

An empirical example: *Rhynoclemmys* data.

Daniel R. Miranda-Esquivel

2020 - 10 - 18

We read the data, distribution and tree.

```
## for reproductibility purposes

set.seed(121)

options(warn=-1)
suppressMessages(library(ggtree))
#~ library(ape)
suppressMessages(library(blepd))
library(ggplot2)
#~ library(ggtree)

## Version

cat("Analyses made with blepd version:",unlist(packageVersion("blepd")))

## Analyses made with blepd version: 0 1 8 2020 10 21
## Create an object to place the distribution and the tree

RhynoclemmysData <- list()

## Read data

## distribution is a csv file areas x terminals, labeled

setwd("../data/")

csvFile <- dir(pattern=".csv")

RhynoclemmysData$distribution <- as.matrix(read.table(csvFile,
  stringsAsFactors=TRUE,
  header=TRUE,
  row.names=1,
  sep=",")
)
```

```

##print(t(RhinoclemmysData$distribution))

## tree(s) in nexus or newick format
treeFiles <- dir(pattern=".tre")

##treeFiles

RhinoclemmysData$tree <- read.tree(treeFiles)

## name of tree(s)
treeFiles <- gsub(".tre","",treeFiles)

## Plotting

#par(mfrow=c(2,1))

## The tree

## using ggtree
plotTree <- ggtree(RhinoclemmysData$tree, ladderize=TRUE,
                   color="black", size=0.51, linetype="solid") +
  geom_tiplab(size=4, color="black") +
  xlim(0,0.035) +
  theme_tree2() +
  ggtitle(treeFiles[1])

##print(plotTree)

#~ Alternatively, we can plot the trees using APE
##
## plot.phylo(RhinoclemmysData$tree)
##
####nodelabels()
####tiplabels()

```

```

## the distribution

distXY <- matrix2XY(RhinoclemmysData$distribution)

terminals <- colnames(RhinoclemmysData$distribution)

realOrder <- match(terminals, RhinoclemmysData$tree$tip.label)

equivalencias <- data.frame(terminals, realOrder)

dXY2 <-      distXY

for(cambiar in terminals){

  dXY2$Terminal[distXY$Terminal == cambiar]      <-
  equivalencias$realOrder[equivalencias$terminals==cambiar]

}

distGraficar <- dXY2

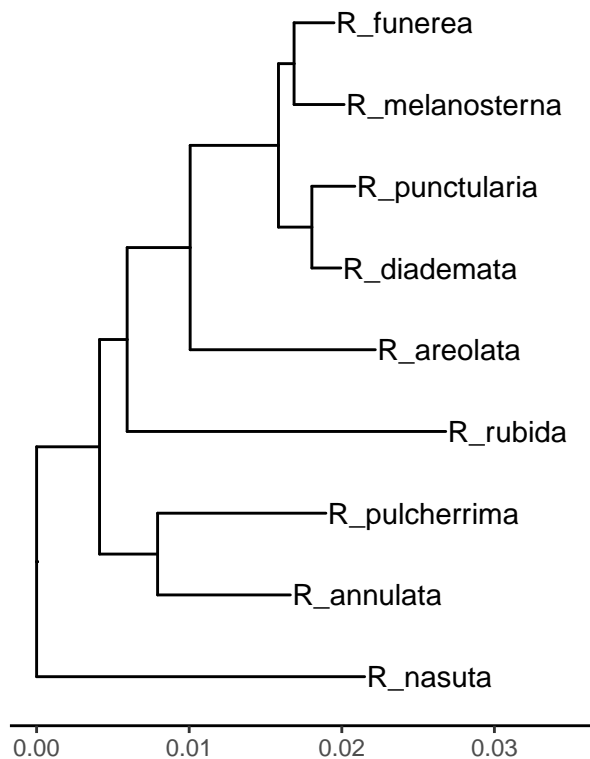
plotDistrib <- ggplot(distGraficar,
  aes(x= Area, y=Terminal), size =35) +
  geom_point(shape=19, fill="white", color="darkred", size=7) +
  labs(title = "Distributions",
    y = "",
    x = "Area") +
  theme(axis.line=element_blank(),
    axis.text.y=element_blank(),
    axis.ticks=element_blank(),
    axis.title.y=element_blank(),
    legend.position="none",
    panel.background=element_blank(),
    panel.border=element_blank(),
    panel.grid.major=element_blank(),
    panel.grid.minor=element_blank(),
    plot.background=element_blank()
  )

##print(plotDistrib)

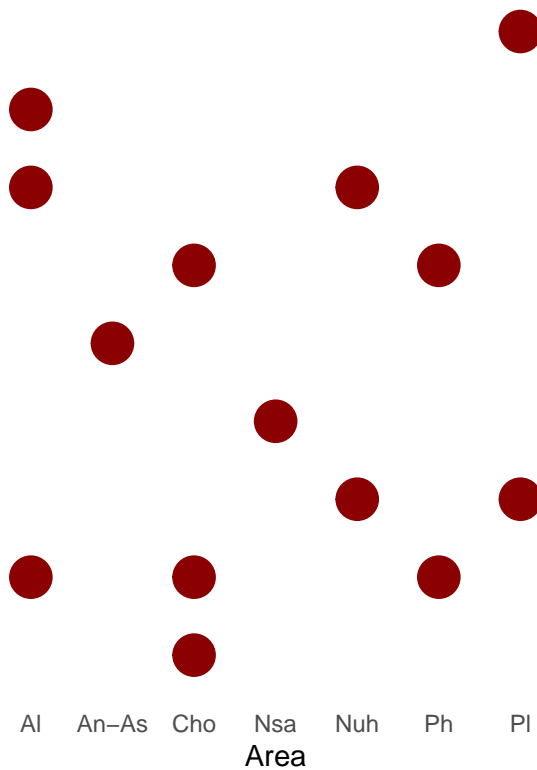
##
cowplot::plot_grid(plotTree, plotDistrib, ncol=2)

