PhyloMorphoSpace Tutorial

Ian G. Brennan

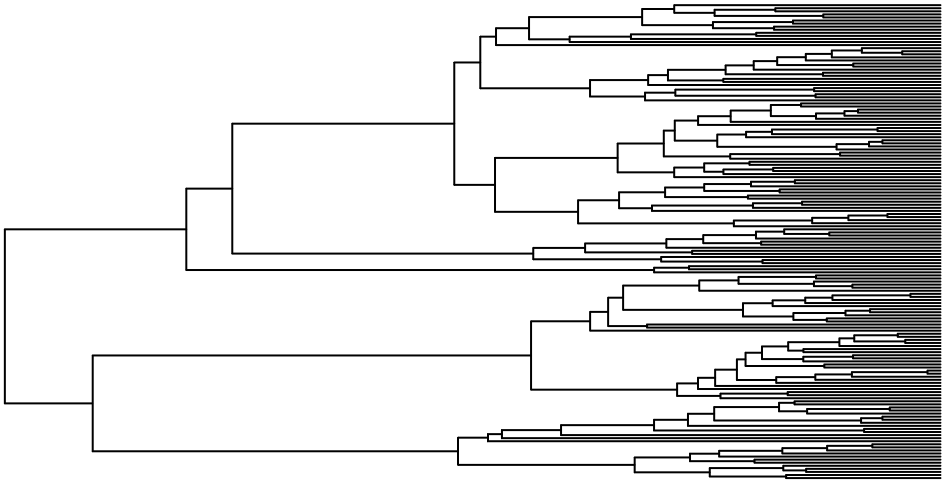
04/09/2018

We're interested in plotting our tree and data together to look for signature of ecology, phylogeny, et al. Our first step is to install or load in some dependencies.

```
library(phytools) # 'ape' is a dependency of phytools, so will load too
library(dplyr)
library(devtools)
# devtools::install_github("awhstin/awtools") # you only need to do this once
library(awtools)
```

We'll start by reading in the tree of interest. Trees are usually in one of two formats: 'newick' (use 'read.tree') or 'nexus' (use 'read.nexus').

```
all.tree <- read.tree("/Users/Ian/Desktop/Olori.Pygopodoidea.tre")
plot(all.tree, show.tip.label=F);
```



```
## [1] "155 taxa in the tree"
```

Now, try reading in the data.

```
all.data <- read.csv("/Users/Ian/Desktop/Olori.Pygopodidae.Data.csv", header=T)
all.data[1:2,] # this will show us the top two rows of the data frame
```

```
##       Name_in_Tree   Genus   Ecology Head_Width Head_Length Trunk_Length
## 1   Aprasia.aurita Aprasia fossorial        2.9         3.8        106.2
## 2 Aprasia.inaurita Aprasia fossorial        2.8         3.6        132.4
##   SVL Tail_Length Total_Length
## 1 110       72.60       182.60
## 2 136      103.36       239.36
```

```
## [1] "42 taxa in the data"
```

You might notice the tree is for all the pygopodoid geckos (Carphodatylidae, Diplodactylidae, Pygopodidae), but the data is just for the Pygopodidae. We'll trim the data and tree down to match one another.

```r
keepers <- intersect(all.tree$tip.label, all.data$Name_in_Tree) # find overlapping taxa
tree <- drop.tip(all.tree, setdiff(all.tree$tip.label, keepers)) # trim with 'drop.tip'
pdata <- filter(all.data, Name_in_Tree %in% keepers) # trim the data with 'filter'
```

```
## [1] "41 taxa in the tree; 41 taxa in the dataframe"
```

log transform the data, then extract the columns (traits) of interest

```r
pdata[,7:8] <- log(pdata[,7:8]) # traits of interest, you'd use PC1 + PC2
rownames(pdata) <- pdata$Name_in_Tree # provide names for the data
morpho <- pdata[,c("SVL", "Tail_Length")] # keep these traits (PC1 + PC2)
```

We will want to color our plot by some factor (ecology, Genus, diet, diel activity…). Stary by extracting the factor, associating it with tip names, then applying a color.

