coralMicrobiomes

Matthew J. Neave

April 22, 2015

This is an R Markdown document detailing the statistical and graphical steps for reproducting the results in:

Neave, M.J., Rachmawati, R., Xun, L., Michell, C.T., Barber, P.H., Bourne, D.G., McCulloch, M.T., Apprill, A., Voolstra, C.R. A global microbiome analysis reveals that closely related corals exhibit fine-scale differences in their association with Endozoicomonas symbionts.

Load required libraries

```
library("phyloseq"); packageVersion("phyloseq")
```

```
## [1] '1.9.15'
```

```
library("ggplot2"); packageVersion("ggplot2")
```

```
## [1] '1.0.0'
```

```
library("plyr"); packageVersion("plyr")
```

```
## [1] '1.8.1'
```

```
library("vegan"); packageVersion("vegan")
```

```
## [1] '2.0.10'
```

```
library("grid"); packageVersion("grid")
```

```
## [1] '3.1.1'
```

```
library("knitr"); packageVersion("knitr")
```

```
## [1] '1.6'
```

```
library("clustsig"); packageVersion("clustsig")
```

```
## Warning: package 'clustsig' was built under R version 3.1.2
```

```
## [1] '1.1'
```

```
library('ape'); packageVersion("ape")
```

```
## [1] '3.1.4'
```

```
library('RColorBrewer'); packageVersion("RColorBrewer")
```

```
## [1] '1.0.5'
```

```
library("dunn.test"); packageVersion("dunn.test")
```

```
## Warning: package 'dunn.test' was built under R version 3.1.2
```

```
## [1] '1.2.3'
```

```
library("DESeq2"); packageVersion("DESeq2")
```

```
## Warning: package 'RcppArmadillo' was built under R version 3.1.2
```

```
## [1] '1.4.5'
```

```
setwd("./data")
opts_knit$set(root.dir = "./data")
```

Import data

First the matrix percent file generated by the minimum entropy decomposition (MED) pipeline, subsampled to 7974 reads per sample, and the accociated taxonony file

```
allShared = read.table("all.7974.matrixPercent.txt", header=T, row.names=1)
allTax = read.table('all.7974.nodeReps.nr_v119.knn.taxonomy', header=T, sep='\t', row.names = 1)
```

```
## Warning: number of items read is not a multiple of the number of columns
```

```
allTax = allTax[,2:8]
allTax = as.matrix(allTax)
```

Import the shared and taxonomy files generated in mothur for 3% and 1% pairwise similarity, in order to calculate alpha diversity measures and to compare to the MED procedure. Also import the 3% OTU file without any subsampling for alpha diversity calculations.

```
all3OTUshared = read.table("all.7974.0.03.pick.shared", header=T, row.names=2)
all3OTUshared = all3OTUshared[,3:length(all3OTUshared)]

alpha3OTUshared = read.table("all.7974.0.03.shared", header=T)
rownames(alpha3OTUshared) = alpha3OTUshared[,2]
alpha3OTUshared = alpha3OTUshared[,4:length(alpha3OTUshared)]
```

```
all1OTUshared = read.table("all.7974.0.01.pick.shared", header=T, row.names=2)
all1OTUshared = all1OTUshared[,3:length(all1OTUshared)]

all3OTUtax = read.table('all.7974.0.03.taxonomy', header=T, sep='\t', row.names=1)
all3OTUtax = all3OTUtax[,2:8]
all3OTUtax = as.matrix(all3OTUtax)

all1OTUtax = read.table('all.7974.0.01.taxonomy', header=T, sep='\t', row.names=1)
all1OTUtax = all1OTUtax[,2:8]
all1OTUtax = as.matrix(all1OTUtax)
```

Import Endozoicomonas phylogenetic tree (exported from ARB) using the APE package (Fig. 3). Also import a MED percent matrix that is slightly modified to accomodate the tree

```
endoTreeFile = read.tree(file='MEDNJ5.tree')
allSharedTree = read.table("all.7974.matrixPercent.tree.txt", header=T, row.names=1)
```

Import meta data for the samples, including metaData3.txt, which is slightly modified to accomodate heatmap sample ordering, and metaDataChem which contains additional columns of physiochemical data

```
metaFile = read.table('metaData2.MED', header=T, sep='\t', row.names=1)
metaFile3 = read.table('metaData3.txt', header=T, sep='\t', row.names=1)
metaFileChem = read.table('metaDataChem.txt', header=T, sep='\t', row.names=1)
```

Create phyloseq objects and add consistent coloring for sites

```
OTU = otu_table(allShared, taxa_are_rows = FALSE)
OTUs3 = otu_table(all3OTUshared, taxa_are_rows = FALSE)
OTUs3alpha = otu_table(alpha3OTUshared, taxa_are_rows = FALSE)
OTUs1 = otu_table(all1OTUshared, taxa_are_rows = FALSE)
OTUtree = otu_table(allSharedTree, taxa_are_rows = FALSE)

TAX = tax_table(allTax)
TAX3 = tax_table(all3OTUtax)
TAX1 = tax_table(all1OTUtax)

META = sample_data(metaFile)
METAchem = sample_data(metaFileChem)
TREE = phy_tree(endoTreeFile)

allPhylo = phyloseq(OTU, TAX, META)
all3OTUphylo = phyloseq(OTUs3, TAX3, META)
alpha3OTUphylo = phyloseq(OTUs3alpha, META)
all1OTUphylo = phyloseq(OTUs1, TAX1, META)
allPhyloChem = phyloseq(OTU, TAX, METAchem)
endoTree = phyloseq(OTUtree, META, TREE)

cols <- c("AmericanSamoa" = "#D95F02", "Indonesia" = "#A6761D", "MaggieIs" = "#666666", "Maldives" = "#
```

Ordinations to compare MED vs pairwise OTUs

Subset samples for the two corals, remove taxa with 0s, create relative abundance and square-root sample counts

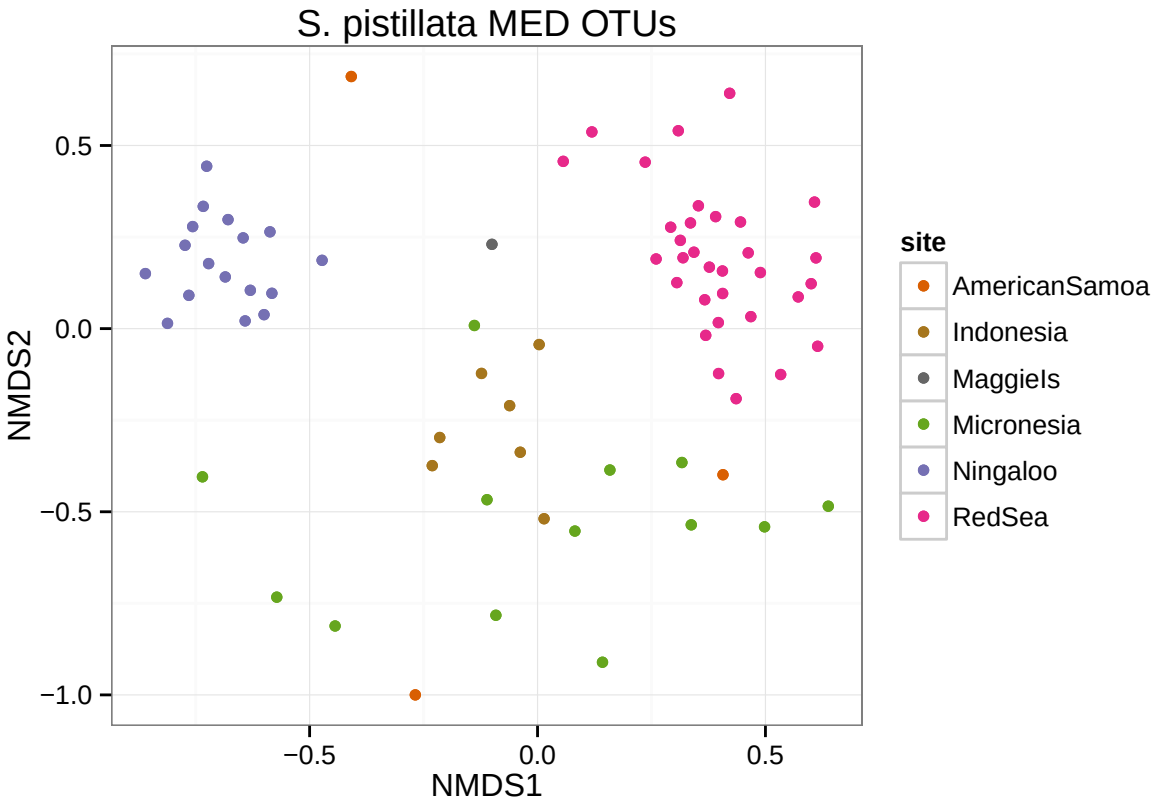
```
filter_stylo_data <- function(initial_matrix){
  initial_coral <- subset_samples(initial_matrix, species=="Stylophora pistillata")
  coral_filt = filter_taxa(initial_coral, function(x) mean(x) > 0, TRUE)
  coral_filt_rel = transform_sample_counts(coral_filt, function(x) x / sum(x) )
  coral_filt_rel_sqrt = transform_sample_counts(coral_filt_rel, function(x) sqrt(x) )
  return(coral_filt_rel_sqrt)
}

filter_pverr_data <- function(initial_matrix){
  initial_coral <- subset_samples(initial_matrix, species=="Pocillopora verrucosa")
  coral_filt = filter_taxa(initial_coral, function(x) mean(x) > 0, TRUE)
  coral_filt_rel = transform_sample_counts(coral_filt, function(x) x / sum(x) )
  coral_filt_rel_sqrt = transform_sample_counts(coral_filt_rel, function(x) sqrt(x) )
  return(coral_filt_rel_sqrt)
}

spistPhyloRelSqrt <- filter_stylo_data(allPhylo)
spist3OTUphyloRelSqrt <- filter_stylo_data(all3OTUphylo)
spist1OTUphyloRelSqrt <- filter_stylo_data(all1OTUphylo)

pverrPhyloRelSqrt <- filter_pverr_data(allPhylo)
pverr3OTUphyloRelSqrt <- filter_pverr_data(all3OTUphylo)
pverr1OTUphyloRelSqrt <- filter_pverr_data(all1OTUphylo)
```

Now do ordinations for each

```
compOrdinations <- function(sample_data, sample_name){
theme_set(theme_bw())
sample_dataOrd <- ordinate(sample_data, "NMDS", "bray")
plot_ordination(sample_data, sample_dataOrd, type = 'samples', color='site', title=sample_name) +
  geom_point(size=2) +
  scale_color_manual(values=cols)
}

compOrdinations(spistPhyloRelSqrt, "S. pistillata MED OTUs")
```

