

portal-phylogenies.Rmd

Azul Carrillo

2022-10-25

The Newick tree format allows us to represent trees in a readable way for computers. This uses the correspondance between the trees and the nested parenthesis. In other words this format is used to describe in detail a phylogenetic tree as text. In the format the parenthesis group the sequence names. The branch lengths can be included using a colon followed by the length. Finally this is finished by a semicolon no branch length `ex(,(,))`; this will just add the sequence names. Ex. branch length included `(B:1.0,(A:2.0,C:3.0,E:4.0):5.0,D:6.0);`.

package ‘ape’

```
#install.packages("ape")
library(ape)
```

```
portal_tree <- read.tree(file = "../raw-data/portal-tree.tre")
```

branch length Dna- will be propotional to mutation rate by some unit of time, change in DNA per kind

```
small_tree <- read.tree(file = "http://ape-package.ird.fr/APER/APER2/primfive.tre")
small_tree
```

```
##
## Phylogenetic tree with 5 tips and 4 internal nodes.
##
## Tip labels:
##   Homo, Pongo, Macaca, Ateles, Galago
##
## Rooted; includes branch lengths.
```

The structure of a tree in R

classes we’ve been working with data.frames next we will work with is phylo

```
class(portal_tree)
```

```
## [1] "phylo"
```

```
length(portal_tree) # gives the number of
```

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

```
names(portal_tree)
```

```
## [1] "edge"      "Nnode"      "node.label" "tip.label"
```

```
#portal_tree$Nnode
```

```
head(portal_tree["tip.label"])
```

```
## $tip.label
```

```
## [1] "Sigmodon_ochrognathus"      "Sigmodon_hispidus"
## [3] "Sigmodon_fulviventer"      "Neotoma_albigula"
## [5] "Onychomys_leucogaster"      "Onychomys_torridus"
## [7] "Peromyscus_maniculatus"     "Peromyscus_leucopus"
## [9] "Peromyscus_eremicus"       "Reithrodontomys_fulvescens"
## [11] "Reithrodontomys_montanus"   "Reithrodontomys_megalotis"
## [13] "Baiomys_taylori"           "Chaetodipus_intermedius"
## [15] "Chaetodipus_penicillatus"   "Chaetodipus_baileyi"
## [17] "Chaetodipus_hispidus"       "Perognathus_flavus"
## [19] "Dipodomys_ordii"           "Dipodomys_merriami"
## [21] "Dipodomys_spectabilis"     "Xerospermophilus_spilosoma"
## [23] "Xerospermophilus_tereticaudus" "Ammospermophilus_harrisii"
## [25] "Sylvilagus_audubonii"       "Campylorhynchus_brunneicapillus"
## [27] "Ammodramus_savannarum"      "Kieneria_fusca"
## [29] "Pipilo_chlorurus"           "Pooecetes_gramineus"
## [31] "Zonotrichia_leucophrys"     "Spizella_breweri"
## [33] "Amphispiza_bilineata"       "Calamospiza_melanocorys"
## [35] "Zenaida_macroura"           "Callipepla_squamata"
## [37] "Crotalus_viridis"           "Crotalus_scutulatus"
## [39] "Sceloporus_undulatus"        "Sceloporus_clarkii"
## [41] "Gambelia_sila"              "Aspidoscelis_uniparens"
## [43] "Aspidoscelis_tigris"
```

```
head(portal_tree[[3]])
```

```
## [1] "Amniota"
## [2] "Mammalia"
## [3] "'Theria (subclass in Deuterostomia)'"
## [4] "'Eutheria (in Deuterostomia)'"
## [5] "Boreoeutheria"
## [6] "Euarchontoglires"
```

```
summary(portal_tree)
```

```
##
## Phylogenetic tree: portal_tree
##
##   Number of tips: 43
##   Number of nodes: 345
##   No branch lengths.
##   No root edge.
##   First ten tip labels: Sigmodon_ochrognathus
##                         Sigmodon_hispidus
##                         Sigmodon_fulviventer
##                         Neotoma_albigula
##                         Onychomys_leucogaster
##                         Onychomys_torridus
##                         Peromyscus_maniculatus
##                         Peromyscus_leucopus
##                         Peromyscus_eremicus
##                         Reithrodontomys_fulvescens
##   First ten node labels: Amniota
##                         Mammalia
##                         'Theria (subclass in Deuterostomia)'
##                         'Eutheria (in Deuterostomia)'
##                         Boreoeutheria
##                         Euarchontoglires
##                         mrcaott42ott30082
##                         Glires
##                         mrcaott42ott29157
##                         Rodentia
```

```
str(portal_tree) # $ to access elements
```