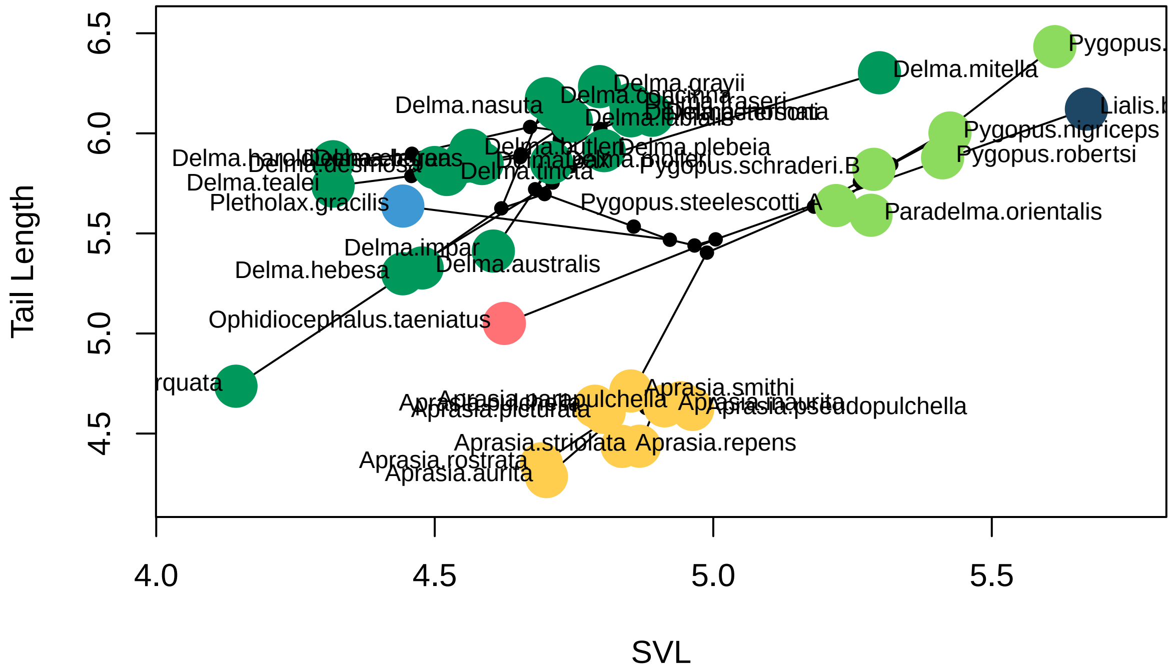
```r
genus <- pdata$Genus
names(genus) <- pdata$Name_in_Tree
mycol <- character(length(genus))
mycol[genus=="Aprasia"] <- mpalette[2]
mycol[genus=="Delma"] <- mpalette[1]
mycol[genus=="Lialis"] <- mpalette[3]
mycol[genus=="Ophidiocephalus"] <- mpalette[4]
mycol[genus=="Paradelma"] <- mpalette[5]
mycol[genus=="Pygopus"] <- mpalette[5]
mycol[genus=="Pletholax"] <- mpalette[6]
```

This next bit gets a little ugly. We need to give 'phylomorphospace' not only the colors of the tips, but the internal nodes of the tree too. The most important part is making sure that the order of the tip colors and data matches the order of the tips in the tree. This is annoying, but necessary.

```r
# add the colors back to the data frame
pdata$color <- mycol
# match the tree order
pdata <- pdata[match(tree$tip.label, pdata$Name_in_Tree),]
# extract the colors in proper order
group.colors <- pdata$color
# supply the tip names (numbers)
names(group.colors) <- 1:Ntip(tree)
# color all internal nodes black
nodecols <- rep("black", tree$Nnode)
# supply internal node numbers
names(nodecols) <- (Ntip(tree)+1) : (Ntip(tree)+tree$Nnode)
# combine the two back together
colorz <- c(group.colors, nodecols)
```

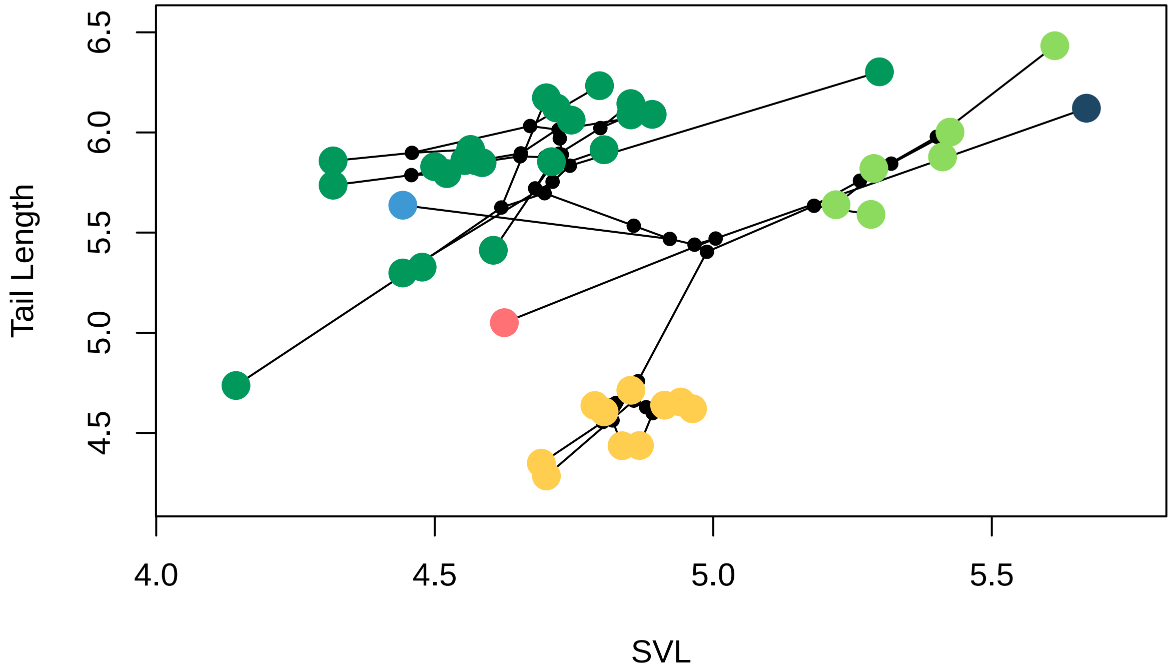
Finally, we're ready to build our plot.

```r
phylomorphospace(tree, morpho,
                 label = "horizontal",
                 node.size = c(1, 3),
                 control=list(col.node=colorz),
                 xlab = "SVL", ylab = "Tail Length")
```



If we want it to look a little prettier, we can remove the labels, shrink the nodeballs, etc.

```r
phylomorphospace(tree, morpho,
                 label = "off",
                 node.size = c(1, 2),
                 control=list(col.node=colorz),
                 xlab = "SVL", ylab = "Tail Length")
```



It looks like there's pretty strong phylogenetic signal in the proportion of tail length to SVL. Look how tightly Aprasia are clustered. Signal of strong functional constraint? Remember, this is really just a visualization, if we really wanted to test for separation between groups, we could use a clustering method to determine the optimum number of splits in the data.

Also, keep in mind this assumes we have a fully-resolved phylogeny. In actuality, the inter-generic relationships of pygopodids are largely equivocal, which is pretty unsatisfying.
We could also easily trim this tree and data to just Aprasia to look at how the genus is split in phenotypic space, for your data, this would likely be along that E/W split.