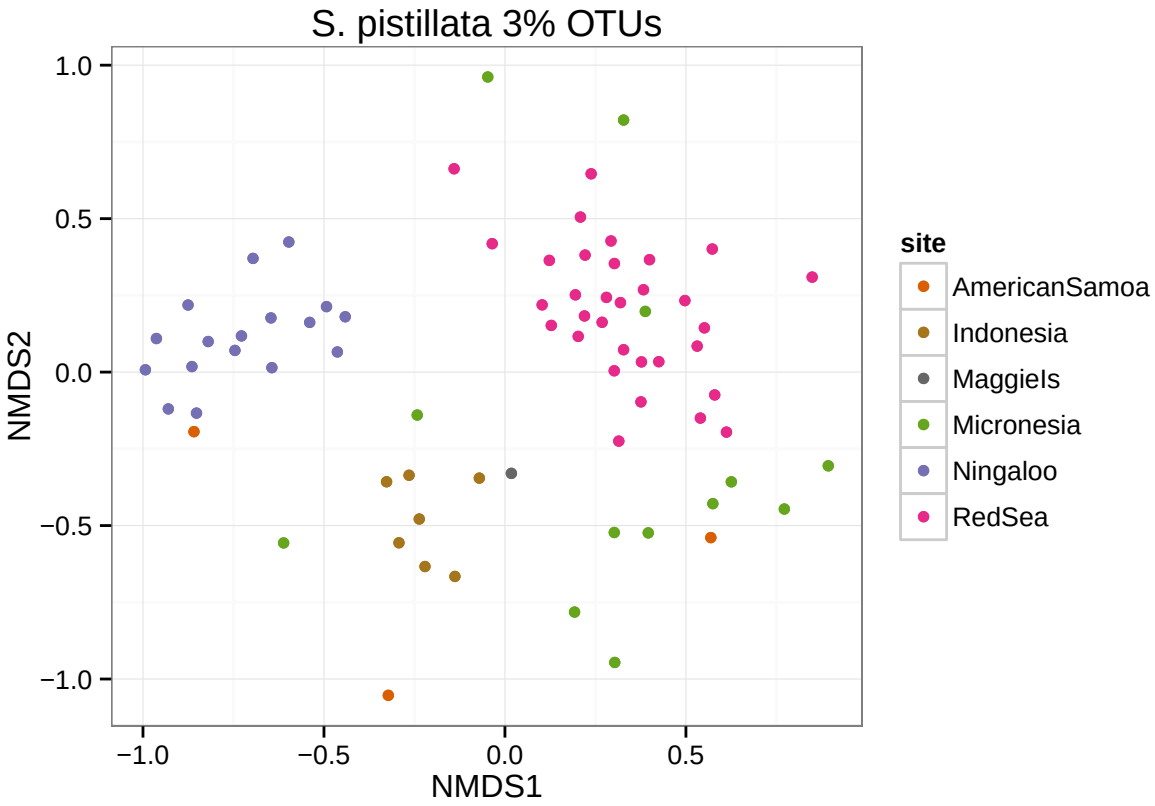
```
## Run 0 stress 0.2254
## Run 1 stress 0.2459
## Run 2 stress 0.2413
## Run 3 stress 0.2399
## Run 4 stress 0.2389
## Run 5 stress 0.2373
## Run 6 stress 0.2344
## Run 7 stress 0.2328
## Run 8 stress 0.2371
## Run 9 stress 0.229
## Run 10 stress 0.2273
## Run 11 stress 0.2341
```

```
## Run 12 stress 0.2302
## Run 13 stress 0.2352
## Run 14 stress 0.2288
## Run 15 stress 0.2503
## Run 16 stress 0.2381
## Run 17 stress 0.2486
## Run 18 stress 0.2409
## Run 19 stress 0.2242
## ... New best solution
## ... procrustes: rmse 0.02453  max resid 0.1974
## Run 20 stress 0.2292
```



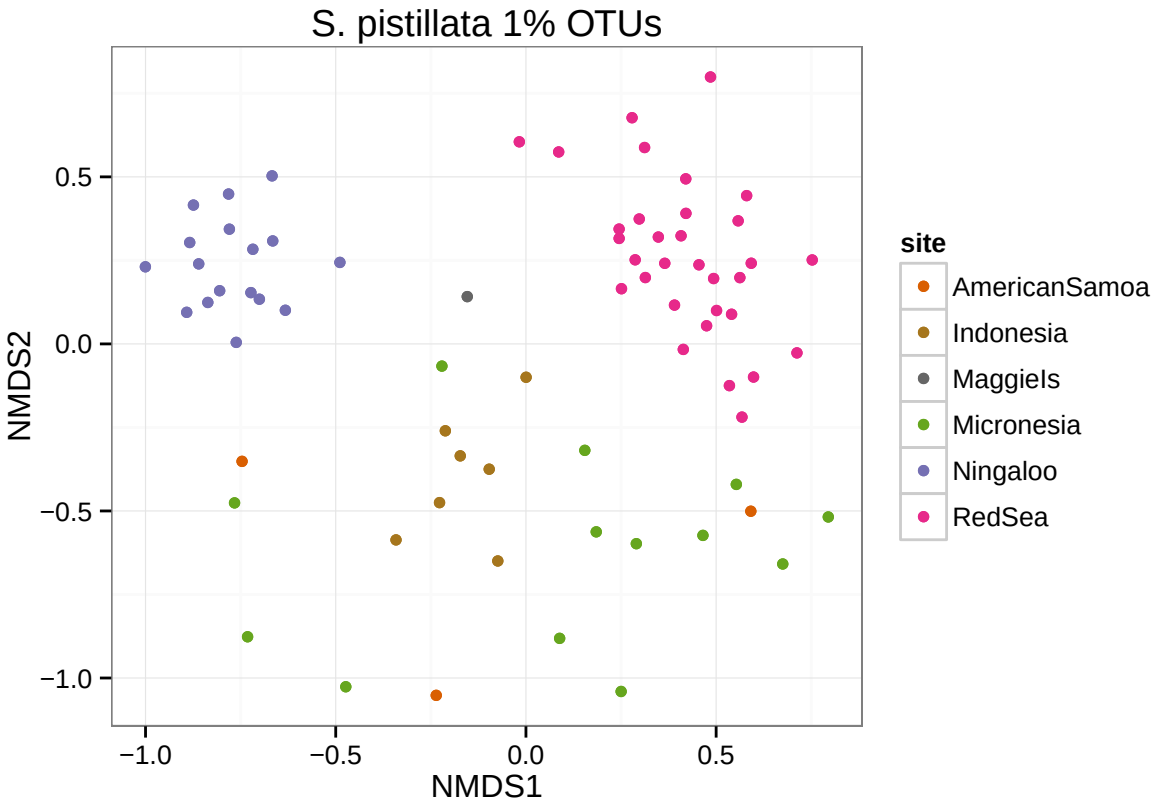

```
compOrdinations(spist3OTUphyloRelSqrt, "S. pistillata 3% OTUs")
```

```
## Run 0 stress 0.2273
## Run 1 stress 0.235
## Run 2 stress 0.248
## Run 3 stress 0.2386
## Run 4 stress 0.2395
## Run 5 stress 0.2663
## Run 6 stress 0.2542
## Run 7 stress 0.2273
## ... New best solution
## ... procrustes: rmse 0.0004645  max resid 0.003106
## *** Solution reached
```




```
compOrdinations(spist1OTUphyloRelSqrt, "S. pistillata 1% OTUs")
```

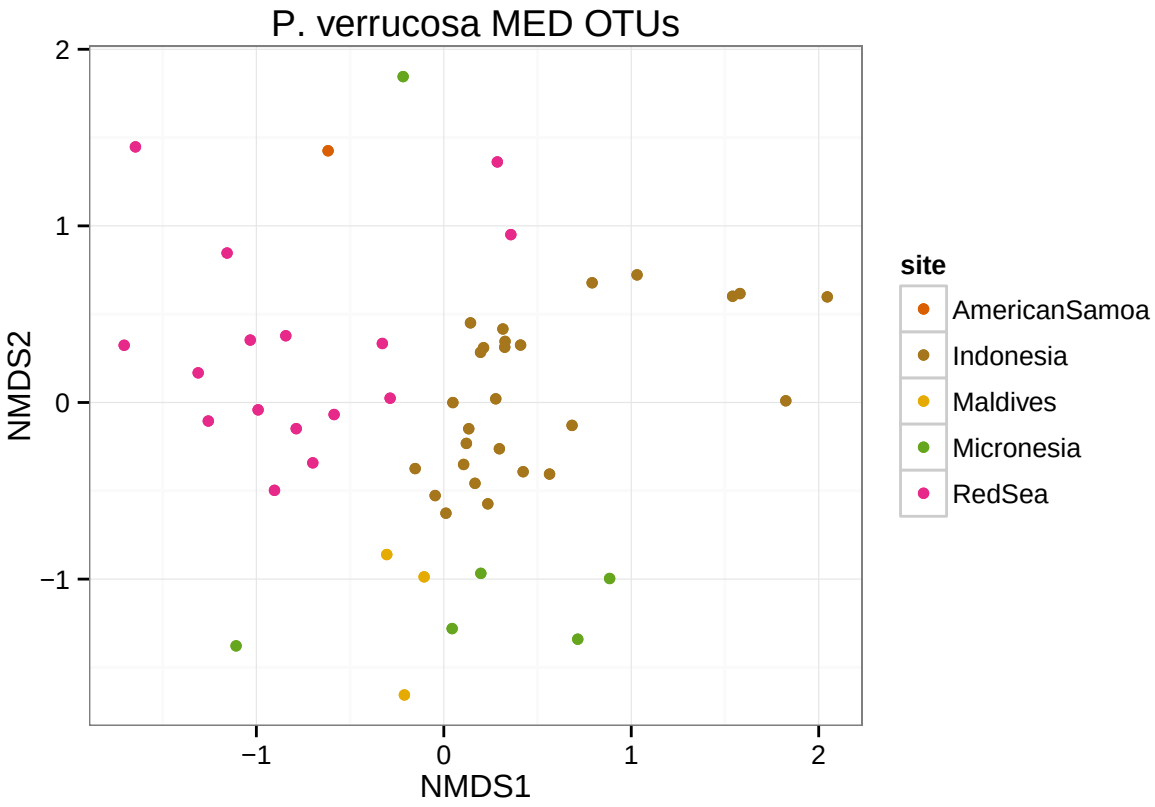
```
## Run 0 stress 0.2225
## Run 1 stress 0.2575
## Run 2 stress 0.2512
## Run 3 stress 0.2274
## Run 4 stress 0.2645
## Run 5 stress 0.2512
## Run 6 stress 0.2366
## Run 7 stress 0.2274
## Run 8 stress 0.2351
## Run 9 stress 0.2532
## Run 10 stress 0.2304
## Run 11 stress 0.2582
## Run 12 stress 0.2338
## Run 13 stress 0.2318
## Run 14 stress 0.2331
## Run 15 stress 0.2399
## Run 16 stress 0.2586
## Run 17 stress 0.2571
## Run 18 stress 0.2417
## Run 19 stress 0.2287
## Run 20 stress 0.2559
```




```
compOrdinations(pverrPhyloRelSqrt, "P. verrucosa MED OTUs")
```

```
## Run 0 stress 0.2437
## Run 1 stress 0.2485
## Run 2 stress 0.271
## Run 3 stress 0.2362
## ... New best solution
## ... procrustes: rmse 0.09863  max resid 0.3381
## Run 4 stress 0.2446
## Run 5 stress 0.2349
## ... New best solution
## ... procrustes: rmse 0.07606  max resid 0.353
## Run 6 stress 0.2133
## ... New best solution
## ... procrustes: rmse 0.08037  max resid 0.3474
## Run 7 stress 0.2259
## Run 8 stress 0.2227
## Run 9 stress 0.2438
## Run 10 stress 0.2136
## ... procrustes: rmse 0.02494  max resid 0.146
## Run 11 stress 0.2247
## Run 12 stress 0.2454
## Run 13 stress 0.2186
## Run 14 stress 0.2393
## Run 15 stress 0.2393
## Run 16 stress 0.2206
## Run 17 stress 0.2315
## Run 18 stress 0.2264
```