```

Rhinoclemmys_TE_likelihood



Distributions



```
##dev.off()
```

We want to test whether the branch lengths will have any effect in the area(s) chosen for conservation using a Phylogenetic Diversity (PD) analysis.

The topology corresponds to a Total Evidence analysis for the *Rhinoclemmys* genus, from R-Alarcon and Miranda-Esquivel (2020), and the distribution is taken and modified from Le and McCord (2008).

```
#par(mfrow=c(2,1))
```

```
#par(mfrow=c(2,1))
```

```
a <- qplot(RhinoclemmysData$tree$edge.length,
  bins=12,
  main=treeFiles,
  xlab="Branch length: internals and terminals") +
  xlim(c(0.001,0.025)) +
  ylim(c(0,5))
```

```
terminals <- RhinoclemmysData$tree$edge[,2] < 1+length(RhinoclemmysData$tree$tip.label)
```

```
b <- qplot(RhinoclemmysData$tree$edge.length[terminals],
  bins=12,
  #main=treeFiles,
```

```

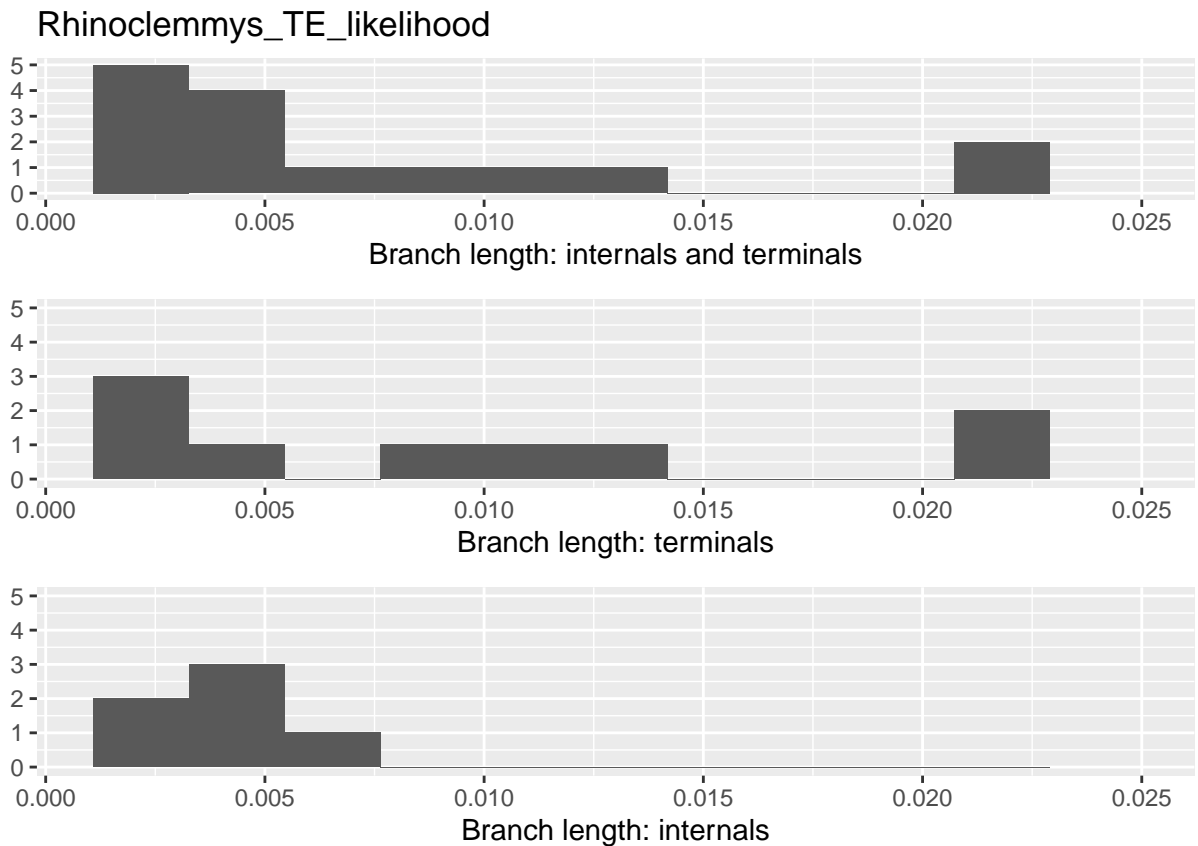
xlab="Branch length: terminals") +
xlim(c(0.001,0.025)) +
ylim(c(0,5))

c <- qplot(RhinoclemmysData$tree$edge.length[!terminals],
  bins=12,
  #main=treeFiles,
  xlab="Branch length: internals") +
  xlim(c(0.001,0.025)) +
  ylim(c(0,5))

#~

cowplot::plot_grid(a, b, c, nrow=3)

```



```

###dev.off()

```

The branch length histogram and the tree plot, show the internal length branches are similar, nad different to terminals'; there are two longer branches, *R. aerolata* (inhabiting AI) and *R. rubida* (PI), while AI and Cho are the richest areas.