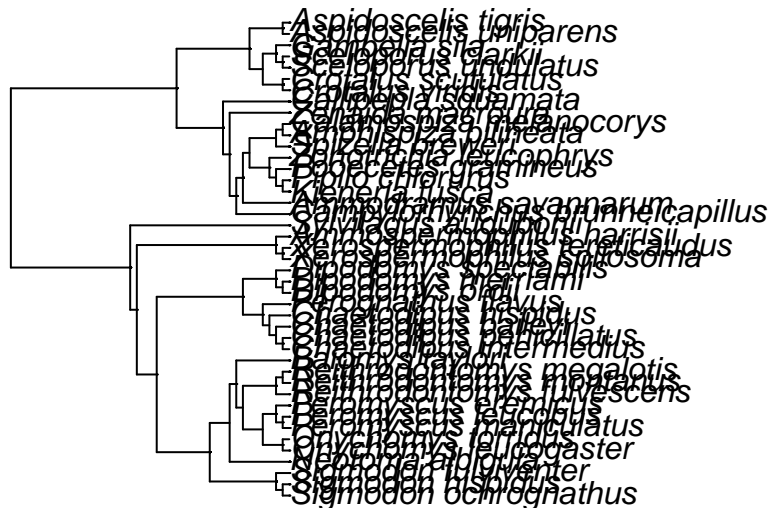
```
## List of 4
## $ edge      : int [1:387, 1:2] 44 45 46 47 48 49 50 51 52 53 ...
## $ Nnode     : int 345
## $ node.label: chr [1:345] "Amniota" "Mammalia" "'Theria (subclass in Deuterostomia)'" "'Eutheria (in Deuterostomia)'"
## $ tip.label : chr [1:43] "Sigmodon_ochrognathus" "Sigmodon_hispidus" "Sigmodon_fulviventer" "Neotoma_albigula"
## - attr(*, "class")= chr "phylo"
## - attr(*, "order")= chr "cladewise"
```

```
head(portal_tree$edge)
```

```
##      [,1] [,2]
## [1,]  44  45
## [2,]  45  46
## [3,]  46  47
## [4,]  47  48
## [5,]  48  49
## [6,]  49  50
```

list is similar to a data frame but it allows combining object of any kind

```
plot.phylo(x = portal_tree)
```



An extension of gg plt for phylogenetic visualization **ggtree** to install packages from Bioconductor repo we need a CRAN package called BiocManager

```
#install.packages("BiocManager")
library(BiocManager)
install("ggtree")
```

```
## Bioconductor version 3.15 (BiocManager 1.30.19), R 4.2.1 (2022-06-23)
```

```
## Warning: package(s) not installed when version(s) same as or greater than current; use
## 'force = TRUE' to re-install: 'ggtree'
```

```
## Old packages: 'cluster', 'foreign', 'jsonlite', 'MASS', 'Matrix', 'mgcv',  
## 'nlme', 'nnet', 'rpart', 'survival', 'vctrs', 'xfun'
```

```
library(ggtree)
```

```
## ggtree v3.4.4 For help: https://yulab-smu.top/treedata-book/
```

