of the previous codes chunks.