Fist we calculate the PD value for the tree, we save the values as a table and later we convert the table to a matrix.

```

RhinoclemmysData$tablePD <- PDindex(RhinoclemmysData$tree,
                                     distribution=RhinoclemmysData$distribution,
                                     root=TRUE)

RhinoclemmysData$matrixPD <- as.data.frame(
  matrix(
    unlist(RhinoclemmysData$tablePD),
    nrow= length(treeFiles),
    byrow=TRUE))

colnames(RhinoclemmysData$matrixPD) <- row.names(RhinoclemmysData$distribution)

row.names(RhinoclemmysData$matrixPD) <- "PD value"

print(t(sort(RhinoclemmysData$matrixPD,decreasing = TRUE)))

```

```

##      PD value
## Cho  0.054150
## Al   0.044127
## Pl   0.041646
## Nuh  0.034323
## Ph   0.032655
## An-As 0.020853
## Nsa  0.019937

```

Now, we test the effect of branch length swapping, terminal and internal branch lengths.

```

for( modelo in c("simpleswap","allswap","uniform") ){

  for( rama in c("terminals","internals") ){

    val <- swapBL(RhinoclemmysData$tree,
                  RhinoclemmysData$distribution,
                  model = modelo,
                  branch = rama
                )

    cat("\n\tTree=",treeFiles,"\n\tModel=",modelo,"\n\tBranchs swapped=",rama,"\n")

    print.blepd(val)

    cat("\n\n\n")

  }

}