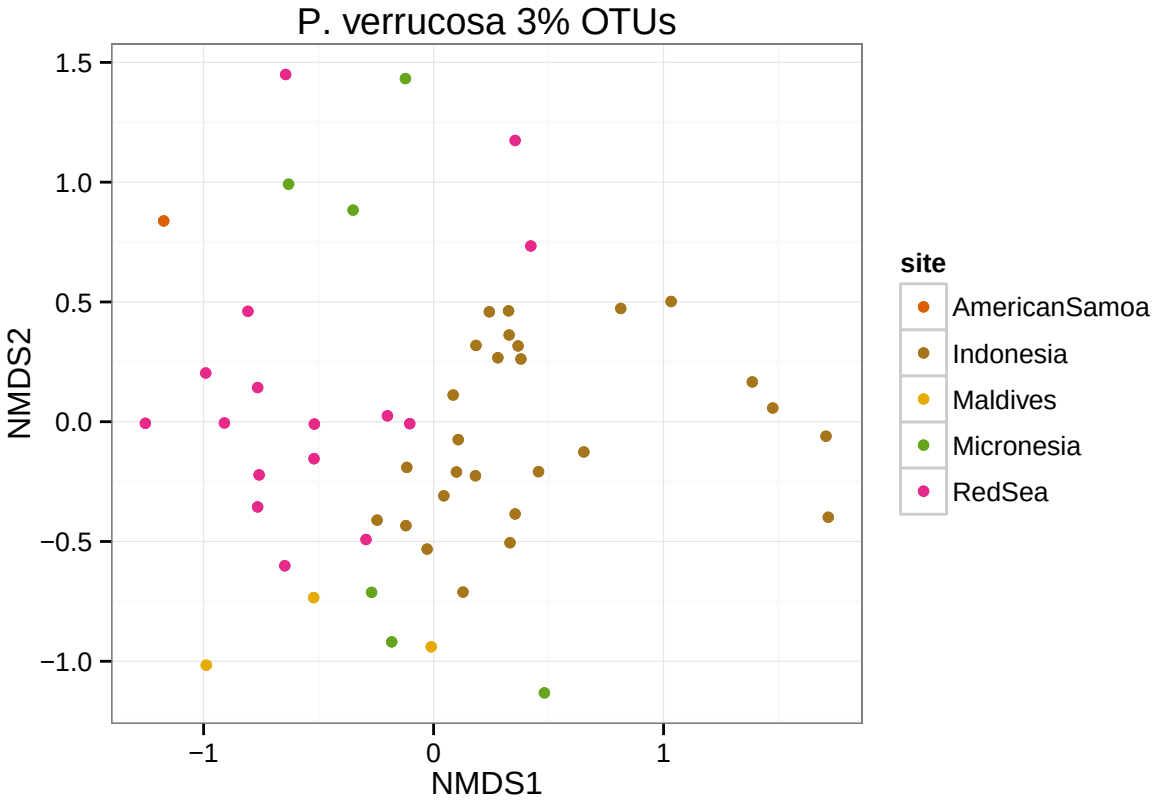
```
## Run 19 stress 0.2572
## Run 20 stress 0.239
```




```
compOrdinations(pverr3OTUphyloRelSqrt, "P. verrucosa 3% OTUs")
```

```
## Run 0 stress 0.2409
## Run 1 stress 0.2354
## ... New best solution
## ... procrustes: rmse 0.09003  max resid 0.3035
## Run 2 stress 0.2482
## Run 3 stress 0.2598
## Run 4 stress 0.2349
## ... New best solution
## ... procrustes: rmse 0.09954  max resid 0.4044
## Run 5 stress 0.23
## ... New best solution
## ... procrustes: rmse 0.1085  max resid 0.3136
## Run 6 stress 0.2283
## ... New best solution
## ... procrustes: rmse 0.09872  max resid 0.2776
## Run 7 stress 0.224
## ... New best solution
## ... procrustes: rmse 0.0909  max resid 0.2662
## Run 8 stress 0.2289
## Run 9 stress 0.2407
## Run 10 stress 0.2288
## Run 11 stress 0.2238
```

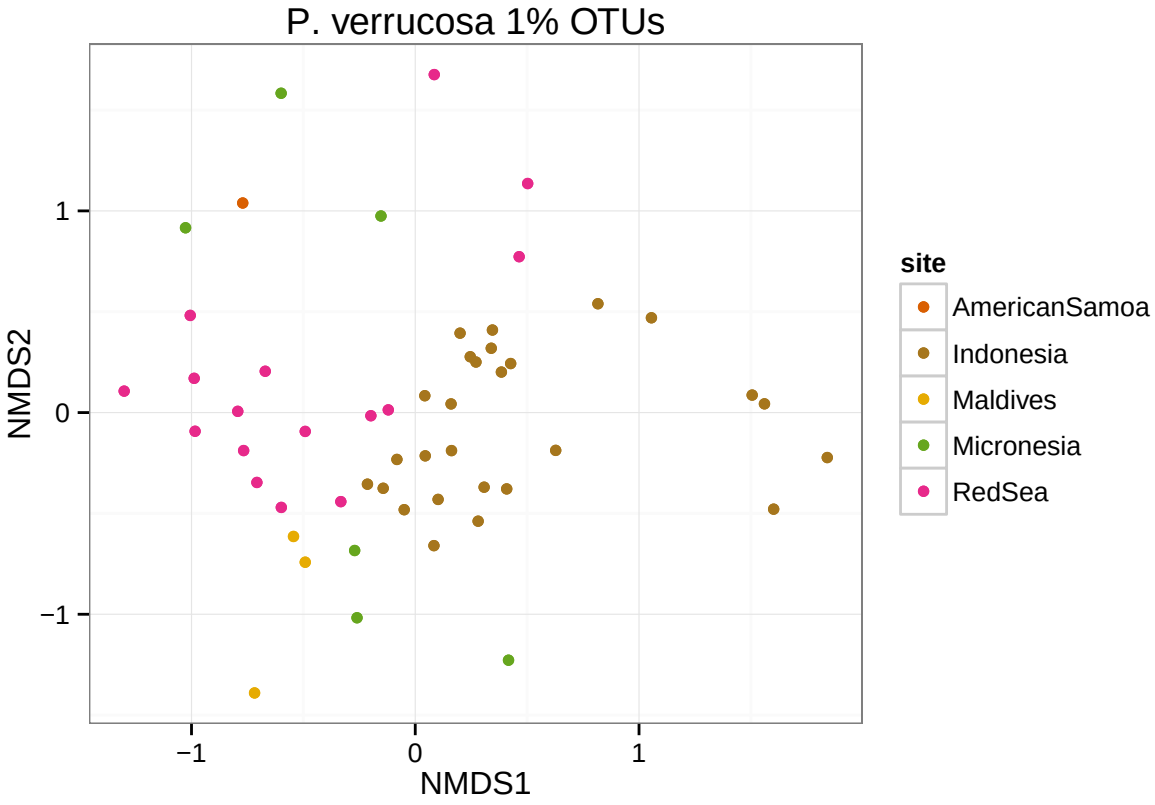
```
## ... New best solution
## ... procrustes: rmse 0.04183  max resid 0.2614
## Run 12 stress 0.2593
## Run 13 stress 0.2852
## Run 14 stress 0.2277
## Run 15 stress 0.227
## Run 16 stress 0.2358
## Run 17 stress 0.2468
## Run 18 stress 0.2353
## Run 19 stress 0.2316
## Run 20 stress 0.4043
```




```
compOrdinations(pverr1OTUphyloRelSqrt, "P. verrucosa 1% OTUs")
```

```
## Run 0 stress 0.234
## Run 1 stress 0.2289
## ... New best solution
## ... procrustes: rmse 0.08099  max resid 0.2296
## Run 2 stress 0.2458
## Run 3 stress 0.217
## ... New best solution
## ... procrustes: rmse 0.09259  max resid 0.4139
## Run 4 stress 0.2297
## Run 5 stress 0.2492
## Run 6 stress 0.2473
## Run 7 stress 0.228
## Run 8 stress 0.2341
```

```
## Run 9 stress 0.2263
## Run 10 stress 0.2306
## Run 11 stress 0.2187
## Run 12 stress 0.2263
## Run 13 stress 0.2331
## Run 14 stress 0.2356
## Run 15 stress 0.2355
## Run 16 stress 0.2172
## ... procrustes: rmse 0.04413  max resid 0.297
## Run 17 stress 0.2188
## Run 18 stress 0.2283
## Run 19 stress 0.2181
## Run 20 stress 0.2322
```



Alpha diversity measures

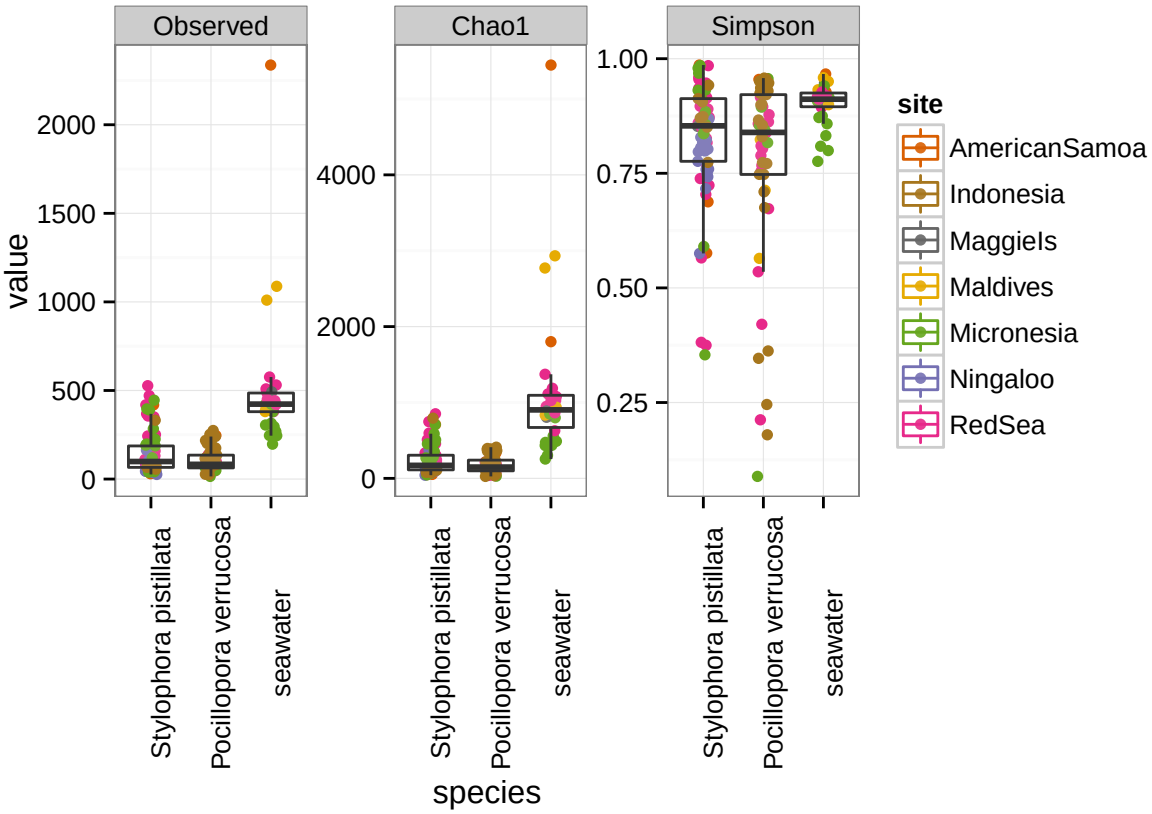
First subset the corals, then plot using phyloseq and ggplot2

Note: I'll use unsubampled 3% pairwise OTUs for calculation of alpha diversity measures as this will make them more comparable to other studies, plus the MED pipeline is has not yet implemented alpha diversity

```
allAlphaTmp <- subset_samples(alpha3OTUphylo, species=="seawater")
allAlphaTmp2 <- subset_samples(alpha3OTUphylo, species=="Stylophora pistillata")
allAlphaTmp3 <- subset_samples(alpha3OTUphylo, species=="Pocillopora verrucosa")
allAlpha2 <- merge_phyloseq(allAlphaTmp, allAlphaTmp2, allAlphaTmp3)

allAlphaPlot2 <- plot_richness(allAlpha2, x = 'species', measures = c('Chao1', 'Simpson', 'observed'),
```

```
ggplot(data = allAlphaPlot2$data) +
  geom_point(aes(x = species, y = value, color = site), position=position_jitter(width=0.1, height=0)) +
  geom_boxplot(aes(x = species, y = value, color = NULL), alpha = 0.1, outlier.shape = NA) +
  scale_color_manual(values=cols) +
  theme(axis.text.x = element_text(angle = 90)) +
  facet_wrap(~variable, scales='free_y') +
  scale_x_discrete(limits=c("Stylophora pistillata", "Pocillopora verrucosa", "seawater"))
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 7 rows containing missing values (geom_point).
## Warning: Removed 4 rows containing missing values (geom_point).
## Warning: Removed 6 rows containing missing values (geom_point).
## Warning: Removed 4 rows containing missing values (geom_point).
## Warning: Removed 7 rows containing missing values (geom_point).
## Warning: Removed 4 rows containing missing values (geom_point).
## Warning: Removed 4 rows containing missing values (geom_point).
```



Check for significant differences in the alpha diversity measures using a kruskal-wallis test and a dunn post-hoc test to check which specific groups were different

```
alphaObserved = estimate_richness(allAlpha2, measures="Observed")
alphaSimpson = estimate_richness(allAlpha2, measures="Simpson")
alphaChao = estimate_richness(allAlpha2, measures="Chao1")

alpha.stats <- cbind(alphaObserved, sample_data(allAlpha2))
alpha.stats2 <- cbind(alpha.stats, alphaSimpson)
alpha.stats3 <- cbind(alpha.stats2, alphaChao)
```

```r
kruskal.test(Observed~species, data = alpha.stats3)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Observed by species
## Kruskal-Wallis chi-squared = 61.88, df = 2, p-value = 3.662e-14
```

```r
dunn.test(alpha.stats3$Observed, alpha.stats3$species, method="bonferroni")
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 61.8764, df = 2, p-value = 0
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   Pocillop   seawater
## ---------+----------------------
## seawater |   7.510384
##          |     0.0000
##          |
## Stylopho |   1.357184  -6.783011
##          |     0.2621     0.0000
```

```r
kruskal.test(Simpson~species, data = alpha.stats3)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Simpson by species
## Kruskal-Wallis chi-squared = 12.25, df = 2, p-value = 0.002193
```

```r
dunn.test(alpha.stats3$Simpson, alpha.stats3$species, method="bonferroni")
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 12.2453, df = 2, p-value = 0
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   Pocillop   seawater
## ---------+----------------------
## seawater |   3.397898
##          |     0.0010
```

```
##          |
## Stylopho |   0.811204  -2.904738
##          |     0.6259     0.0055
```

```
kruskal.test(Chao1~species, data = alpha.stats3)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Chao1 by species
## Kruskal-Wallis chi-squared = 64.31, df = 2, p-value = 1.086e-14
```

```
dunn.test(alpha.stats3$Chao1, alpha.stats3$species, method="bonferroni")
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 64.3067, df = 2, p-value = 0
##
##
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |   Pocillop   seawater
## ---------+----------------------
## seawater |   7.581725
##          |     0.0000
##          |
## Stylopho |   1.146749  -7.033279
##          |     0.3772     0.0000
```

In each case, the seawater was signficiantly different to the corals, while the corals were not different to each other. This suggests the corals have a more 'selective' community of microbes compared to the surrounding seawater.