Let's quickly do the same exercise, but color the nodes by ecology instead, to see if there's a clear pattern.

log transform the data, then extract the columns (traits) of interest

```
pdata[,7:8] <- log(pdata[,7:8]) # traits of interest, you'd use PC1 + PC2
rownames(pdata) <- pdata$Name_in_Tree # provide names for the data
morpho <- pdata[,c("SVL", "Tail_Length")] # keep these traits (PC1 + PC2)
```

Set the factor of interest, color the states

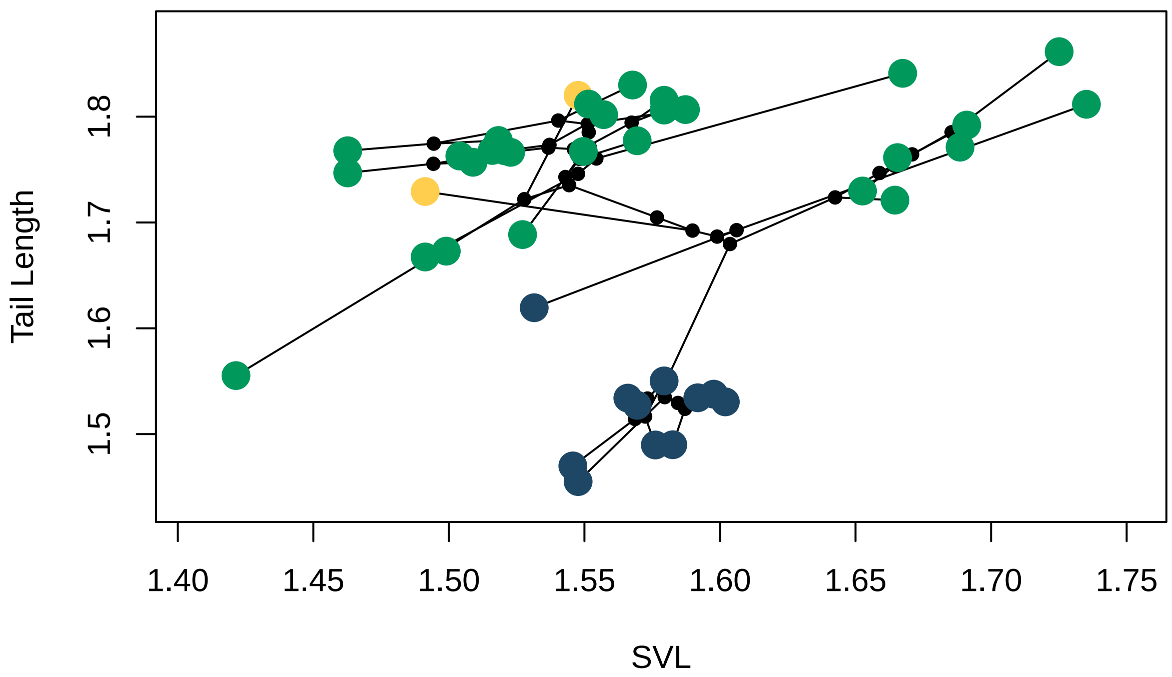
```
ecology <- pdata$Ecology
names(ecology) <- pdata$Name_in_Tree
mycol <- character(length(ecology))
mycol[ecology=="arboreal"] <- mpalette[2]
mycol[ecology=="terrestrial"] <- mpalette[1]
mycol[ecology=="fossorial"] <- mpalette[3]
```

Organize the data and assign colors appropriately

```
pdata$color <- mycol
pdata <- pdata[match(tree$tip.label, pdata$Name_in_Tree),]
group.colors <- pdata$color
names(group.colors) <- 1:Ntip(tree)
nodecols <- rep("black", tree$Nnode)
names(nodecols) <- (Ntip(tree)+1) : (Ntip(tree)+tree$Nnode)
colorz <- c(group.colors, nodecols)
```

Run the visualization

```
phylomorphospace(tree, morpho,
                 label = "off",
                 node.size = c(1, 2),
                 control=list(col.node=colorz),
                 xlab = "SVL", ylab = "Tail Length")
```



If we were taking this more seriously we'd probably want to regress traits like head or tail size against SVL or some measure of length to remove the effect of size, as it's a pretty strong linear relationship between body size and other trait size. Alternatively we could use log-shape ratios, but that's another discussion.

Hopefully this might prove useful.