```

```

## model to test simpleswap reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= simpleswap

```

```

## Branchs swapped= terminals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al   13      13
## 2           Cho  86      86
## 3           Pl   1       1
##
##
##
## model to test simpleswap reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= simpleswap
## Branchs swapped= internals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Cho 100      100
##
##
##
## model to test allswap reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= allswap
## Branchs swapped= terminals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al   34      34
## 2           Cho  54      54
## 3           Nuh  10      10
## 4           Pl   2       2
##
##
##
## model to test allswap reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= allswap
## Branchs swapped= internals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Cho 100      100
##
##
##
## model to test uniform reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= uniform
## Branchs swapped= terminals

```

```
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al   59      59
## 2           Cho   38      38
## 3           Nuh    3       3
##
##
##
## model to test uniform reps 100
##
## Tree= Rhinoclemmys_TE_likelihoood
## Model= uniform
## Branchs swapped= internals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Cho  100     100
```

The terminal branch length is critical in the decision taken, if we swap all terminal branch lengths (model=*allswap*), or if we replace them with a uniform distribution (model=*uniform*), the area selected might change from Cho to Al. As the terminal branch lengths are distributed unequally, we might suspect that the results depend heavily on the longest branches that inhabit the area.

But first, let us see if the number of replicates has any effect.

```
for(repetir in 1:3){

  val <- swapBL(RhinoclemmysData$tree,
                RhinoclemmysData$distribution,
                model = "allswap",
                branch = "terminals",
                nTimes = 10**repetir
                )

  print.blepd(val)

}
```

```
## model to test allswap reps 10
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al    5      50
## 2           Cho    5      50
## model to test allswap reps 100
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al   41      41
## 2           Cho   43      43
## 3           Nsa    1       1
## 4           Nuh   11      11
## 5           Pl    4       4
## model to test allswap reps 1000
##
```



```
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1           Al  379    37.9
## 2        An-As   13     1.3
## 3           Cho 429    42.9
## 4           Nsa   9     0.9
## 5           Nuh 148    14.8
## 6           Pl  22     2.2
```

Roughly speaking from 1000 – 10000 the results are alike, and the largest difference is in 10 –100. As a rule of thumb, we most use at least 100 replicates, but 1000 will be better.

now, let's see if the possible difference in results could be assigned to both of our longest branches.

```
## Testing the effect if we reduce the branch lenght form the actual
## value to 0.0 using evalTerminal {lower}
```

```
testEvalTerminalALL <- evalTerminal(RhinoclemmysData$tree,
                                     RhinoclemmysData$distribution,
                                     tipToEval = "all",
                                     approach = "lower")
```

```
##
## Terminal: R_annulata, has NO effect on branch length == 0 or Max
## Terminal: R_pulcherrima, has NO effect on branch length == 0 or Max
## Terminal: R_diademata, has NO effect on branch length == 0 or Max
## Terminal: R_punctularia, has NO effect on branch length == 0 or Max
## Terminal: R_melanosterna, has NO effect on branch length == 0 or Max
## Terminal: R_funerea, has NO effect on branch length == 0 or Max
## Terminal: R_areolata, has NO effect on branch length == 0 or Max
## Terminal: R_rubida, has NO effect on branch length == 0 or Max
```

```
print.multiBlepd(testEvalTerminalALL)
```

```
##
## Delta (%)   Terminal   Initial Selected   Evaluating lower
## -0.0047    R_nasuta    Cho Al
## 0          R_annulata    Cho Cho
## 0          R_pulcherrima Cho Cho
## 0          R_diademata    Cho Cho
## 0          R_punctularia Cho Cho
## 0          R_melanosterna Cho Cho
## 0          R_funerea    Cho Cho
## 0          R_areolata    Cho Cho
## 0          R_rubida     Cho Cho
```

```
## Testing the effect if we increase the branch lenght form the actual
## value to the sum of all branch length using evalTerminal {upper}
```

```
testEvalTerminalALL <- evalTerminal(RhinoclemmysData$tree,
                                     RhinoclemmysData$distribution,
                                     tipToEval = "all",
                                     approach = "upper")
```

```
##
## Terminal: R_nasuta, has NO effect on branch length == 0 or Max
```

```
## Terminal: R_annulata, has NO effect on branch length == 0 or Max
## Terminal: R_melanosterna, has NO effect on branch length == 0 or Max
```

```
print.multiBlepd(testEvalTerminalALL)
```

```
##
## Delta (%)      Terminal      Initial Selected      Evaluating upper
## 0      R_nasuta    Cho Cho
## 0      R_annulata    Cho Cho
## 0.0116  R_pulcherrima  Cho Pl
## 0.182   R_diademata   Cho Nsa
## 0.1203  R_punctularia  Cho An-As
## 0      R_melanosterna  Cho Cho
## 0.0391  R_funerea    Cho Al
## 0.0085  R_areolata    Cho Al
## 0.0061  R_rubida     Cho Pl
```

The area selected depends heavily in the branch length of *R. aerolata* and *R. rubida*, if the branch length is just a little shorter, the area selected will change from area A to area B.

To test whether internal branches have more impact than terminal branches, we can use the function eval-TerminalvsInternal.

```
#~ ## calculate values
```

```
terminals <- evalTerminalvsInternal(RhinoclemmysData$tree,
                                     RhinoclemmysData$distribution,
                                     nTimes=100,
                                     branch = "terminals")
```

```
## model to test allswap reps 100
## model to test allswap reps 100
## model to test allswap reps 100
```

```
internals <- evalTerminalvsInternal(RhinoclemmysData$tree,
                                     RhinoclemmysData$distribution,
                                     nTimes=100,
                                     branch = "internals")
```

```
## model to test allswap reps 100
## model to test allswap reps 100
## model to test allswap reps 100
```

```
for( i in 1:3 ){
#   cat("name ",i,"\n")

  cat("Evaluated:",terminals[[i]]$evaluated,"\tBranch swapped:",terminals[[i]]$branch,"\n")
  print.blepd(terminals[[i]])

  cat("Evaluated:",internals[[i]]$evaluated,"\tBranch swapped:",internals[[i]]$branch,"\n")
  print.blepd(internals[[i]])
}
```

```
## Evaluated: All equal      Branch swapped: terminals
```

```

##
## BestInitial:AlCho
##   AreaSelected Freq Percent
## 1      AlCho  100      100
## Evaluated: All equal      Branch swapped: internals
##
## BestInitial:AlCho
##   AreaSelected Freq Percent
## 1      AlCho  100      100
## Evaluated: Terminals      Branch swapped: terminals
##
## BestInitial:AlCho
##   AreaSelected Freq Percent
## 1      AlCho  100      100
## Evaluated: Terminals      Branch swapped: internals
##
## BestInitial:AlCho
##   AreaSelected Freq Percent
## 1      Al    40      40
## 2      AlCho  52      52
## 3      Cho    8       8
## Evaluated: Internals      Branch swapped: terminals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1      Al    43      43
## 2      Cho   40      40
## 3      Nuh   17      17
## Evaluated: Internals      Branch swapped: internals
##
## BestInitial:Cho
##   AreaSelected Freq Percent
## 1      Cho  100      100

```

Le, Minh, and William P. McCord. 2008. "Phylogenetic relationships and biogeographical history of the genus *Rhinoclemmys* Fitzinger, 1835 and the monophyly of the turtle family Geoemydidae (Testudines: Testudinoidea)." *Zoological Journal of the Linnean Society* 153 (4): 751–67. <https://doi.org/10.1111/j.1096-3642.2008.00413.x>.

R-Alarcon, L. Viviana, and Daniel Rafael Miranda-Esquivel. 2020. "A Total-Evidence Phylogeny of the Crown and Stem-Groups of Turtles (Pantestudines: Testudinata)." *Cladistics: Submitted*.