Similarity Profile Analysis (SIMPROF)

This will show how the samples cluster without any a priori assumptions regarding sample origin

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

```
# spist <- subset_samples(allPhylo, species=='Stylophora pistillata')
# spistShared = otu_table(spist)
# class(spistShared) <- "numeric"
#
# spistSIMPROF <- simprof(spistShared, num.expected=1000, num.simulated=999, method.cluster='average', 
#
# simprof.plot(spistSIMPROF, leafcolors=NA, plot=TRUE, fill=TRUE, leaflab="perpendicular", siglinetype=
#
# pVerr <- subset_samples(allPhylo, species=='Pocillopora verrucosa')
# pVerrShared = otu_table(pVerr)
```

```
# class(pVerrShared) <- "numeric"
#
# pVerrSIMPROF <- simprof(pVerrShared, num.expected=1000, num.simulated=999, method.cluster='average', 
#
# simprof.plot(pVerrSIMPROF, leafcolors=NA, plot=TRUE, fill=TRUE, leaflab="perpendicular", siglinetype=
```

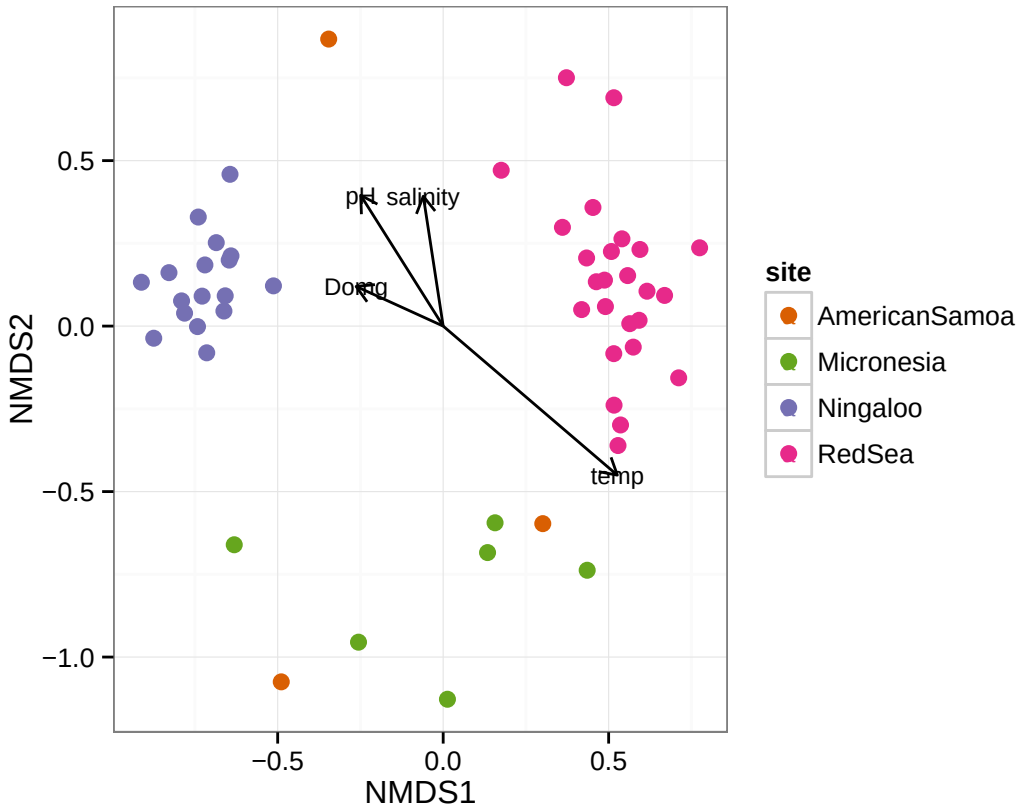
Chemical and biological correlations

Use the envfit function from the Vegan package to test if any environmental variables are significantly correlated with microbiome differences in the corals

```
draw_envfit_ord <- function(coral_chem, env_data){
  chemNoNA <- na.omit(metaFileChem[sample_names(coral_chem),env_data])
  coralNoNA <- prune_samples(rownames(chemNoNA), coral_chem)

  theme_set(theme_bw())
coralNoNAOrd <- ordinate(coralNoNA, "NMDS", "bray")
coralNoNAOrdPlot <- plot_ordination(coralNoNA, coralNoNAOrd, type = 'samples', color='site') +
  geom_point(size=3) +
  scale_color_manual(values=c(cols))

  # get point for ggplot
  pointsNoNA <- coralNoNAOrd$points[rownames(chemNoNA),]
  chemFit <- envfit(pointsNoNA, env = chemNoNA, na.rm=TRUE)
  print(chemFit)
  chemFit.scores <- as.data.frame(scores(chemFit, display= "vectors"))
  chemFit.scores <- cbind(chemFit.scores, Species = rownames(chemFit.scores))

  # create arrow info
chemNames <- rownames(chemFit.scores)
arrowmap <- aes(xend = MDS1, yend = MDS2, x = 0, y = 0, shape = NULL, color = NULL)
  labelmap <- aes(x = MDS1, y = MDS2 + 0.04, shape = NULL, color = NULL, size=1.5, label = chemNames)
  arrowhead = arrow(length = unit(0.25, "cm"))

  # note: had to use aes_string to get ggplot to recognize variables
coralNoNAOrdPlot +
  coord_fixed() +
  geom_segment(arrowmap, size = 0.5, data = chemFit.scores, color = "black",  arrow = arrowhead, show_g
  geom_text(aes_string(x = "MDS1", y = "MDS2", shape = NULL, color = NULL, size=1.5, label = "Species")
}

waterQual <- c("temp", "salinity", "Domg", "pH")
nutrients <- c("PO4", "N.N", "silicate", "NO2", "NH4")
FCM <- c("prok", "syn", "peuk", "pe.peuk", "Hbact")

spistChem <- subset_samples(allPhyloChem, species=='Stylophora pistillata')
pverrChem <- subset_samples(allPhyloChem, species=='Pocillopora verrucosa')

draw_envfit_ord(spistChem, waterQual)
```

```
## Square root transformation
## Wisconsin double standardization
```

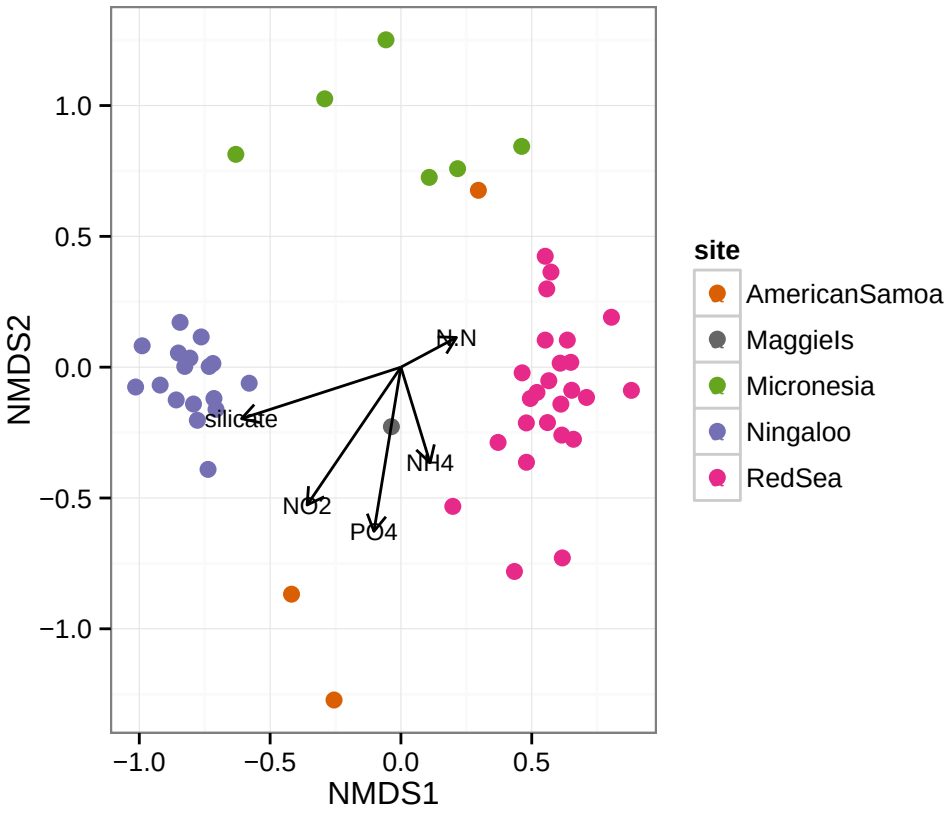
```
## Run 0 stress 0.189
## Run 1 stress 0.1729
## ... New best solution
## ... procrustes: rmse 0.04878  max resid 0.2672
## Run 2 stress 0.2191
## Run 3 stress 0.2
## Run 4 stress 0.2133
## Run 5 stress 0.2092
## Run 6 stress 0.2003
## Run 7 stress 0.1853
## Run 8 stress 0.1848
## Run 9 stress 0.2156
## Run 10 stress 0.2121
## Run 11 stress 0.1828
## Run 12 stress 0.2202
## Run 13 stress 0.2
## Run 14 stress 0.1931
## Run 15 stress 0.1848
## Run 16 stress 0.2178
## Run 17 stress 0.2018
## Run 18 stress 0.2141
## Run 19 stress 0.1885
## Run 20 stress 0.1941
##
## ***VECTORS
##
##            MDS1   MDS2   r2 Pr(>r)
## temp      0.760 -0.650 0.48  0.001 ***
## salinity -0.151  0.989 0.16  0.020 *
## Domg     -0.910  0.415 0.08  0.122
## pH       -0.533  0.846 0.22  0.004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```




