Branch length evaluation for Phylogenetic Diversity: a worked example

Daniel R. Miranda-Esquivel

2018 - 01 - 22

Four taxa and two areas

Preparing the data space

First, we load the required libraries:

```
## cleaning
rm(list = ls())

## libraries

## installing and loading the package

##install.packages("../../blepd_0.1.1.tar.gz", repos = NULL, type="source")

library(blepd)
```

```
## Loading required package: ape

## Loading required package: picante

## Loading required package: vegan

## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.4-5

## Loading required package: nlme
```

```
packageVersion("blepd")
```

```
## [1] '0.1.4.2018.1.22.2218'
## To plot trees; you can use ape or phytools
```

```
library(ggtree)
```

```
## Loading required package: ggplot2

## Loading required package: treeio

##
## Attaching package: 'treeio'

## The following objects are masked from 'package:ape':
##
##     drop.tip, Nnode, Ntip
```

```
## ggtree v1.10.2  For help: https://guangchuangyu.github.io/ggtree
##
## If you use ggtree in published research, please cite:
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam. ggtree: an R package for visualizat
##
## Attaching package: 'ggtree'

## The following object is masked from 'package:nlme':
##
##     collapse

## The following object is masked from 'package:ape':
##
##     rotate
```

```r
library(gridExtra)
```

