

# Dragon Phylogeny

Roshael Chellappah (20103016)

09/03/2022

*GitHub* Username: RoshaelC Repository Link: <https://github.com/RoshaelC/DNAalignment.git>  
(<https://github.com/RoshaelC/DNAalignment.git>)

## Import Data

```
library(ape)
DragonNexus <- read.nexus.data("input/DragonMatrix.nex")

DragonNexusDF <- data.frame(matrix(unlist(DragonNexus), ncol = 78, byrow = T))
row.names(DragonNexusDF) <- names(DragonNexus)
head(DragonNexusDF)
```

```
##           X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X18 X19
## 0.1FishXXX 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 0.2SnakeXX 1 1 1 1 0 0 0 0 1 1 1 0 0 1 1 1 0 1 1
## 0.3MammalX 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0
## 1GermanXXX 0 1 0 0 1 1 1 0 0 1 1 1 1 0 0 0 0 1 1
## 2FrenchXXX 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 1 0 1 0
## 3FrenchXXX 0 1 0 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 ?
##           X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30 X31 X32 X33 X34 X35 X36
## 0.1FishXXX 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 0.2SnakeXX 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0
## 0.3MammalX 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 1 1
## 1GermanXXX 0 0 0 0 1 0 1 1 1 1 1 0 0 0 0 0 1 1
## 2FrenchXXX 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 1 1
## 3FrenchXXX ? ? 1 0 1 0 0 1 1 1 1 0 0 0 0 0 1 1
##           X37 X38 X39 X40 X41 X42 X43 X44 X45 X46 X47 X48 X49 X50 X51 X52 X53
## 0.1FishXXX 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 0.2SnakeXX 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0
## 0.3MammalX 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0
## 1GermanXXX 0 1 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0
## 2FrenchXXX 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0
## 3FrenchXXX 0 1 1 1 0 1 1 1 0 1 0 0 0 0 0 0 0
##           X54 X55 X56 X57 X58 X59 X60 X61 X62 X63 X64 X65 X66 X67 X68 X69 X70
## 0.1FishXXX 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 0.2SnakeXX 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 1 0
## 0.3MammalX 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0
## 1GermanXXX 1 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0
## 2FrenchXXX 1 1 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0
## 3FrenchXXX 1 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0
##           X71 X72 X73 X74 X75 X76 X77 X78
## 0.1FishXXX 0 0 0 0 0 0 0 0
## 0.2SnakeXX 0 0 0 0 0 0 0 0
## 0.3MammalX 0 0 0 1 0 0 0 0
## 1GermanXXX ? ? ? 0 0 1 1 1
## 2FrenchXXX 0 1 1 0 0 1 1 0
## 3FrenchXXX 0 1 1 0 0 1 1 1
```

```
DragonDist <- dist(DragonNexusDF, method = "binary")
```

```
## Warning in dist(DragonNexusDF, method = "binary"): NAs introduced by coercion
```

```
DragonDistMat <- as.matrix(DragonDist)
```

## Visualizing the Matrix

```
library(reshape2)
PDat <- melt(DragonDistMat)
```

# Weights

Setting up the weight distribution as done in Dragon Phylogeny Tutorial

```
WeightsDat <- read.csv("input/Weights.csv")

Weights <- paste0(WeightsDat$Weight, collapse = "")
Weights <- strsplit(Weights, split = "") [[1]]

WeightsNum <- rep(NA, length(Weights))

for (i in 1:length(WeightsNum)){
  if (Weights[i] %in% LETTERS){
    WeightsNum[i] <- which(LETTERS == Weights[i]) + 9
  } else {
    WeightsNum[i] <- Weights[i]
  }
}
WeightsNum <- as.numeric(WeightsNum)

length(WeightsNum)
```

```
## [1] 78
```

```
WtDragonNexus = DragonNexus

for (i in 1:length(DragonNexus)){
  RepWeight <- DragonNexus[[i]] == 1
  WtDragonNexus[[i]][RepWeight] <- WeightsNum[RepWeight]
  RepWeight <- NA
}
```

