

# Microbial Innovations

*Mario E. Muscarella, James P. O'Dwyer*

*Last updated on 08 September, 2016*

## Initial Setup

```
# Clear Environment and Set Working Directory
rm(list=ls())
setwd("~/GitHub/microbial-innovations/analyses")

# Add Basic Functions
se <- function(x, ...){sd(x, na.rm = TRUE)/sqrt(length(na.omit(x)))}
ci <- function(x, ...){1.96 * sd(x, na.rm = TRUE)}

# Load Packages
require("ape"); require("png"); require("grid")

## Loading required package: ape
## Loading required package: png
## Loading required package: grid
require("ggtree")

## Loading required package: ggtree
## Loading required package: ggplot2
## 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 visualization and annotation of phylogenetic trees with their covariates and
## Methods in Ecology and Evolution 2016, doi:10.1111/2041-210X.12628
##
## Attaching package: 'ggtree'
## The following object is masked from 'package:ape':
##
##     rotate
require("phytools")

## Loading required package: phytools
## Loading required package: maps
##
## Attaching package: 'phytools'
## The following object is masked from 'package:ggtree':
##
##     reroot
library("colorspace")
```

```

ggplotColours <- function(n = 6, h = c(0, 360) + 15){
  if ((diff(h) %% 360) < 1) h[2] <- h[2] - 360/n
  hcl(h = (seq(h[1], h[2], length = n)), c = 100, l = 65)
}

# Load Source Functions
source("../bin/MothurTools.R")

## Loading required package: reshape

##
## Attaching package: 'reshape'

## The following object is masked from 'package:ggtree':
##
##     expand

# Save Default Plot Settings
opar <- par(no.readonly = TRUE) # Saves plot defaults

```

## Import Tree Files

```

LTP <- read.tree("../data/LTPs123.tree")
NTL <- read.tree("../data/NTL.tree")
RDP <- read.tree("../data/RDP.tree")

```

## Import Taxonomy Information

```

LTP.tax <- read.tax("../data/LTPs123_unique.pds.wang.taxonomy", format = "rdp", col.tax = 2)
NTL.tax <- read.tax("../data/nmicrobiol201648_s7.pds.wang.taxonomy", format = "rdp", col.tax = 2)
RDP.tax <- read.tax("../data/rdp_download_9752seqs.pds.wang.taxonomy", format = "rdp", col.tax = 2)

```

## Edge Length Distribution

```

LTP.edges <- LTP$edge.length
NTL.edges <- NTL$edge.length
RDP.edges <- RDP$edge.length

png(filename="../figures/BranchLengths.png",
     width = 800, height = 1600, res = 96*2)

layout(matrix(1:3, 3, 1))
par(mar = c(5, 6, 3, 1) + 0.1, oma = c(1,1,1,1))

hist(log10(LTP.edges), main = "Silva: Living Tree Project",
     axes = F, xlab = "", ylab = "", xlim = c(-4, 1))

axis(1, at = c(-4, -3, -2, -1, 0, 1),
     labels = expression(10^-4, 10^-3, 10^-2, 10^-1, 1, 10^1),
     las = 1, lwd = 1.5)

```

```

axis(2, labels = T, las = 1, lwd = 1.5)

mtext("Edge Length", side = 1, cex = 1, line = 2.5)
mtext("Frequency", side = 2, cex = 1, line = 4)

hist(log10(NTL.edges), main = "Hug et al. 2016: New Tree of Life",
      axes = F, xlab = "", ylab = "", xlim = c(-4, 1))

axis(1, at = c(-4, -3, -2, -1, 0, 1),
      labels = expression(10^-4, 10^-3, 10^-2, 10^-1, 1, 10^1),
      las = 1, lwd = 1.5)
axis(2, labels = T, las = 1, lwd = 1.5)

mtext("Edge Length", side = 1, cex = 1, line = 2.5)
mtext("Frequency", side = 2, cex = 1, line = 4)

hist(log10(RDP.edges), main = "Ribosomal Database Project",
      axes = F, xlab = "", ylab = "", xlim = c(-4, 1))

axis(1, at = c(-4, -3, -2, -1, 0, 1),
      labels = expression(10^-4, 10^-3, 10^-2, 10^-1, 1, 10^1),
      las = 1, lwd = 1.5)
axis(2, labels = T, las = 1, lwd = 1.5)

mtext("Edge Length", side = 1, cex = 1, line = 2.5)
mtext("Frequency", side = 2, cex = 1, line = 4)

# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/BranchLengths.png")
grid.raster(img)

```

## Edge Length vs Support

```

node.tab <- data.frame(matrix(NA, LTP$Nnode, 5))
colnames(node.tab) <- c("node", "support", "edge_up", "edge_down1", "edge_down2")
node.tab$node <- as.numeric(c(1:LTP$Nnode))
node.tab$support <- LTP$node.label
node.tab[1,3] <- NA
for (i in 2:LTP$Nnode){
  edge.up <- which(LTP$edge[,2] == length(LTP$tip.label) + i)
  node.tab[i,3] <- LTP$edge.length[edge.up]
  edge.down <- which(LTP$edge[,1] == length(LTP$tip.label) + i)
  node.tab[i,4] <- LTP$edge.length[edge.down[1]]
  node.tab[i,5] <- LTP$edge.length[edge.down[2]]
}
node.tab$edge_down_min <- apply(cbind(node.tab$edge_down1, node.tab$edge_down2), 1, min)

```

```

scatterhist = function(x, y, xlab="", ylab=""){
  zones=matrix(c(2,0,1,3), ncol=2, byrow=TRUE)
  par(oma = c(1,1,1,1))
  layout(zones, widths=c(4/5,1/5), heights=c(1/5,4/5))
  xhist = hist(x, plot=FALSE)
  yhist = hist(y, plot=FALSE)
  top = max(c(xhist$counts, yhist$counts))
  par(mar=c(5,5,1,1))
  plot(x,y, las = 1, xlab = "", ylab = "")
  mtext(xlab, side=1, line=3, outer=F, cex = 1.25)
  mtext(ylab, side=2, line=3, outer=F, cex = 1.25)
  par(mar=c(0,5,1,1))
  barplot(xhist$counts, axes=FALSE, space=0)
  par(mar=c(5,0,1,1))
  barplot(yhist$counts, axes=FALSE, space=0, horiz=TRUE)
}

png(filename="../figures/NodeSupport.png",
     width = 800, height = 800, res = 96*2)

# Upstream
scatterhist(x = log10(node.tab$edge_up), y = as.numeric(node.tab$support),
            xlab = "Upstream Edge Length", ylab = "Node Support")

# Downstream
scatterhist(log10(node.tab$edge_down_min), as.numeric(node.tab$support),
            xlab = "Downstream Edge Length", ylab = "Node Support")

mean(as.numeric(node.tab$support), na.rm = T)

## [1] 0.8306474
mean(as.numeric(node.tab$support[log10(node.tab$edge_up) < -1.75]), na.rm = T)

## [1] 0.614402
mean(as.numeric(node.tab$support[log10(node.tab$edge_up) > 0.75]), na.rm = T)

## [1] 1
mean(as.numeric(node.tab$support[log10(node.tab$edge_up) < -2.25]), na.rm = T)

## [1] 0.4839011
# Close Plot Device
dev.off()

## pdf
## 2
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/NodeSupport.png")
grid.raster(img)

```

## LTP Exploratory Tree with Domain Grouping

```
png(filename="../figures/LTP_explore.png",
      width = 1600, height = 1600, res = 96*2)

LTP.2 <- LTP
LTP.tax2 <- LTP.tax[match(LTP.2$tip.label, LTP.tax$OTU), ]
groupInfo <- split(LTP.2$tip.label, LTP.tax2$Domain)
livingTree <- groupOTU(LTP.2, groupInfo, group_name = "Domain")
levels(attributes(livingTree)$Domain) <- names(groupInfo)
ggtree(livingTree, aes(color = Domain), layout="rectangular") +
  geom_text(aes(subset=!isTip, label = node), show.legend = FALSE,
            hjust=-0.1, vjust = -0.5, size = 2) +
  theme(legend.position="top", legend.key = element_rect(colour = NA)) +
  scale_color_manual(values=c(ggplotColours(n = 4)[2:3]))

# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/LTP_explore.png")
grid.raster(img)
```