Now, we load the data included in the package: tree and distributions

```r
## trees

data(package = "blepd")

data(tree)

initialTree <- tree




## distributions

data(distribution)

str(distribution)
```

```
##  int [1:2, 1:4] 1 0 0 1 1 0 0 1
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:2] "A1" "A2"
##   ..$ : chr [1:4] "t1" "t2" "t3" "t4"
```

```r
dist4taxa <- distribution

## distribution to XY

distXY <- matrix2XY(dist4taxa)


## plotting

## the tree

plotTree <-  ggtree(initialTree, ladderize=TRUE,
            color="black", size=1 , linetype="dotted") +
            geom_tiplab(size=6, color="black") +
            theme_tree2() +
            labs(title = "Four terminals, equal branch length")
```
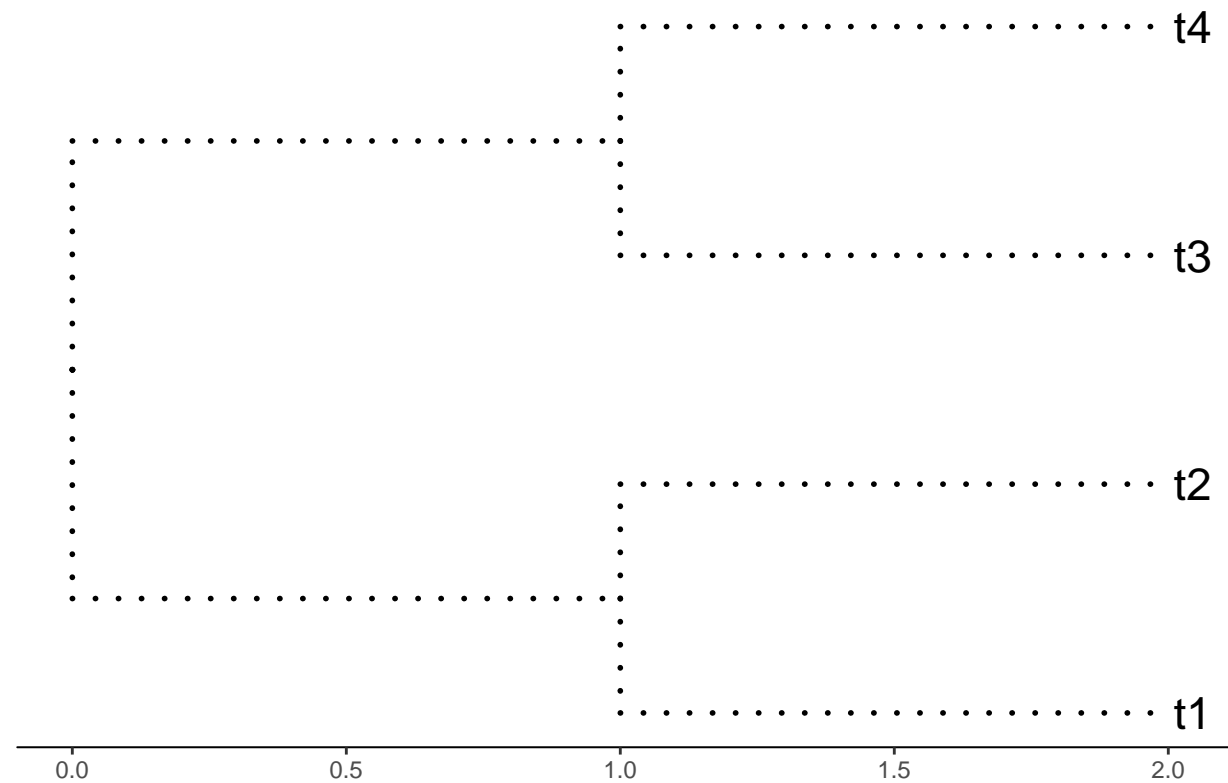
```
print(plotTree)
```
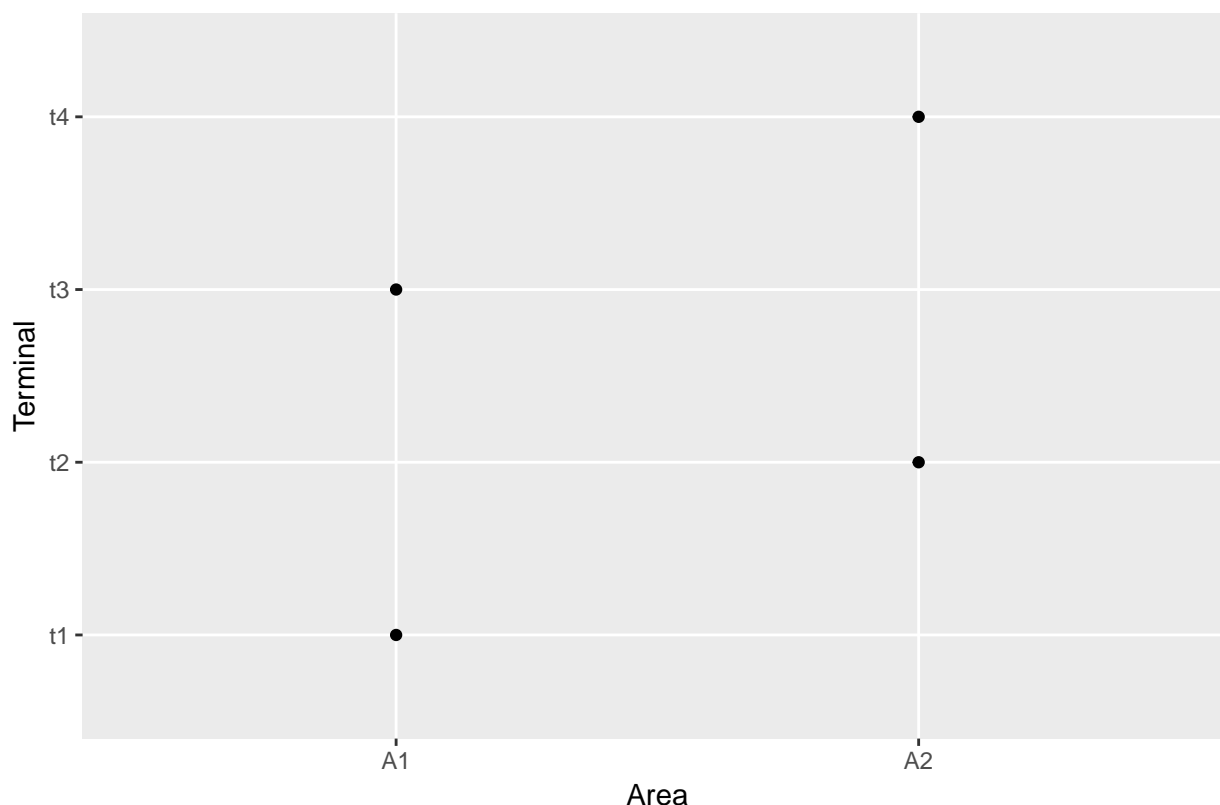
Four terminals, equal branch length


```
## the distribution

plotDistrib <- ggplot(data=distXY,
                      aes(x= Area, y= Terminal),
                      size =11) +
              geom_point() +
              labs(title = "B. Terminals and Distributions",
                   y = "Terminal",
                   x = "Area")

print(plotDistrib)
```