```
draw_envfit_ord(spistChem, nutrients)
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1927
## Run 1 stress 0.1979
## Run 2 stress 0.1848
## ... New best solution
## ... procrustes: rmse 0.0305  max resid 0.1973
## Run 3 stress 0.1837
## ... New best solution
## ... procrustes: rmse 0.01026  max resid 0.06491
## Run 4 stress 0.2098
## Run 5 stress 0.2
## Run 6 stress 0.2092
## Run 7 stress 0.1963
## Run 8 stress 0.1874
## Run 9 stress 0.1957
## Run 10 stress 0.2218
## Run 11 stress 0.2114
## Run 12 stress 0.1888
## Run 13 stress 0.1955
## Run 14 stress 0.2078
## Run 15 stress 0.1821
## ... New best solution
## ... procrustes: rmse 0.07277  max resid 0.4077
## Run 16 stress 0.2004
```

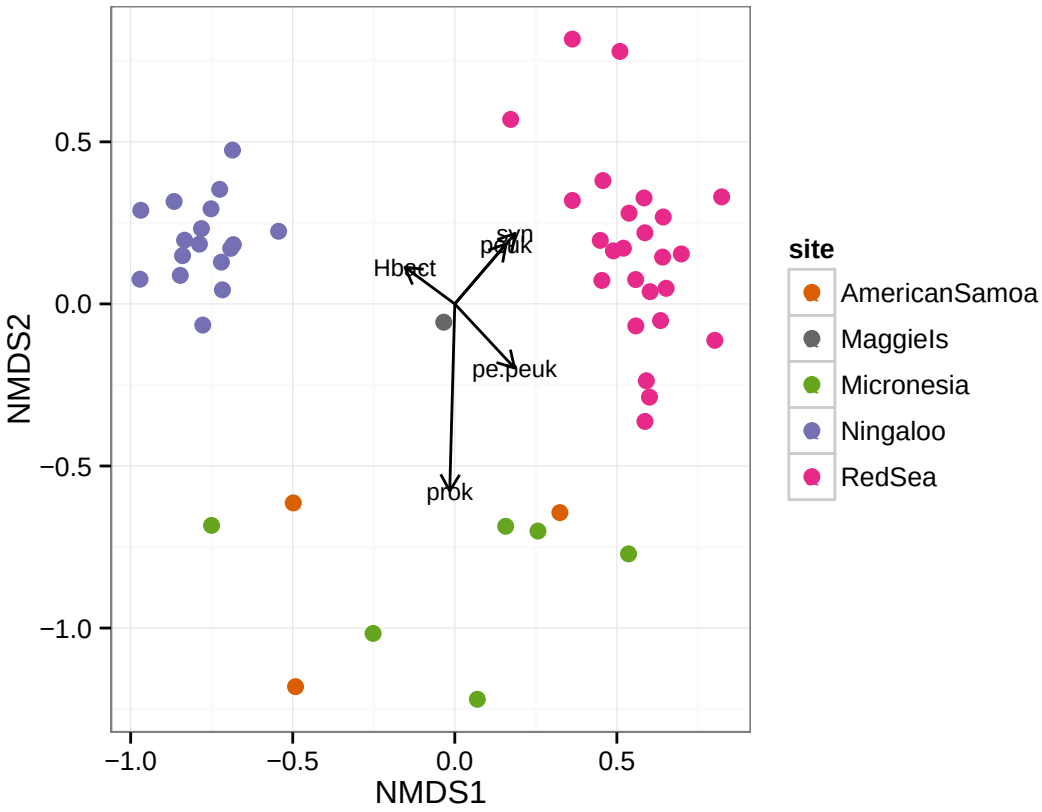
```
## Run 17 stress 0.1837
## Run 18 stress 0.2134
## Run 19 stress 0.2139
## Run 20 stress 0.1955
##
## ***VECTORS
##
##             MDS1   MDS2   r2 Pr(>r)
## PO4      -0.162 -0.987 0.40  0.001 ***
## N.N       0.882  0.472 0.06  0.197
## silicate -0.952 -0.307 0.41  0.001 ***
## NO2      -0.563 -0.827 0.41  0.001 ***
## NH4       0.292 -0.957 0.15  0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```




```
draw_envfit_ord(spistChem, FCM)
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1927
## Run 1 stress 0.2191
## Run 2 stress 0.2193
## Run 3 stress 0.1971
## Run 4 stress 0.2141
## Run 5 stress 0.1848
```

```
## ... New best solution
## ... procrustes: rmse 0.03064  max resid 0.1973
## Run 6 stress 0.2075
## Run 7 stress 0.2075
## Run 8 stress 0.1936
## Run 9 stress 0.1922
## Run 10 stress 0.2043
## Run 11 stress 0.216
## Run 12 stress 0.196
## Run 13 stress 0.206
## Run 14 stress 0.2108
## Run 15 stress 0.1943
## Run 16 stress 0.1949
## Run 17 stress 0.2147
## Run 18 stress 0.2017
## Run 19 stress 0.2089
## Run 20 stress 0.2179
##
## ***VECTORS
##
##           MDS1   MDS2   r2 Pr(>r)
## prok    -0.027 -1.000 0.33  0.002 **
## syn      0.650  0.760 0.08  0.092 .
## peuk     0.650  0.760 0.06  0.214
## pe.peuk  0.682 -0.731 0.07  0.163
## Hbact   -0.805  0.593 0.04  0.426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```



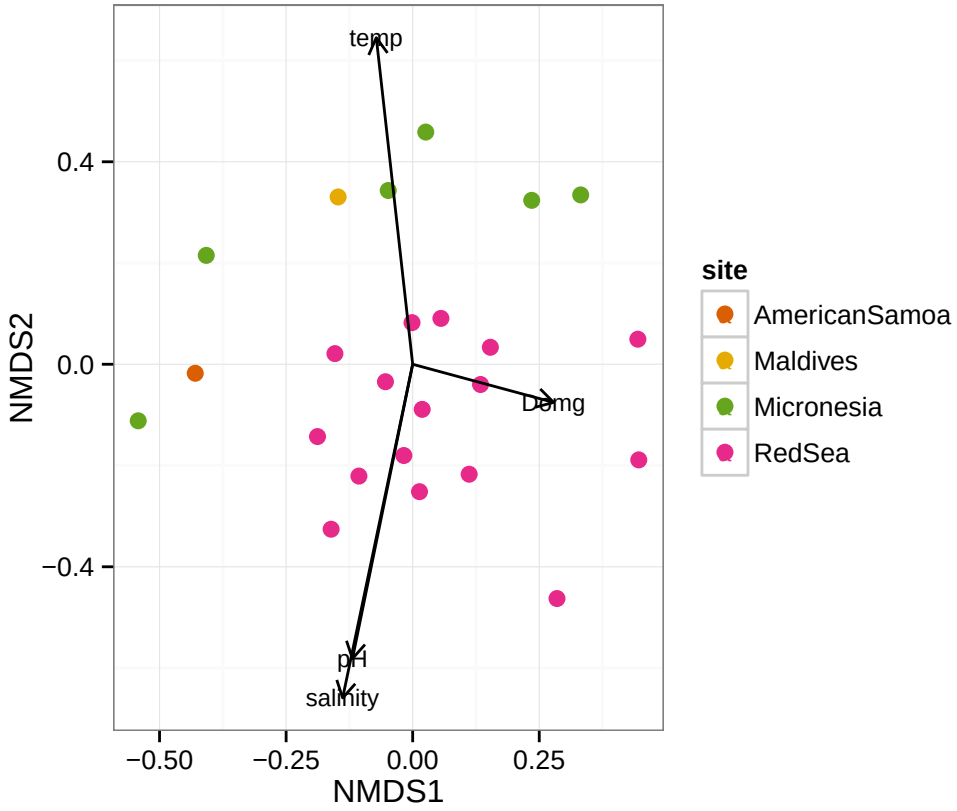

```
draw_envfit_ord(pverrChem, waterQual)
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.2599
## Run 1 stress 0.2639
## Run 2 stress 0.2739
## Run 3 stress 0.2583
## ... New best solution
## ... procrustes: rmse 0.1098  max resid 0.2726
## Run 4 stress 0.263
## Run 5 stress 0.2699
## Run 6 stress 0.2583
## ... New best solution
## ... procrustes: rmse 0.1761  max resid 0.4596
## Run 7 stress 0.259
## Run 8 stress 0.2535
## ... New best solution
## ... procrustes: rmse 0.1793  max resid 0.436
## Run 9 stress 0.2579
## Run 10 stress 0.2545
## Run 11 stress 0.247
## ... New best solution
## ... procrustes: rmse 0.1713  max resid 0.384
## Run 12 stress 0.2708
## Run 13 stress 0.2762
## Run 14 stress 0.2536
```

```
## Run 15 stress 0.2559
## Run 16 stress 0.2555
## Run 17 stress 0.263
## Run 18 stress 0.2692
## Run 19 stress 0.2582
## Run 20 stress 0.2581
```

```
## Warning: skipping half-change scaling: too few points below threshold
```

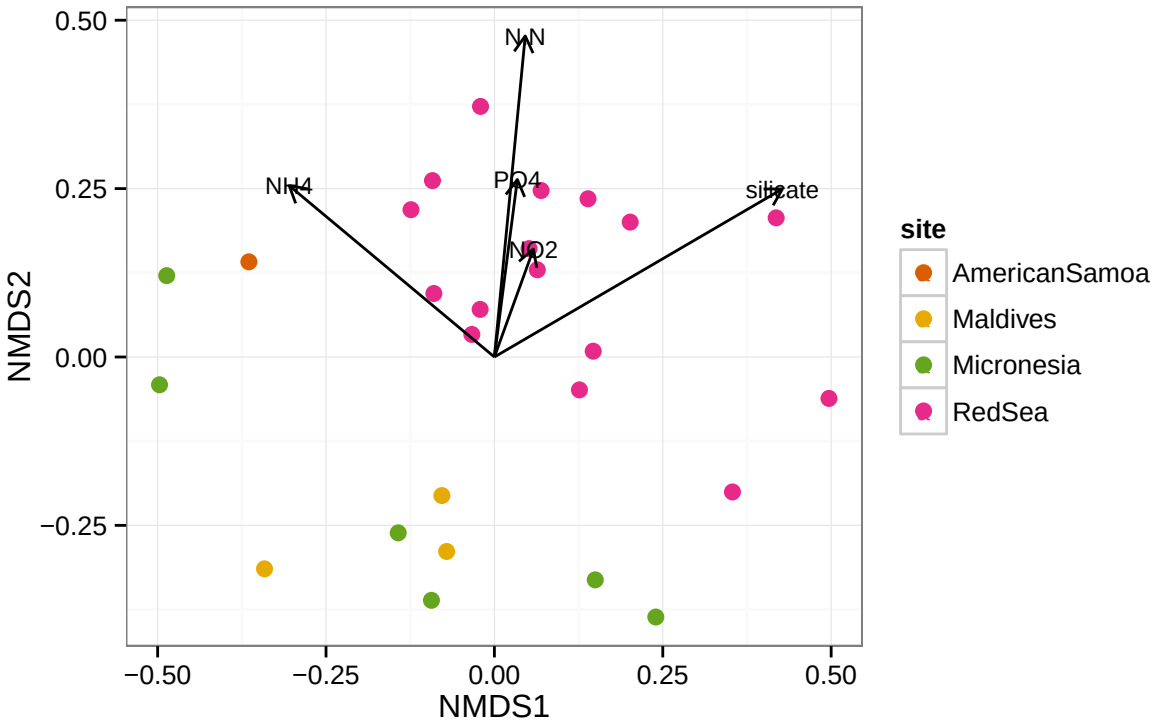
```
##
## ***VECTORS
##
##            MDS1   MDS2   r2 Pr(>r)
## temp     -0.111  0.994 0.42  0.006 **
## salinity -0.205 -0.979 0.45  0.007 **
## Domg      0.965 -0.261 0.08  0.369
## pH       -0.200 -0.980 0.35  0.010 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```




```
draw_envfit_ord(pverrChem, nutrients)
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.262
```

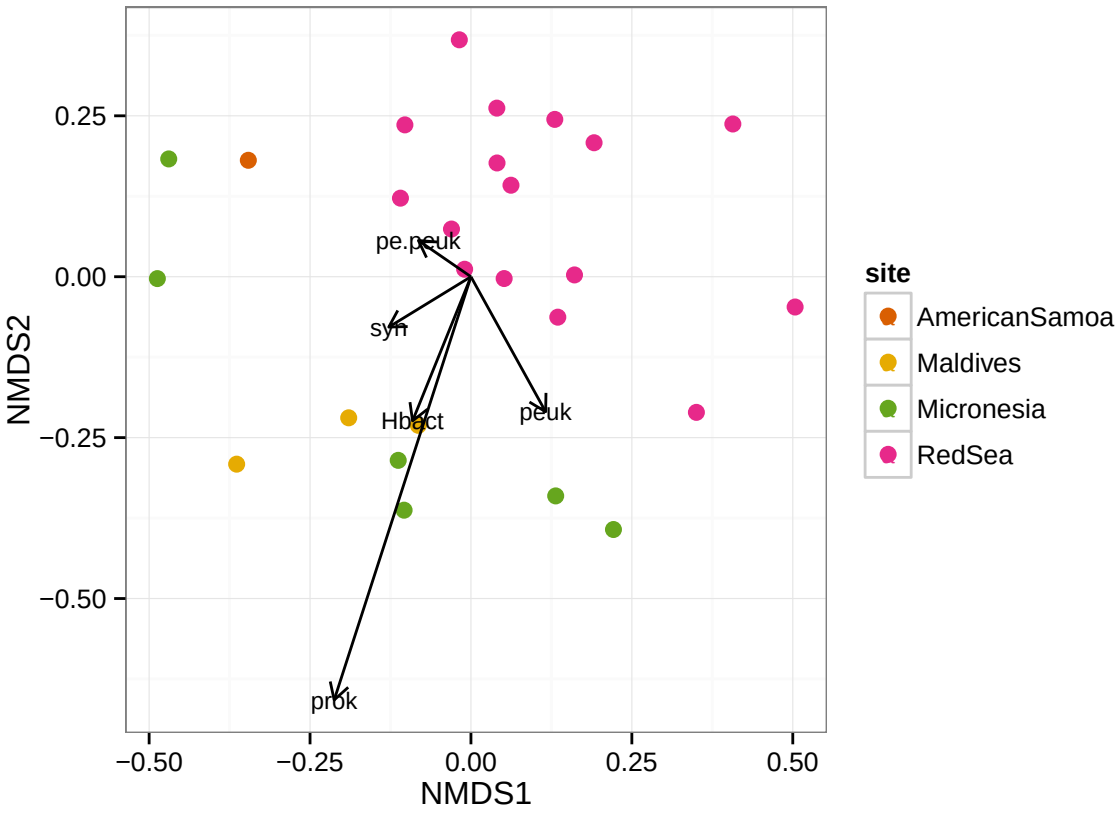
```
## Run 1 stress 0.2529
## ... New best solution
## ... procrustes: rmse 0.1795  max resid 0.3817
## Run 2 stress 0.2605
## Run 3 stress 0.276
## Run 4 stress 0.2585
## Run 5 stress 0.2703
## Run 6 stress 0.282
## Run 7 stress 0.2666
## Run 8 stress 0.2939
## Run 9 stress 0.2439
## ... New best solution
## ... procrustes: rmse 0.08096  max resid 0.2958
## Run 10 stress 0.2542
## Run 11 stress 0.2778
## Run 12 stress 0.2645
## Run 13 stress 0.2743
## Run 14 stress 0.2654
## Run 15 stress 0.2766
## Run 16 stress 0.2719
## Run 17 stress 0.2748
## Run 18 stress 0.2627
## Run 19 stress 0.2657
## Run 20 stress 0.2907

## Warning: skipping half-change scaling: too few points below threshold

##
## ***VECTORS
##
##            MDS1   MDS2   r2 Pr(>r)
## PO4       0.128  0.992 0.07  0.437
## N.N       0.096  0.995 0.23  0.038 *
## silicate  0.864  0.504 0.25  0.027 *
## NO2       0.340  0.940 0.03  0.711
## NH4      -0.768  0.641 0.16  0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```




```
draw_envfit_ord(pverrChem, FCM)
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.262
## Run 1 stress 0.2789
## Run 2 stress 0.2602
## ... New best solution
## ... procrustes: rmse 0.1728  max resid 0.4041
## Run 3 stress 0.2598
## ... New best solution
## ... procrustes: rmse 0.15  max resid 0.4096
## Run 4 stress 0.2633
## Run 5 stress 0.2603
## Run 6 stress 0.2684
## Run 7 stress 0.2588
## ... New best solution
## ... procrustes: rmse 0.1693  max resid 0.3954
## Run 8 stress 0.2437
## ... New best solution
## ... procrustes: rmse 0.1479  max resid 0.3598
## Run 9 stress 0.244
## ... procrustes: rmse 0.01929  max resid 0.08613
## Run 10 stress 0.2873
## Run 11 stress 0.2627
## Run 12 stress 0.2963
## Run 13 stress 0.255
## Run 14 stress 0.2608
## Run 15 stress 0.2957
## Run 16 stress 0.2736
```

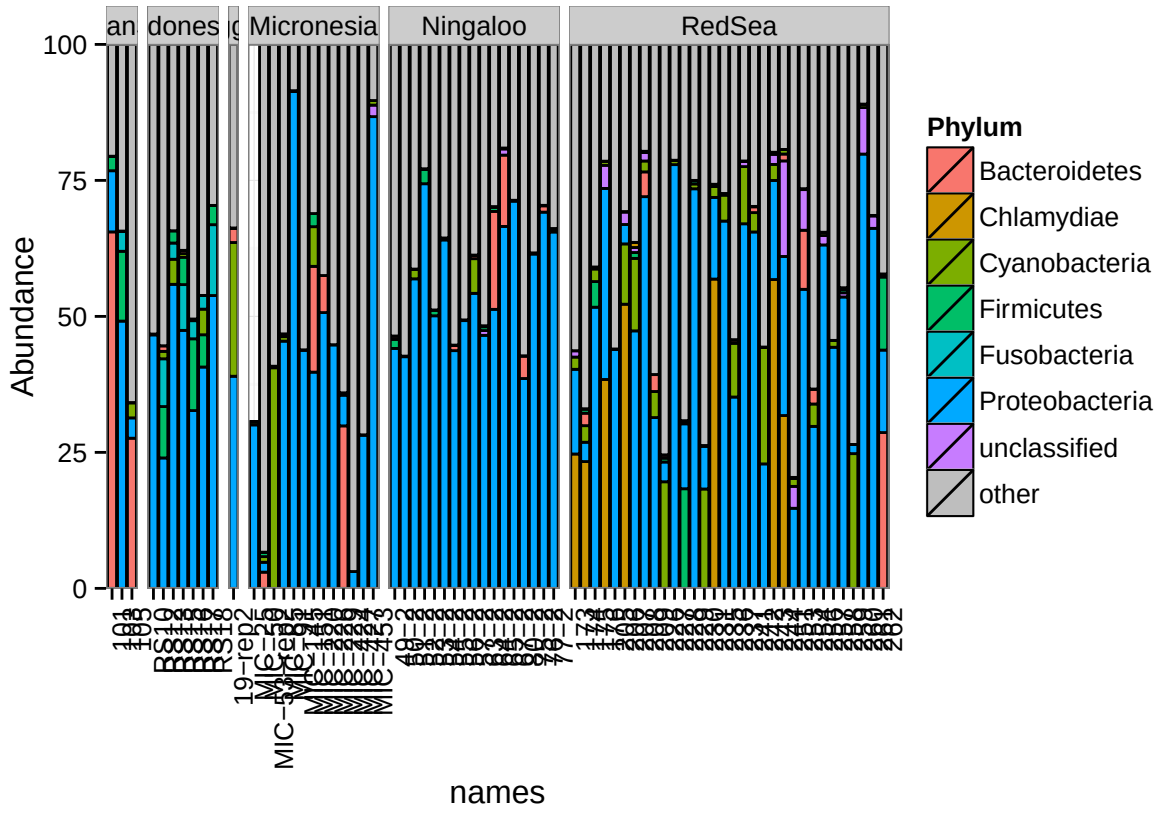
```
## Run 17 stress 0.2652
## Run 18 stress 0.2667
## Run 19 stress 0.2906
## Run 20 stress 0.2621

## Warning: skipping half-change scaling: too few points below threshold

##
## ***VECTORS
##
##           MDS1   MDS2   r2 Pr(>r)
## prok    -0.308 -0.952 0.48  0.001 ***
## syn     -0.853 -0.522 0.02  0.785
## peuk     0.484 -0.875 0.06  0.511
## pe.peuk -0.823  0.568 0.01  0.886
## Hbact   -0.375 -0.927 0.06  0.492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```



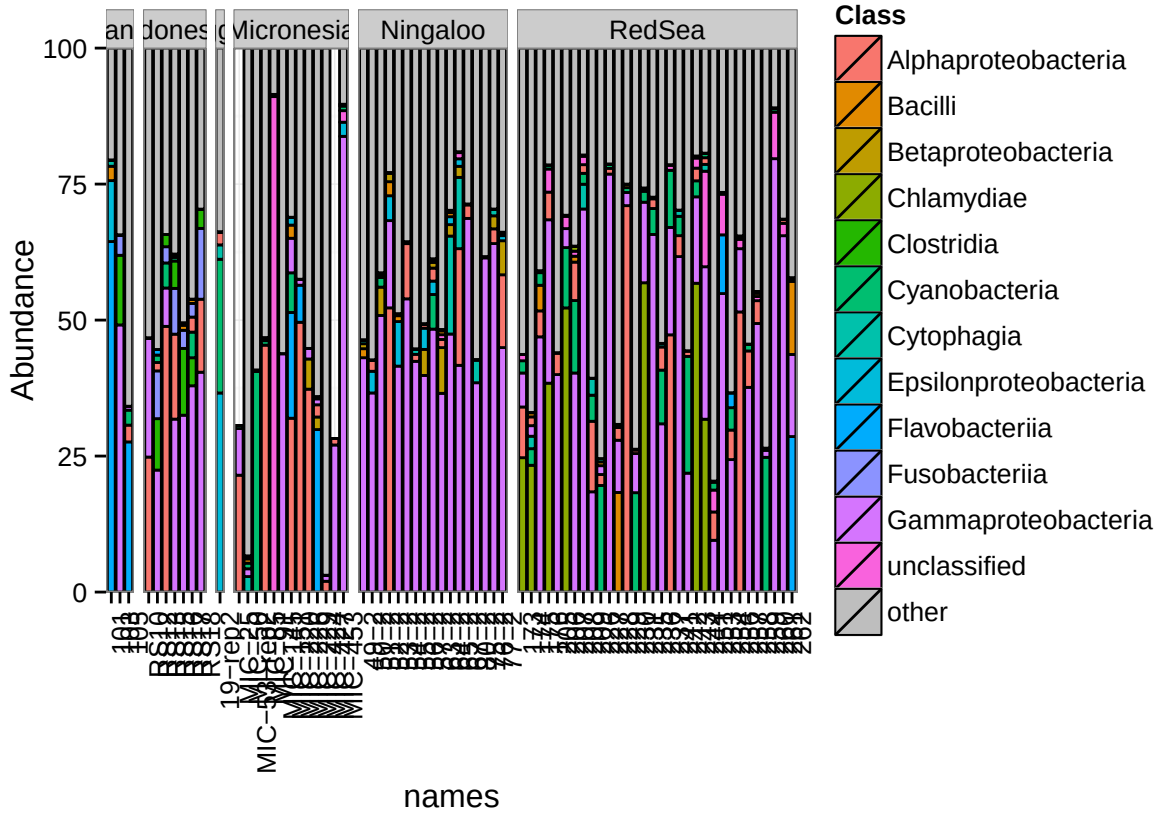
Taxonomic barcharts of bacteria in the corals and seawaters, and core microbiome members

```
# define a function to draw barcharts at a specific taxonomic level
# also need to create my own ggplot colors then replace the last one ('other' column) with gray
```

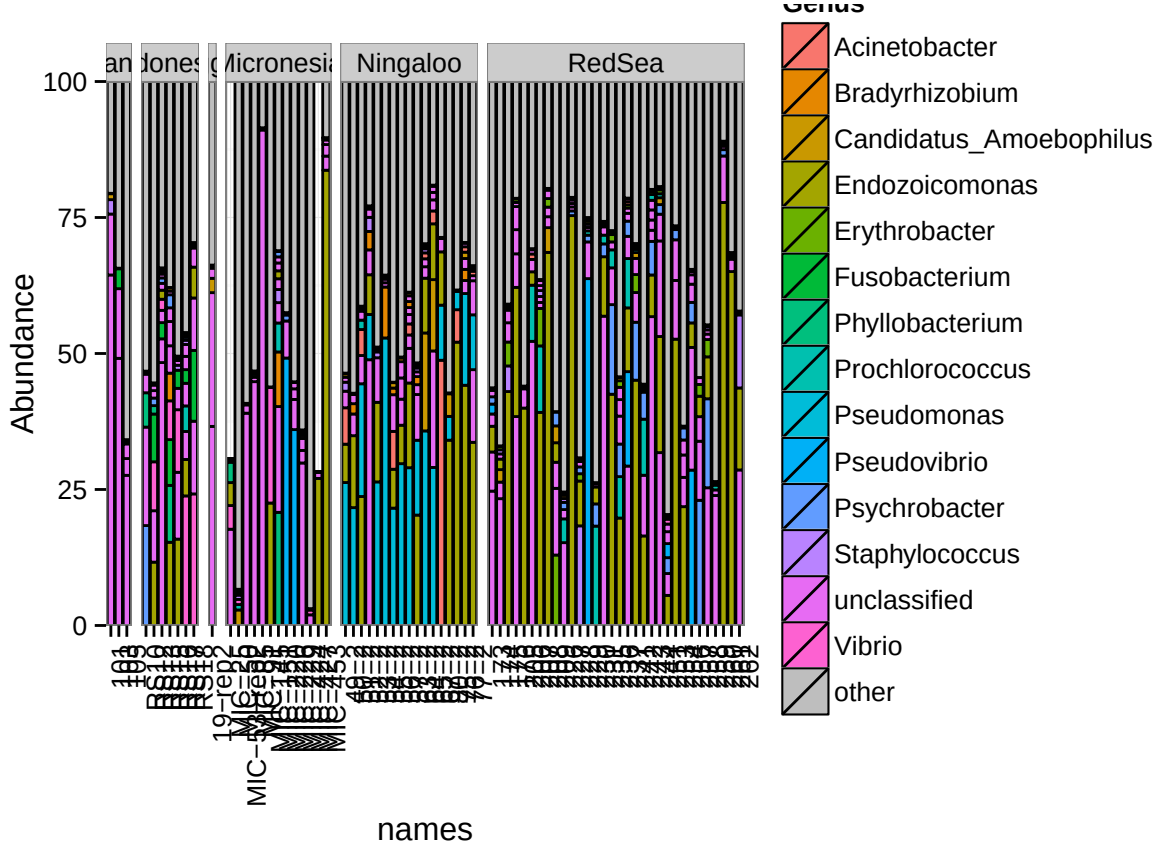
```
gg_color_hue <- function(n) {
  hues = seq(15, 375, length=n+1)
  hcl(h=hues, l=65, c=100)[1:n]
}

draw_barcharts <- function(coral_species, tax_level) {

coralFiltGlom <- tax_glom(coral_species, taxrank=tax_level)
physeqdf <- psmelt(coralFiltGlom)

# get total abundance so can make an 'other' column
# had to add ^ and $ characters to make sure grep matches whole word

physeqdfOther <- physeqdf

for (j in unique(physeqdf$Sample)) {
  jFirst = paste('^', j, sep='')
  jBoth = paste(jFirst, '$', sep='')
  rowNumbers = grep(jBoth, physeqdf$Sample)
  otherValue = 100 - sum(physeqdf[rowNumbers,"Abundance"])
  newRow = (physeqdf[rowNumbers,])[1,]
  newRow[,tax_level] = "other"
  newRow[,"Abundance"] = otherValue
  physeqdfOther <- rbind(physeqdfOther, newRow)
}

ggCols <- gg_color_hue(length(unique(physeqdfOther[,tax_level])))
ggCols <- head(ggCols, n=-1)

physeqdfOther$names <- factor(physeqdfOther$Sample, levels=rownames(metaFile), ordered = TRUE)

theme_set(theme_bw())
ggplot(physeqdfOther, aes_string(x="names", y="Abundance", fill=tax_level, order = as.factor(tax_level))
  geom_bar(stat="identity", colour="black") +
  scale_fill_manual(values=c(ggCols, "gray")) +
  scale_y_continuous(expand = c(0,0), limits = c(0,100)) +
  facet_grid(~site, scales='free', space='free_x') +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
}

# subset coral samples, create names factor for label ordering and filter so the graph isn't too full
spist <- subset_samples(allPhylo, species=='Stylophora pistillata')
sample_data(spist)$names <- factor(sample_names(spist), levels=rownames(metaFile), ordered = TRUE)
spistFilt = filter_taxa(spist, function(x) mean(x) > 0.5, TRUE)
draw_barcharts(spistFilt, "Phylum")
```



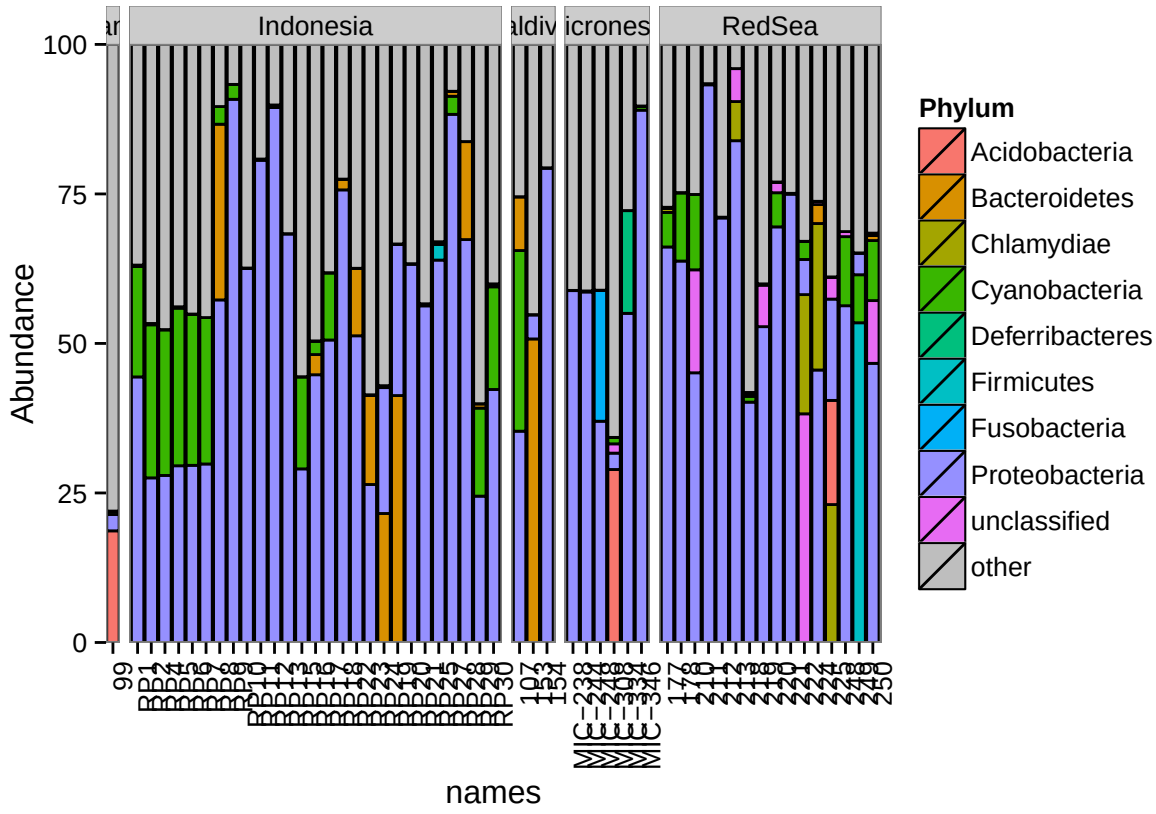

```
draw_barcharts(spistFilt, "Class")
```



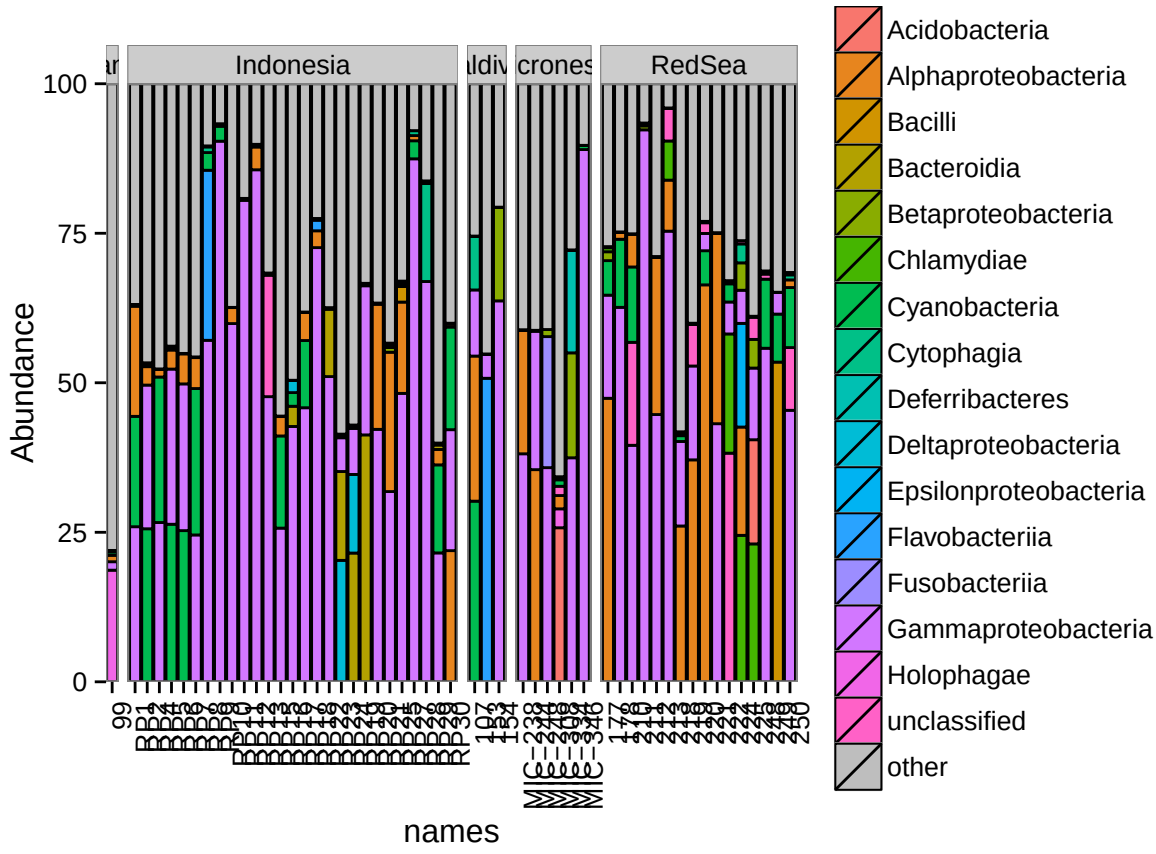
```
draw_barcharts(spistFilt, "Genus")
```



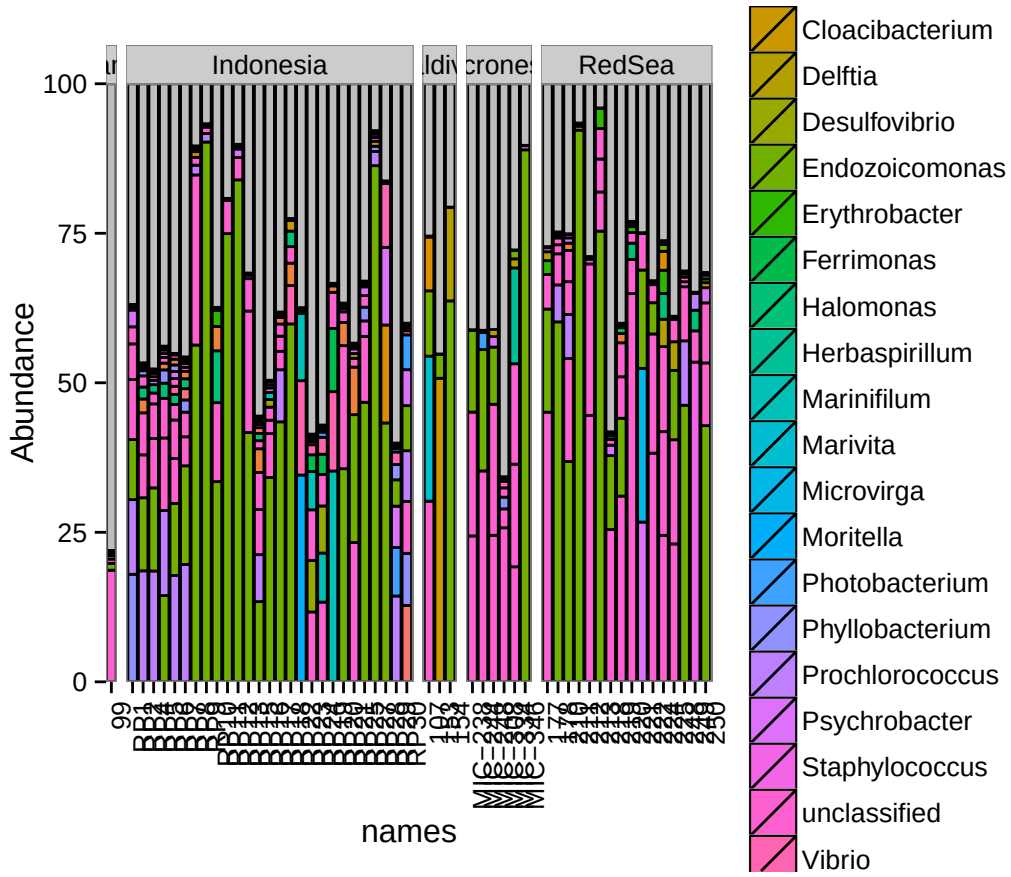

```
pVerr <- subset_samples(allPhylo, species=='Pocillopora verrucosa')
sample_data(pVerr)$names <- factor(sample_names(pVerr), levels=unique(sample_names(pVerr)))
pVerrFilt = filter_taxa(pVerr, function(x) mean(x) > 0.3, TRUE)
draw_barcharts(pVerrFilt, "Phylum")
```



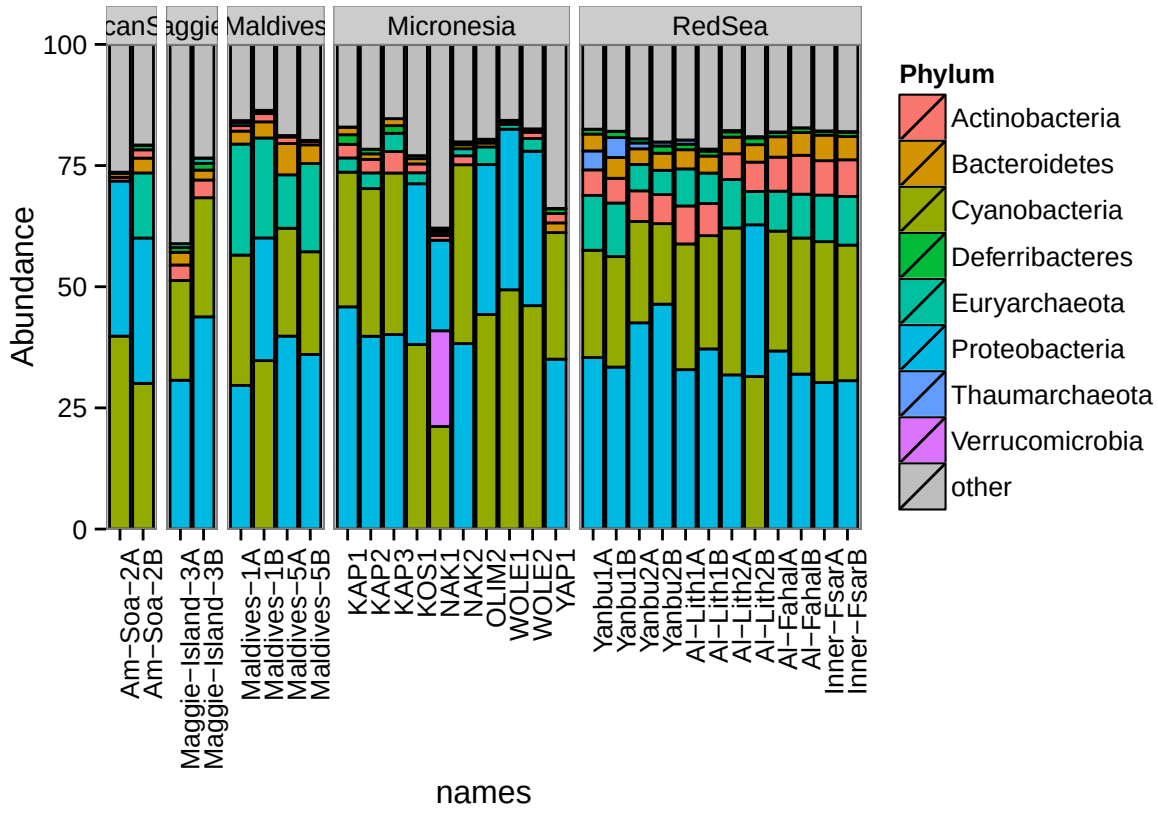

```
draw_barcharts(pVerrFilt, "Class")
```



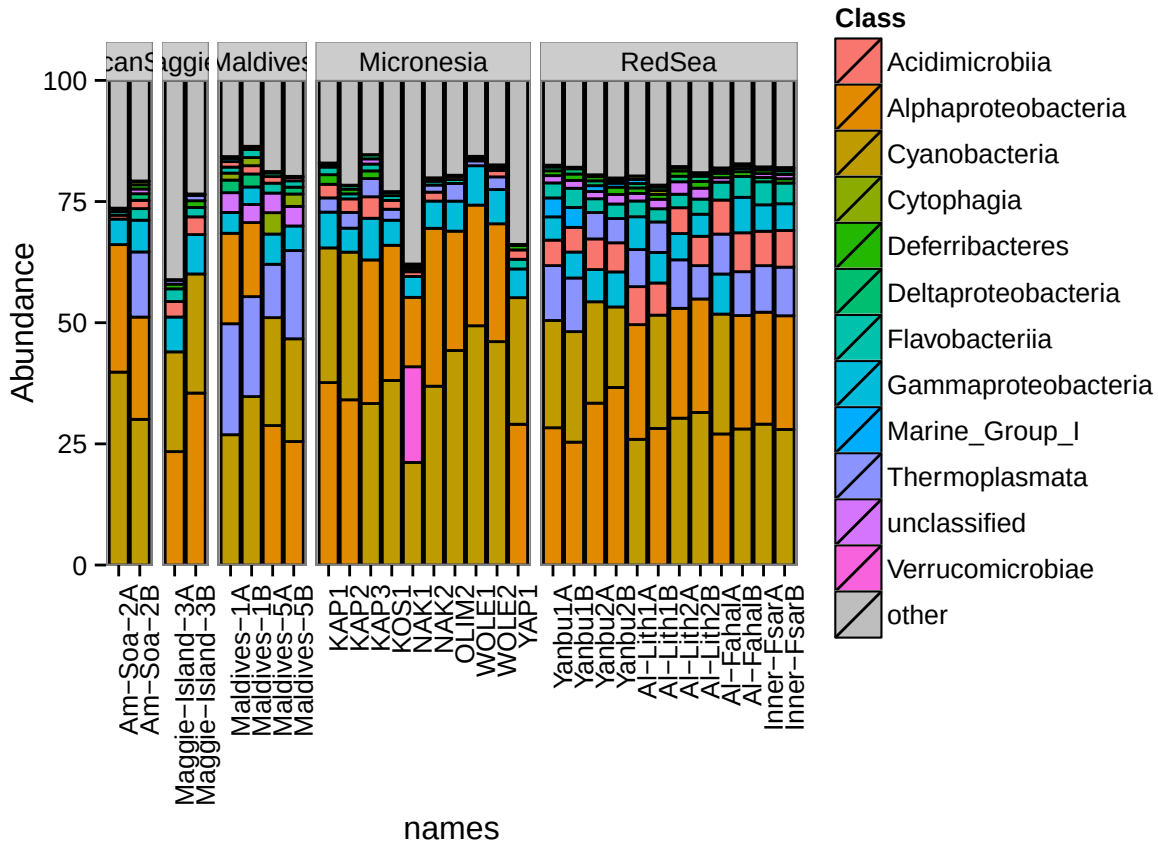

```
draw_barcharts(pVerrFilt, "Genus")
```



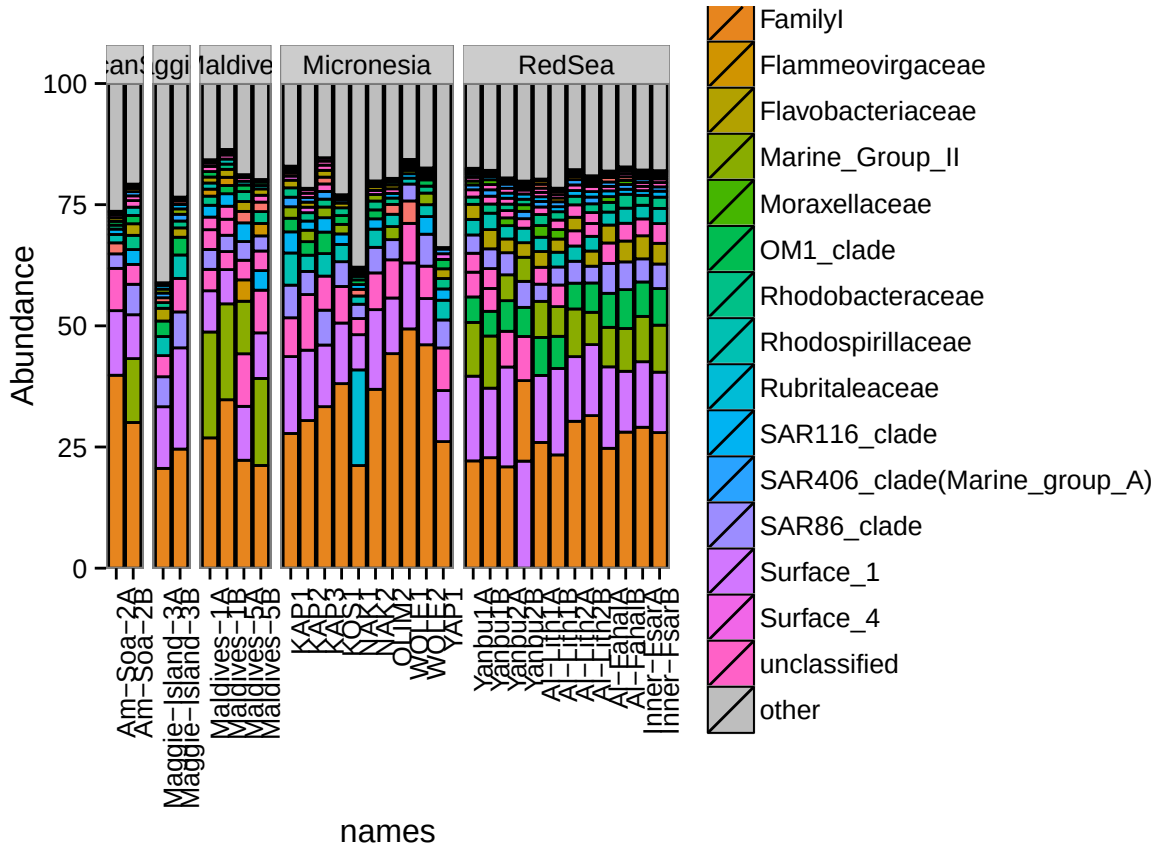

```
sea <- subset_samples(allPhylo, species=='seawater')
sample_data(sea)$names <- factor(sample_names(sea), levels=rownames(metaFile), ordered = TRUE)
seaFilt = filter_taxa(sea, function(x) mean(x) > 0.3, TRUE)
draw_barcharts(seaFilt, "Phylum")
```




```
draw_barcharts(seaFilt, "Class")
```



```
draw_barcharts(seaFilt, "Family")
```



The two corals are both dominated by Gammaproteobacteria at the higher taxonomic levels. At the genus level, there is more variability but Endozoicomonas seem to be fairly prevalent. Let's check which bacterial genera are most consistently associated with the corals and may be considered a 'core' microbiome member.

Core coral microbiome members

```
# check for 'core' microbiome members at the genus level
# which taxa are present at 1% overall abundance and at least 50% of samples in Stylophora pistillata?
spistGenusGlom <- tax_glom(spistFilt, taxrank="Genus")
coreTaxa = filter_taxa(spistGenusGlom, function(x) sum(x > 1) > (0.5*length(x)), TRUE)
tax_table(coreTaxa)
```

```
## Taxonomy Table:     [1 taxa by 7 taxonomic ranks]:
##               Domain     Phylum           Class
## MED000008661 "Bacteria" "Proteobacteria" "Gammaproteobacteria"
##               Order               Family        Genus            Species
## MED000008661 "Oceanospirillales" "Hahellaceae" "Endozoicomonas" NA
```

```
sum(otu_table(coreTaxa) > 1) / nsamples(spist)
```

```
## [1] 0.7671
```

```
# which taxa are present at 1% overall abundance and at least 50% of samples in Pocillopora verrucosa?
pVerrGenusGlom <- tax_glom(pVerrFilt, taxrank="Genus")
coreTaxa = filter_taxa(pVerrGenusGlom, function(x) sum(x > 1) > (0.5*length(x)), TRUE)
tax_table(coreTaxa)
```

```
## Taxonomy Table:     [1 taxa by 7 taxonomic ranks]:
##              Domain     Phylum           Class
## MED000008683 "Bacteria" "Proteobacteria" "Gammaproteobacteria"
##              Order               Family        Genus            Species
## MED000008683 "Oceanospirillales" "Hahellaceae" "Endozoicomonas" NA
```

```
sum(otu_table(coreTaxa) > 1) / nsamples(pVerr)
```

```
## [1] 0.8491
```

Indeed Endozoicomonas were the most prevalent bacteria in the corals and were they only bacterial genera to occur in more than 50% of the colonies sampled. In fact, for both coral species Endozoicomonas occurred in more than 75% of colonies.