##

```
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
```

##

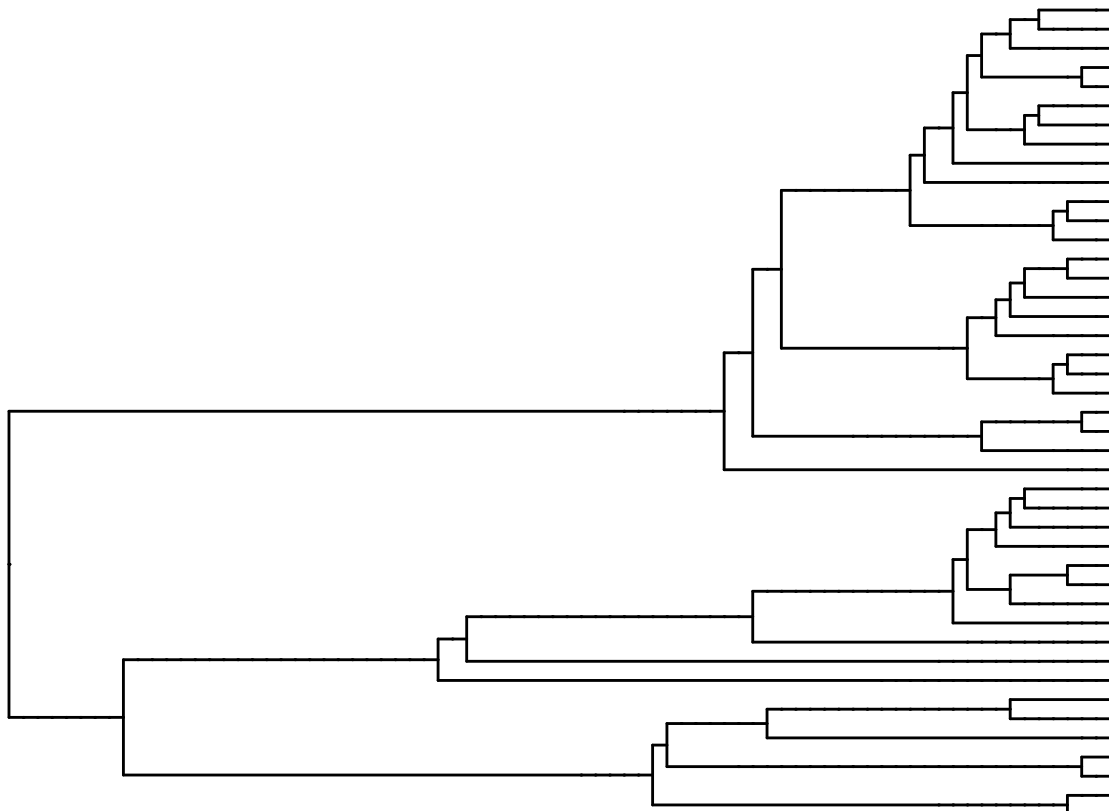
```
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## G Yu. Data Integration, Manipulation and Visualization of Phylogenetic
## Trees (1st ed.). Chapman and Hall/CRC. 2022. ISBN: 9781032233574
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
```

```
##
## Attaching package: 'ggtree'
```

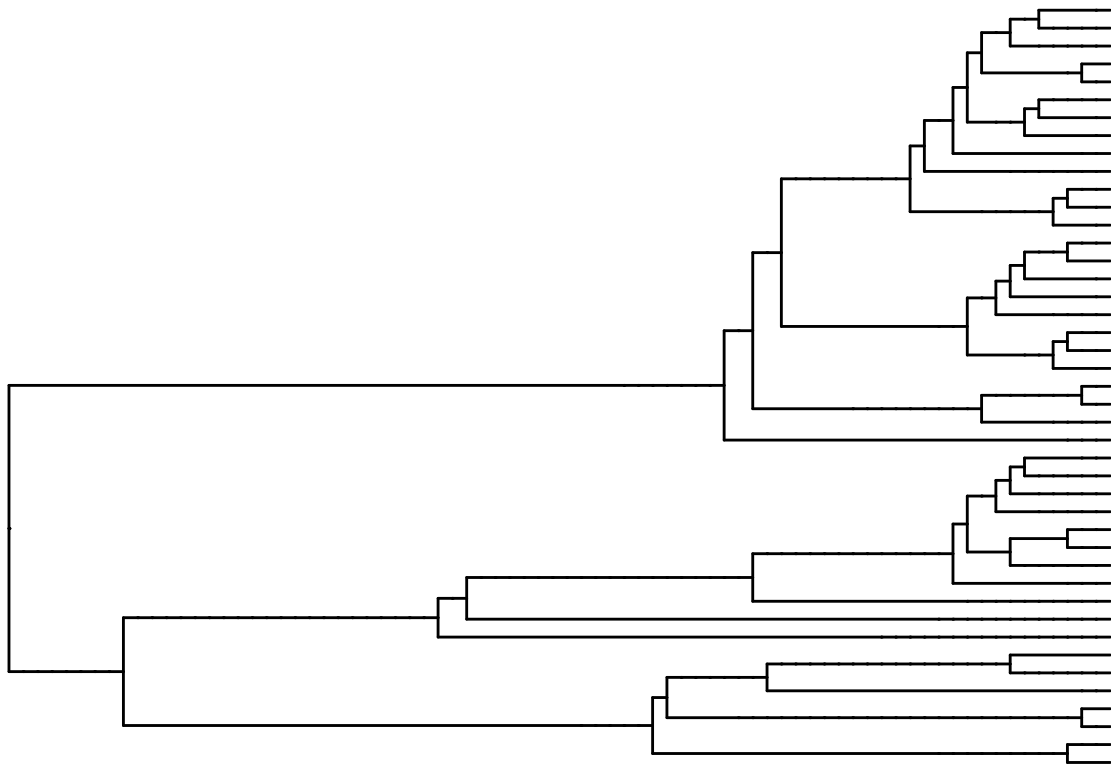
```
## The following object is masked from 'package:ape':
##
## rotate
```

Plotting a Phylogenetic tree

```
ggtree(portal_tree)
```

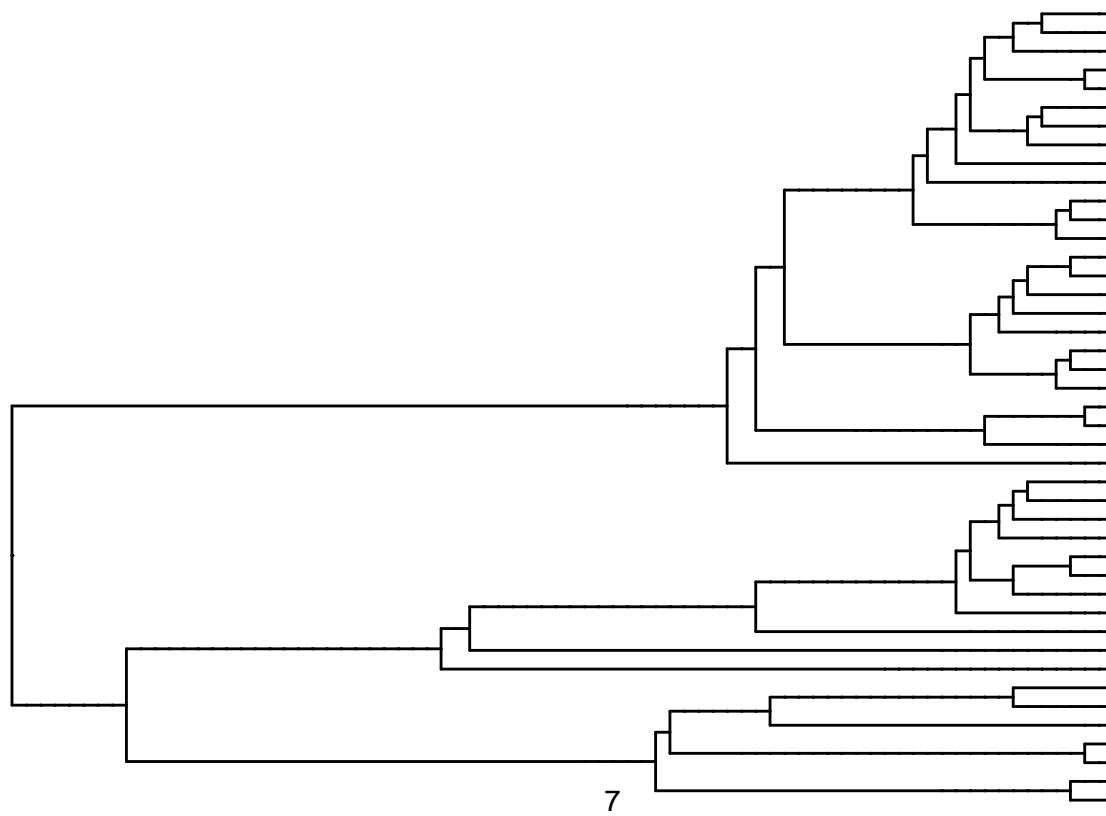


```
#this is equivalent to
ggplot(portal_tree, aes(x,y)) +
  geom_tree() +
  theme_tree()
```



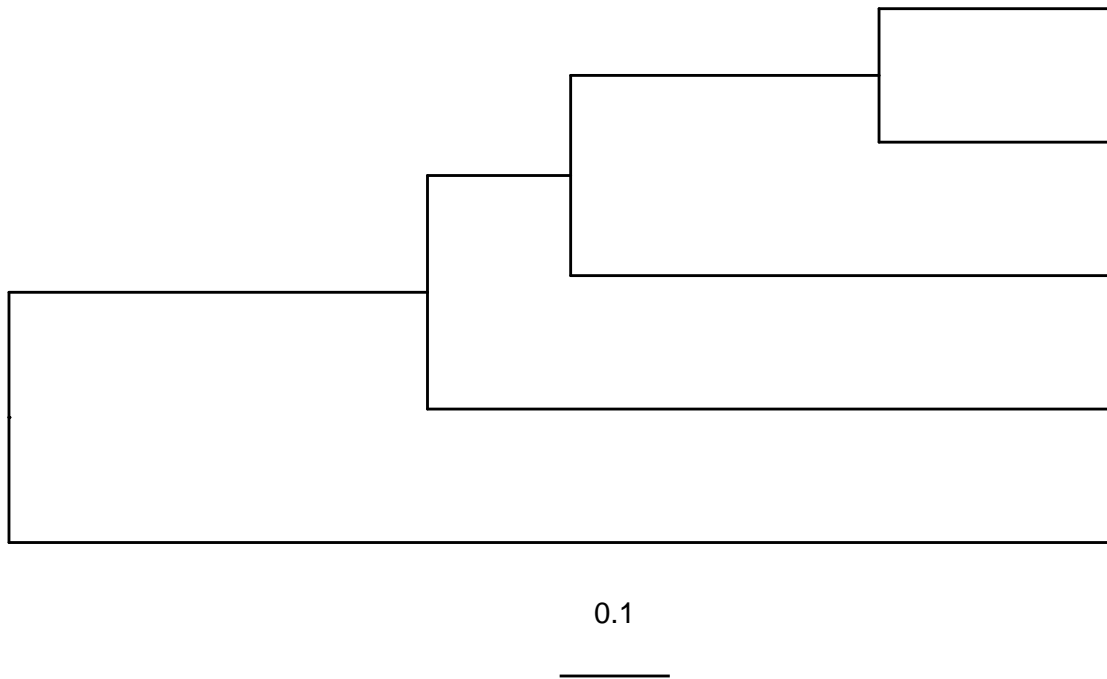
to add a scale use the function `geom_treescales()`:

```
ggtree(portal_tree)+  
  geom_treescale()
```



plot the small tree of five species of primates and include a scale what is the difference between the plot and the small tree

```
ggtree(small_tree) +  
  geom_treescale()
```



```
class(small_tree)
```

```
## [1] "phylo"
```

```
length(small_tree) # gives the number of
```

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

```
names(small_tree)
```

```
## [1] "edge"      "edge.length" "Nnode"      "tip.label"
```

```
small_tree$Nnode
```

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

```
small_tree["tip.label"]
```

```
## $tip.label
```

```
## [1] "Homo"  "Pongo" "Macaca" "Ateles" "Galago"
```



```
head(small_tree[[3]])
```

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

```
summary(small_tree)
```

```
##
## Phylogenetic tree: small_tree
##
##   Number of tips: 5
##   Number of nodes: 4
##   Branch lengths:
##     mean: 0.415
##     variance: 0.08208571
##     distribution summary:
##   Min. 1st Qu.  Median 3rd Qu.    Max.
## 0.1300 0.2100 0.3300 0.5225 1.0000
##   No root edge.
##   Tip labels: Homo
##               Pongo
##               Macaca
##               Ateles
##               Galago
##   No node labels.
```

```
str(small_tree) # $ to access elements
```

```
## List of 4
## $ edge      : int [1:8, 1:2] 6 7 8 9 9 8 7 6 7 8 ...
## $ edge.length: num [1:8] 0.38 0.13 0.28 0.21 0.21 0.49 0.62 1
## $ Nnode      : int 4
## $ tip.label  : chr [1:5] "Homo" "Pongo" "Macaca" "Ateles" ...
## - attr(*, "class")= chr "phylo"
## - attr(*, "order")= chr "cladewise"
```

```
head(small_tree$edge)
```

```
##      [,1] [,2]
## [1,]    6    7
## [2,]    7    8
## [3,]    8    9
## [4,]    9    1
## [5,]    9    2
## [6,]    8    3
```

```
# for portal_tree the Number of tips: 43 Number of nodes: 345 therefor it is larger than the small tree
```

```
small_tree$edge.length
```

```
## [1] 0.38 0.13 0.28 0.21 0.21 0.49 0.62 1.00
```

```
head(small_tree$node.label)
```

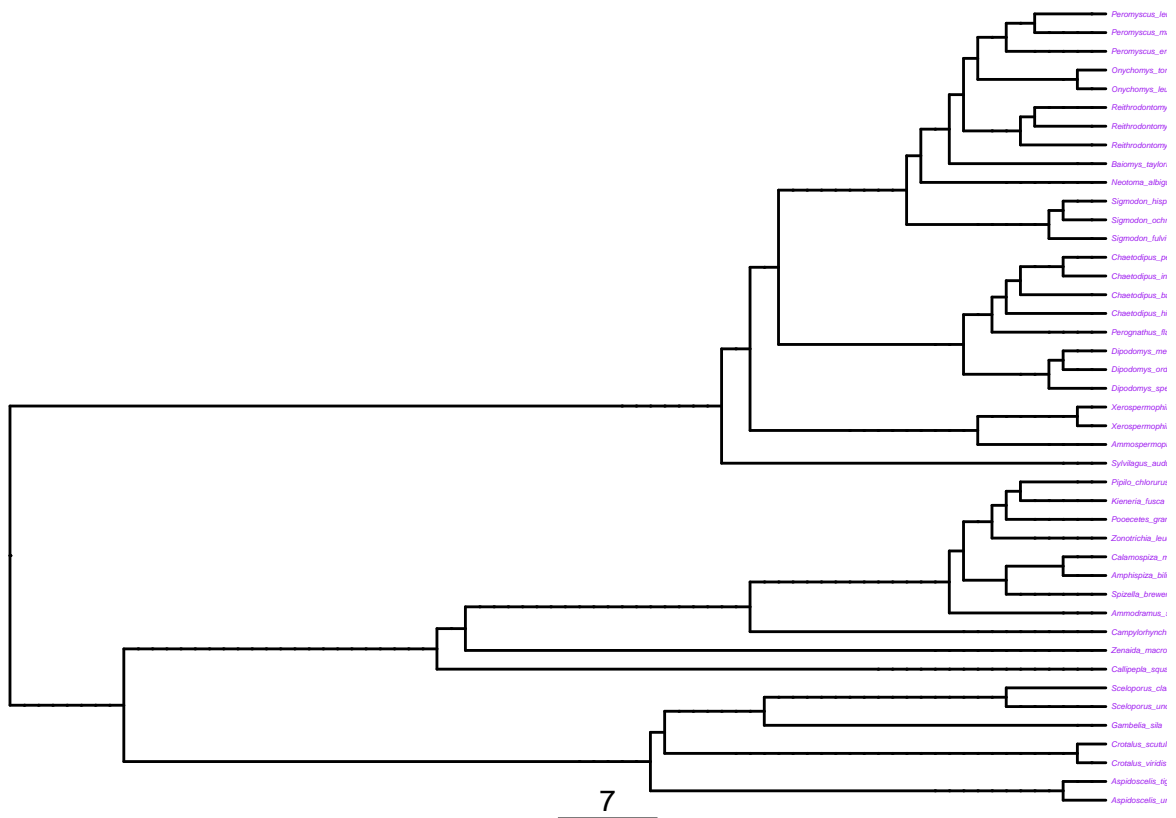
```
## NULL
```

```
head(portal_tree$node.label)
```

```
## [1] "Amniota"
## [2] "Mammalia"
## [3] "'Theria (subclass in Deuterostomia)'"
## [4] "'Eutheria (in Deuterostomia)'"
## [5] "Boreoeutheria"
## [6] "Euarchontoglires"
```

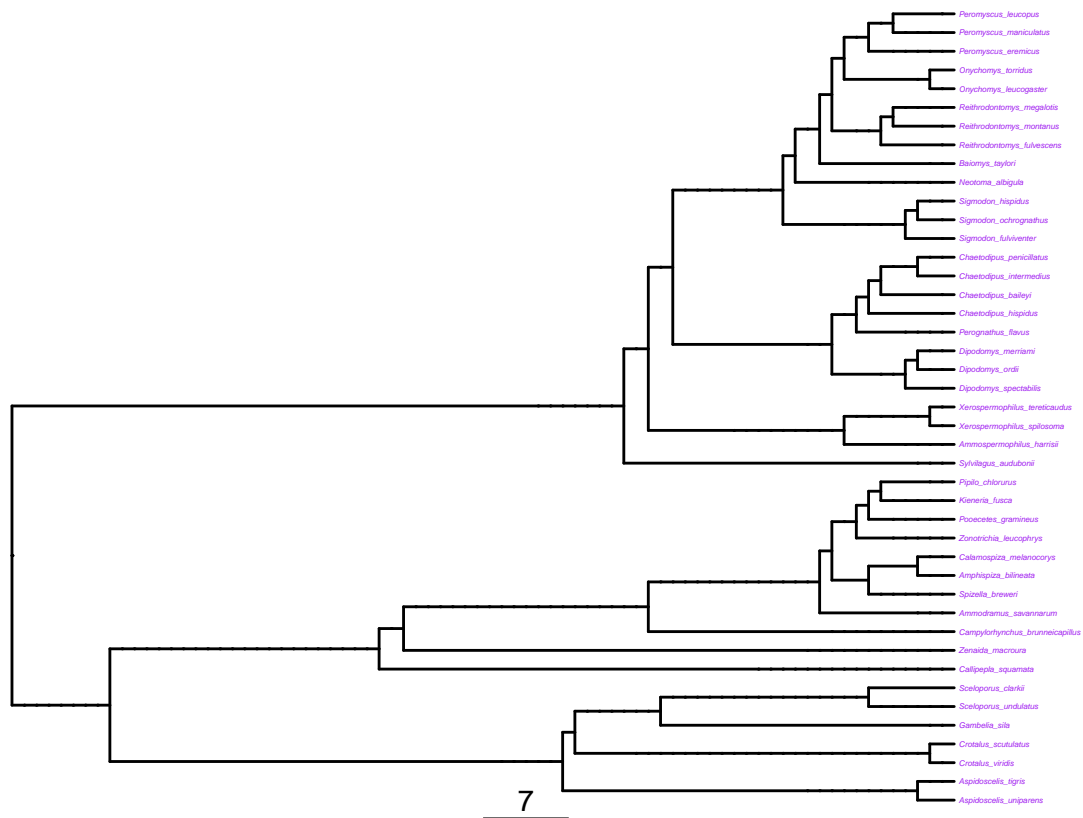
```
#Add tip labels and node labels because a plo is nothing with no labels
```

```
ggtree(portal_tree) +  
  geom_treescale() +  
  geom_tiplab(size = 1, color = "purple", fontface = "italic")
```

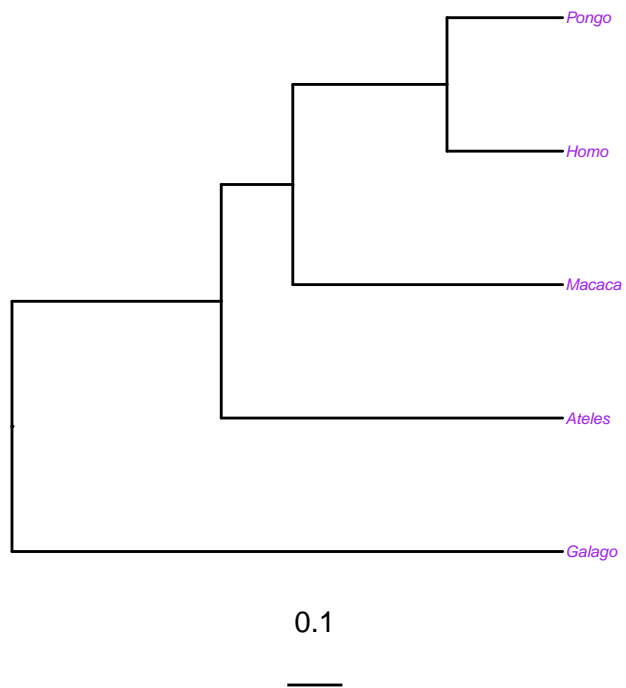


add a limit to the plot so that we can see the labels

```
ggtree(portal_tree) +
  geom_treescale() +
  geom_tiplab(size = 1, color = "purple", fontface = "italic") +
  xlim(NA, 90)
```



```
ggtree(small_tree) +
  geom_treescale() +
  geom_tiplab(size = 2, color= "purple", fontface = "italic") +
  xlim(NA, 2)
```



```
library(ape)
branching.times(small_tree)
```

```
##      6      7      8      9
## 1.00 0.62 0.49 0.21
```