B. Terminals and Distributions

We check whether names in both objects, trees and distributions are the same:

```
all(colnames(dist4taxa) == initialTree$tip.label)
```

```
## [1] TRUE
```

We report all branches' length and calculate the PD values.

```
initialTree$edge.length
```

```
## [1] 1 1 1 1 1 1
```

```
initialPD <- myPD(tree=initialTree, distribution = dist4taxa)

initialPD
```

```
## [1] 4 4
```

Single taxon evaluation function

To test the effect of changing the branch length for a single terminal, we will use evalTerminal:

```
tipToEval <- "t1"

evalTerminal(tree = initialTree, distribution = dist4taxa, tipToEval = "t1", approach = "lower" )
```

```
## branchLengthChange    bestInitialArea    bestModifiedArea
##          "0.9999"            "A1A2"                "A2"
##      initialLength
##                "1"
```

4

The lower limit when we change the branch length for terminal t1 is 0.99, as any change in branch length will modify the area selected from A1A2 to A2, as the tie between the paths terminals t1/t3 (area A1) vs t2/t4 (area A2) will be solved in favour of t2/t4 when A1 is shorter.

Tree evaluation function

The function to test all terminals at the same time is *evalTree*, with two parameters: the tree and the distribution. The function returns a data.frame object with 14 fields: labelTerminal, lowerBranchLength, InitialArea, lowerFinalArea, initialLength, upperBranchLength, upperFinalArea, changeLower, changeUpper, deltaUpper, deltaLower, deltaPD, areaDelta, and abDelta.

```
finalResults <- evalTree(tree = initialTree, distribution = dist4taxa)

finalResults
```

```
##   labelTerminal lowerBranchLength InitialArea lowerFinalArea initialLength
## 1            t1            0.9999        A1A2             A2             1
## 2            t2            0.9999        A1A2             A1             1
## 3            t3            0.9999        A1A2             A2             1
## 4            t4            0.9999        A1A2             A1             1
##   upperBranchLength upperFinalArea changeLower changeUpper deltaUpper
## 1            1.0001             A1          A2          A1      1e-04
## 2            1.0001             A2          A1          A2      1e-04
## 3            1.0001             A1          A2          A1      1e-04
## 4            1.0001             A2          A1          A2      1e-04
##   deltaLower deltaPD areaDelta abDelta
## 1      1e-04       0  LU_A2_A1       0
## 2      1e-04       0  LU_A1_A2       0
## 3      1e-04       0  LU_A2_A1       0
## 4      1e-04       0  LU_A1_A2       0
```

The extreme sensitivity of the PD results to the terminal branch length is seen in the column absolute length difference (=abDelta), as any length change -larger than 0-, will change the area selected.

We plot the results to see the effect in each terminal, as a table:

```
plotResults <- ggplot(data=finalResults, aes(x= labelTerminal, y= initialLength,
                                 shape="Actual",
                                 colour=InitialArea
                                 ) ) +
   geom_point(size= 7) +
  geom_point(aes(x= labelTerminal, y= lowerBranchLength,
                colour=lowerFinalArea,
                shape="Lower_limit"), size=7) +
  geom_point(aes(x= labelTerminal, y= upperBranchLength,
                colour=upperFinalArea,
                shape="Upper_limit"), size=7) +
  labs(title = "C. Branch length change, All branches are equal",
       colour = "Area selected",
       shape = "Terminal branch length value",
       y = "Terminal branch length",
       x = "Terminal")

print(plotResults)
```