## Identify the LTP Node that separates bacteria and archaea

```
png(filename="../figures/LTP_archaea.png",
      width = 1600, height = 1600, res = 96*2)
par(mar=c(1,1,1,3))
cp <- ggtree(livingTree, aes(color = Domain), layout="rectangular") %>% collapse(node=18228)
cp + geom_point2(aes(subset=(node == 18227)), size=3, shape=18,
                 colour=ggplotColours(n = 4)[1], fill = "gray80") +
  theme(legend.position="top", legend.key = element_rect(colour = NA)) +
  geom_text2(aes(subset=(node == 18228), label = "Archaea"),
            show.legend = FALSE, hjust=-0.1, size = 3) +
  guides(colour = guide_legend(override.aes = list(shape = NA, label = NA))) +
  scale_color_manual(values=c(ggplotColours(n = 4)[2:3]))

# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/LTP_archaea.png")
grid.raster(img)
```

## Reroot LTP Tree

```
png(filename="../figures/LTP_root.png",
     width = 1600, height = 1600, res = 96*2)

livingTree.r <- midpoint.root(livingTree)
livingTree.r2 <- group0TU(livingTree.r, groupInfo, group_name = "Domain")
levels(attributes(livingTree.r2)$Domain) <- c("root", names(groupInfo))
rt <- ggtree(livingTree.r2, aes(color = Domain), layout="rectangular")
rt + geom_point2(aes(subset=(node == 11934)), size=4, shape=18,
                colour=ggplotColours(n = 4)[1], fill = "gray80") +
  theme(legend.position="top", legend.key = element_rect(colour = NA)) +
  scale_color_manual(values=c(ggplotColours(n = 4)[1:3]))

# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/LTP_root.png")
grid.raster(img)
```

## Export Rooted Tree

```
write.tree(livingTree.r, file = "../data/LTP.rooted.tree")
```

## NTL Exploratory Tree with Domain Grouping

```
png(filename="../figures/NTL_explore.png",
     width = 1600, height = 1600, res = 96*2)

NTL.2 <- NTL
NTL.tax2 <- NTL.tax[na.omit(match(NTL.2$tip.label, NTL.tax$OTU)), ]
NTL.tax2[which(NTL.tax2$Domain == "unknown"), ]$Domain <- "Bacteria"
groupInfo <- split(NTL.2$tip.label, NTL.tax2$Domain)
newTree <- group0TU(NTL.2, groupInfo, group_name = "Domain")
levels(attributes(newTree)$Domain) <- c("root", names(groupInfo))
newTree$node.label[which(attributes(newTree)$Domain == "root")]
ggtree(newTree, aes(color = Domain), layout="rectangular") +
  geom_text(aes(subset=!isTip, label = node), show.legend = FALSE,
            hjust=-0.1, vjust = -0.5, size = 2) +
  theme(legend.position="top", legend.key = element_rect(colour = NA))

# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
```

```
img <- readPNG("../figures/NTL_explore.png")
grid.raster(img)
```

Remove Eukaryota and Identify the NTL Node that separates bacteria and archaea

```
png(filename="../figures/NTL_archaea.png",
     width = 1600, height = 1600, res = 96*2)
par(mar=c(1,1,1,3))
NTL.2 <- drop.tip(NTL, c(NTL.tax2$OTU[which(NTL.tax2$Domain == "Eukaryota")]))
NTL.tax3 <- NTL.tax2[-which(NTL.tax2$Domain == "Eukaryota"), ]
groupInfo <- split(NTL.2$tip.label, NTL.tax3$Domain)
newTree.2 <- groupOTU(NTL.2, groupInfo, group_name = "Domain")
cp <- ggtree(newTree.2, aes(color = Domain), layout="rectangular") %>% collapse(node=2420)
cp + geom_point2(aes(subset=(node == 18227)), size=3, shape=23,
                 colour="black", fill = "gray80") +
  theme(legend.position="top", legend.key = element_rect(colour = NA)) +
  geom_text2(aes(subset=(node == 18228), label = "Archaea"),
             show.legend = FALSE, hjust=-0.1, size = 3) +
  geom_text(aes(subset=!isTip, label = node), show.legend = FALSE,
            hjust=-0.1, vjust = -0.5, size = 2) +
  guides(colour = guide_legend(override.aes = list(shape = NA, label = NA)))
```

```
## Warning: Removed 351 rows containing missing values (geom_text).
```

```
# Close Plot Device
dev.off()
graphics.off()
#```
## Show Plot
#```{r}
img <- readPNG("../figures/NTL_archaea.png")
grid.raster(img)
```