Calculate the distance matrix

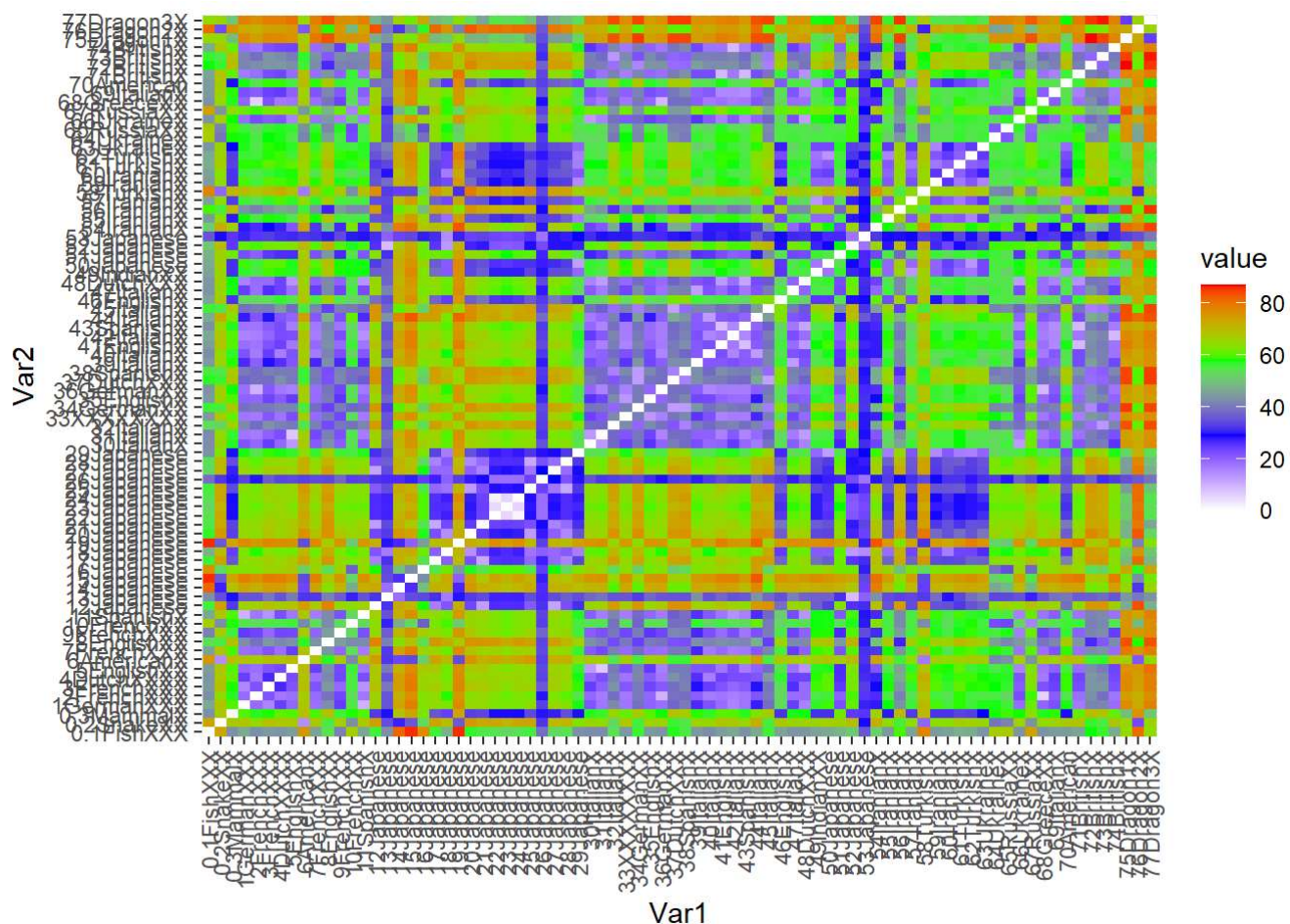
```
WtDragonNexusDF <- data.frame(matrix(unlist(WtDragonNexus), ncol = 78, byrow = T))
row.names(WtDragonNexusDF) <- names(WtDragonNexus)
WtDragonDist <- dist(WtDragonNexusDF, method = 'euclidean')
```

```
## Warning in dist(WtDragonNexusDF, method = "euclidean"): NAs introduced by
## coercion
```

```
WtDragonDistMat <- as.matrix(WtDragonDist)

library(ggplot2)

WtPDat <- melt(WtDragonDistMat)
ggplot(data = WtPDat, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() + scale_fill_gradientn(colours = c("white", "blue", "green", "red")) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



## Plotting the Tree

```
WtDragonTree<-fastme.bal(WtDragonDist)
WtDragonTreeNJ<-nj(WtDragonDist)
str(WtDragonTree)
```

```
## List of 4
## $ edge      : int [1:157, 1:2] 81 81 82 83 84 84 85 86 87 88 ...
## $ edge.length: num [1:157] 25.25 5.37 6.03 2.43 1.77 ...
## $ tip.label  : chr [1:80] "0.1FishXXX" "26Japanese" "0.2SnakeXX" "6AmericanX" ...
## $ Nnode      : int 78
## - attr(*, "class")= chr "phylo"
## - attr(*, "order")= chr "cladewise"
```

## Formatting the Tree

```
head(WtDragonTree$tip.label)
```

```
## [1] "0.1FishXXX" "26Japanese" "0.2SnakeXX" "6AmericanX" "76Dragon2X"  
## [6] "58TurkishX"
```

```
Country <- gsub("[0-9\\.|]+(^[X]*).*", "\\1", WtDragonTree$tip.label) # remove the Xs and Numbers  
from the names
```

```
CountryGroups <- split(WtDragonTree$tip.label, Country)
```

```
library(ggtree)
```

```
## ggtree v3.2.1 For help: https://yulab-smu.top/treedata-book/  
##  
## If you use ggtree in published research, please cite the most appropriate paper(s):  
##  
## 1. Guangchuang Yu. Using ggtree to visualize data on tree-like structures. Current Protocols  
in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96  
## 2. Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods for mapping and visu  
alizing associated data on phylogeny using ggtree. Molecular Biology and Evolution. 2018, 35(1  
2):3041-3043. doi:10.1093/molbev/msy194  
## 3. Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam. ggtree: an R packag  
e for visualization and annotation of phylogenetic trees with their covariates and other associa  
ted data. Methods in Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
```

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

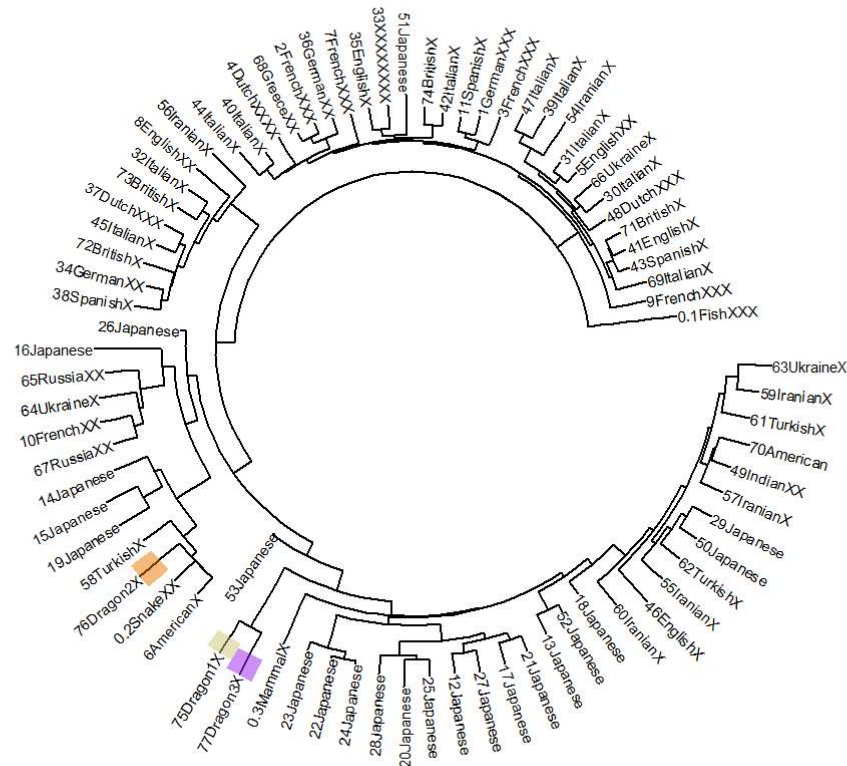
```
## The following object is masked from 'package:ape':  
##  
## rotate
```

```
WtDTcol <- groupOTU(WtDragonTree, CountryGroups)  
str(WtDTcol)
```

```
## List of 4  
## $ edge : int [1:157, 1:2] 81 81 82 83 84 84 85 86 87 88 ...  
## $ edge.length: num [1:157] 25.25 5.37 6.03 2.43 1.77 ...  
## $ tip.label : chr [1:80] "0.1FishXXX" "26Japanese" "0.2SnakeXX" "6AmericanX" ...  
## $ Nnode : int 78  
## - attr(*, "class")= chr "phylo"  
## - attr(*, "order")= chr "cladewise"  
## - attr(*, "group")= Factor w/ 22 levels "", "American", ...: 9 16 19 2 5 21 16 16 16 18 ...
```

```
ggtree(WtDTcol, layout = "circular") + geom_tiplab(size = 2, aes(angle = angle)) + geom_highlight(
  node = 43, fill = "purple") + geom_highlight(node = 42, fill = "khaki3") + geom_highlight(node
  = 5, fill = "darkorange2") + xlim(-50, NA)
```

```
## Warning: The tree contained negative edge length. If you want to ignore the edge, you
## can set 'options(ignore.negative.edge=TRUE)', then re-run ggtree.
```



## Dragon Information





Dragon 1 (highlighted beige/khaki in phylogeny)



Dragon 2 (highlighted orange in phylogeny)



Dragon 3 (highlighted purple in phylogeny)