C. Branch length change, All branches are equal
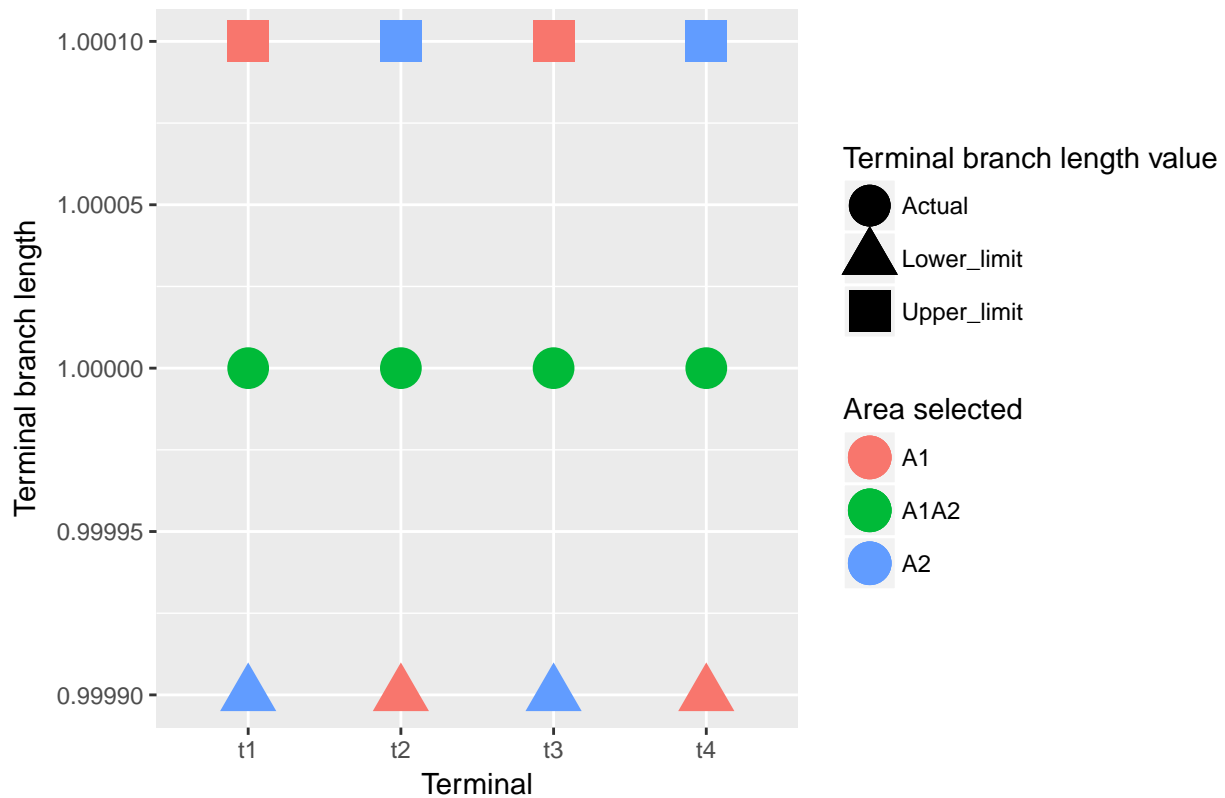
or plotted as a simple table:

```r
countFreqChanges <- table(finalResults$areaDelta)


countFreqChanges <- as.data.frame(countFreqChanges, ncol=1)


colnames(countFreqChanges) <- c("Area change","Freq")

row.names(countFreqChanges) <- NULL


countFreqChanges

##    Area change Freq
## 1    LU_A1_A2    2
## 2    LU_A2_A1    2
```

or plotted into the tree:

```r
p0 <-    ggtree(initialTree, layout="slanted", ladderize=TRUE,
                color="black", size=0.8 ) +
        theme(legend.position="right") +
        labs(title = paste("Initial area selected: ",finalResults$maxPD, sep = ""))


p <- p0 %<+% finalResults + geom_tiplab(aes(color=areaDelta), size =6)
```

```
print(p)
```

Initial area selected: