Global STP Analysis - Stylophora pistillata

Matt

November 16, 2014

Load required libraries

```
library("phyloseq")
```

```
## Warning in fun(libname, pkgname): couldn't connect to di
```

```
library("ggplot2")
library("plyr")
library("vegan")
library("grid")
library("directlabels")
library("knitr")
library("clustsig")
library("ellipse")
setwd("./data")
opts_knit$set(root.dir = "./data", fig.keep='high')
opts_chunk$set(dev = 'pdf', fig.path="figures/spist/")
```

Import data into R studio

First .shared 'otu matrix' from mothur

```
sharedFile = read.table("micro.final.shared")
sharedFile = t(sharedFile)
rownames(sharedFile) = sharedFile[,1]
colnames(sharedFile) = sharedFile[2,]
sharedFile = sharedFile[,2:234]
sharedFile = sharedFile[4:37368,]
class(sharedFile) <- "numeric"
```

Import subsampled otu matrix (7779 seqs)

```
sharedsubFile = read.table('micro.final.clean.0.03.subsampl
sharedsubFile = t(sharedsubFile)
rownames(sharedsubFile) = sharedsubFile[,1]
colnames(sharedsubFile) = sharedsubFile[2,]
sharedsubFile = sharedsubFile[,2:219]
sharedsubFile = sharedsubFile[4:14839,]
class(sharedsubFile) <- "numeric"
```

Create phyloseq object

```
OTU = otu_table(sharedFile, taxa_are_rows = TRUE)
OTUsub = otu_table(sharedsubFile, taxa_are_rows = TRUE)
TAX = tax_table(taxFile)
META = sample_data(metaFile)
physeq = phyloseq(OTU, TAX, META)
physeqSub = phyloseq(OTUsub, TAX, META)
physeqSub
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 14836 taxa and 214 sa
## sample_data() Sample Data:       [ 214 samples by 5 samp
## tax_table()   Taxonomy Table:    [ 14836 taxa by 7 taxon
```

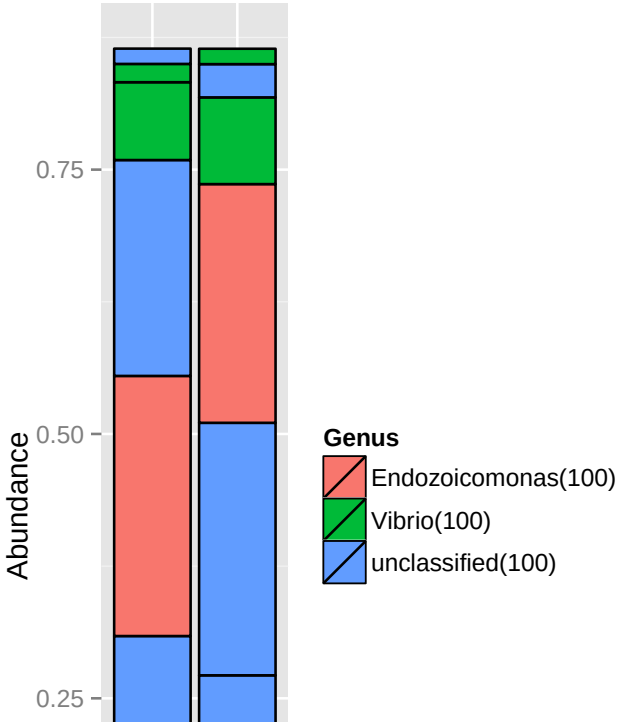
Get rid of any OTUs not present in any samples and get relative abundance

```
microSub <- prune_taxa(taxa_sums(physeqSub) > 0, physeqSub)
microSubRel = transform_sample_counts(microSub, function(x)
```

Stylophora pistillata from aquaria

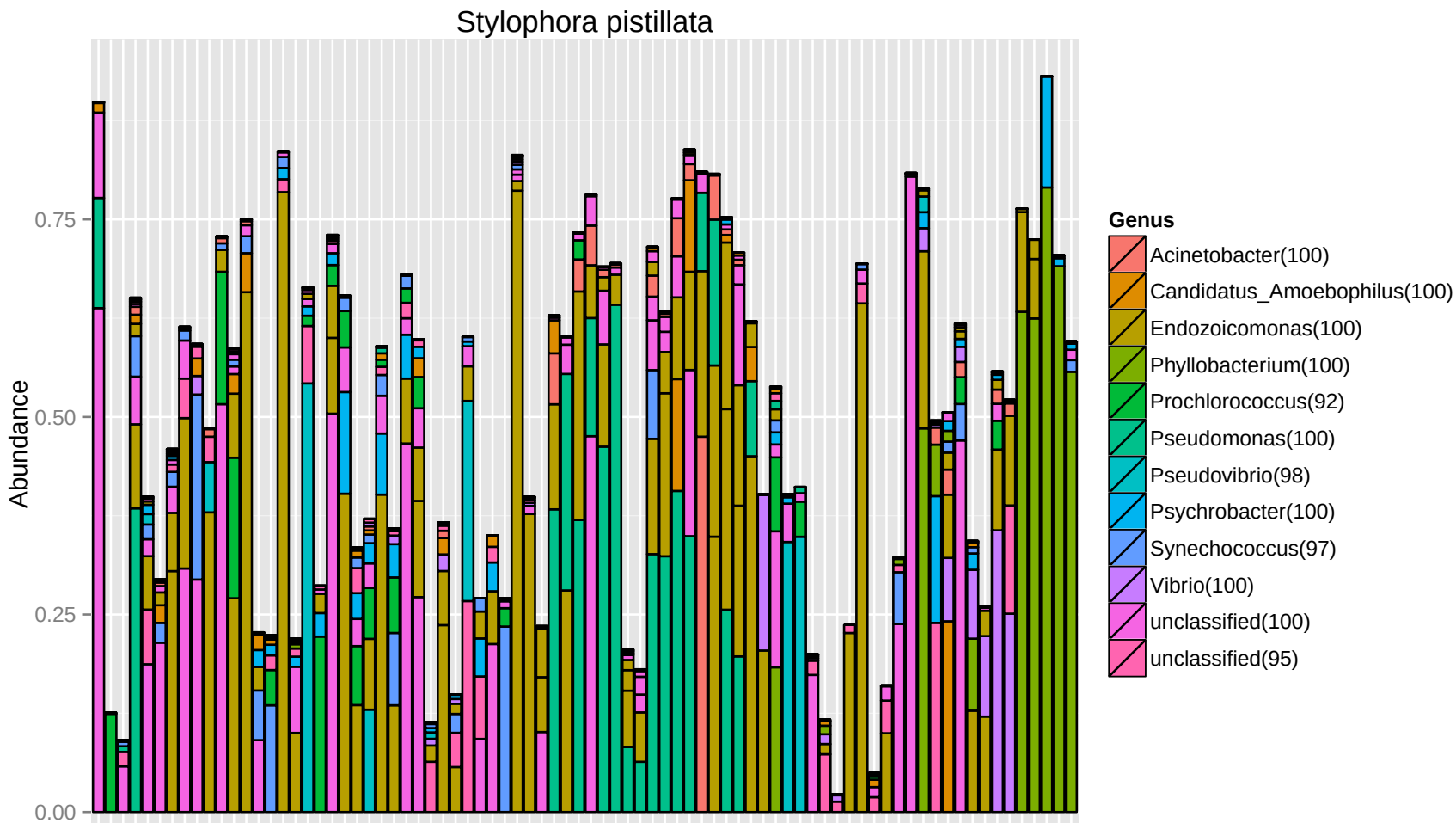
Take a look at the bacteria classified to genus and to family. NOTE: have to do this on non-subsampled data as these aqarium samples get removed in this step because of low sequence numbers (~3000)

```
spistAq = prune_samples(c("SPaq", "SPaq-rep2"), physeq)
spistAqrelAbund = transform_sample_counts(spistAq, functio
spistAqrelAbundFilt = filter_taxa(spistAqrelAbund, functio
plot_bar(spistAqrelAbundFilt, fill="Genus")
```



Relative abundance of otus in all Stylophora pistillata samples

```
microSubRelFiltSpistFilt = filter_taxa(microSubRelFiltSpist
spistBar <- plot_bar(microSubRelFiltSpistFilt, fill="Genus"
spistBar
```



Ordination of Stylophora pistillata samples

Which distance measure to use?

```
dist_methods <- unlist(distance("list"))
dist_methods <- dist_methods[-(1:2)]
dist_methods <- dist_methods[-which(dist_methods == "ANY")]

plist <- vector("list", length(dist_methods))
names(plist) = dist_methods
for (j in dist_methods) {
  iDist <- distance(microSubRelFiltSpistFilt, method = j)
  iMDS <- ordinate(microSubRelFiltSpistFilt, "MDS", distanc
  p <- NULL
  p <- plot_ordination(microSubRelFiltSpistFilt, iMDS, colo
  p <- p + ggtitle(paste("MDS using distance method ", j, s
  plist[[j]] = p
}

df = ldply(plist, function(x) x$data)
```

Correlations between chemical data and microbes

Import nutrient and FCM data and look at Spearman correlations between this data and the sample ordinations

```
Hbact <- as.data.frame(read.table("meta.Hbact.spearman.corr
NH4 <- as.data.frame(read.table("meta.NH4.spearman.corr.axe
NN <- as.data.frame(read.table("meta.NN.spearman.corr.axes"
NO2 <- as.data.frame(read.table("meta.NO2.spearman.corr.axe
PEPeuk <- as.data.frame(read.table("meta.PE+Peuk.spearman.c
Peuk <- as.data.frame(read.table("meta.PEUK.spearman.corr.a
po4 <- as.data.frame(read.table("meta.po4.spearman.corr.axe
PRO <- as.data.frame(read.table("meta.PRO.spearman.corr.axe
silicate <- as.data.frame(read.table("meta.silicate.spearma
SYN <- as.data.frame(read.table("meta.SYN.spearman.corr.axe

nutrients <- rbind(NH4, NN, NO2, po4, silicate)
FCM <- rbind(Hbact, PEPeuk, Peuk, PRO, SYN)

arrowmatrix = nutrients
```

SIMPROF analysis to check which samples fall into 'groups' without any *a priori* assumptions

Need to import the shared file containing just spist OTUs, then calcualte the simprof clusters based on the braycurtis metric.

```
spistShared = read.table('micro.spist.0.03.shared')
spistShared = t(spistShared)
rownames(spistShared) = spistShared[,1]
colnames(spistShared) = spistShared[2,]
spistShared = spistShared[,2:81]
spistShared = spistShared[4:4813,]
class(spistShared) <- "numeric"

spistSIMPROF <- simprof(spistShared, num.expected=10, num.s

simprof.plot(spistSIMPROF, leafcolors=NA, plot=TRUE, fill=T
```