NAAMES 1 and 2 16S rRNA Amplicon Analysis

Luis M. Bolaños

July 2019

Contents

1 Introduction

This pipeline was created and used for the analysis of NAAMES 1 and NAAMES 2 datasets, which results are presented in the manuscript: "Small Phytoplankton Dominate Western North Atlantic Biomass"

2 Samples, data generation and raw data availability

During NAAMES1 and NAAMES2, 56 and 64 microbial biomass samples were collected, respectively (see SOD for more information). 16S rRNA amplicon sequencing was performed on libraries made using 27F (5'-AGAGTTTGATCNTGGCTCAG-3) and 338 RPL (5'-GCWGCCWCCCGTAGGWGT-3') primer set.

Raw 16S rRNA datasets are publicly available at:

– https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/naames_1/archive/NAAMES1_RawFwd.tgz

– https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/naames_1/archive/NAAMES1_RawRev.tgz

– https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/naames_2/archive/NAAMES2_RawFwd.tgz

– https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/naames_2/archive/NAAMES2_RawRev.tgz

If using these datasets for publication please contact

- Stephen Giovannoni (Steve.giovannoni@oregonstate.edu) or
- Luis M Bolanos (bolanosl@oregonstate.edu, lbolanos@lcg.unam.mx)
- and cite: Small Phytoplankton Dominate Western North Atlantic Biomass

3 Pipeline to generate working tables

The following scripts were used sequentially to achieve the results showed in the article. Output processed tables used for R analysis can be found attached to this file.

Task 3.1 Pre-processing sequence files: CUTADAPT

Cutadapt is used to chop the primers from the raw sequences. In this SOP we are using a fixed number of bp to trim each paired-end. The fixed number match the primer length: 27F (20bp) and 338RPL (18bp)

We created two bash scripts (trimf.sh and trimrev.sh) to chop the fixed number of bp from the raw sequences

Listing 1: trimf.sh

```
#!/bin/bash

for i in *_R1.fastq.gz;
do
SAMPLE=$(echo ${i} | sed "s/_R1\.fastq\.gz//")
echo ${SAMPLE}_R1.fastq.gz
cutadapt -u 20 ${SAMPLE}_R1.fastq.gz -o ../${SAMPLE}_R1.fastq #
    output is redirected to the higher class directory. Scripts
    can be modified to direct the output to a specific directory
    created by the user.
done
```

Listing 2: trimrev.sh

```
#!/bin/bash

for i in *_R2.fastq.gz;
do
SAMPLE=$(echo ${i} | sed "s/_R2\.fastq\.gz//")
echo ${SAMPLE}_R2.fastq.gz
```

```
cutadapt -u 18 ${SAMPLE}_R2.fastq.gz -o ../${SAMPLE}_R2.fastq #
    output is redirected to the higher class directory. Scripts
    can be modified to direct the output to a specific directory
    created by the user.
done
```

Task 3.2 Run DADA2 version 1.2

We used dada2 to generate an amplified nucleotide variant table coupled with taxonomic assignation using SILVA database train version 123 (silva_nr_v123_train_set.fa) and the R version used was R-3.4.1

Listing 3: DadaR.R

```
library(dada2)

#filter
path <- "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/DADA2/
    N1N2"

fnFs <- sort(list.files(path, pattern="_R1_001.fastq", full.names
     = TRUE))
fnRs <- sort(list.files(path, pattern="_R2_001.fastq", full.names
     = TRUE))
sample.names <- sapply(strsplit(basename(fnFs), "_L"), `[`, 1)
filt_path <- file.path(path, "filtered")
filtFs <- file.path(filt_path, paste0(sample.names, "_F_filt.
    fastq"))
filtRs <- file.path(filt_path, paste0(sample.names, "_R_filt.
    fastq"))
for(i in seq_along(fnFs)) {
  fastqPairedFilter(c(fnFs[i], fnRs[i]), c(filtFs[i], filtRs[i]),
                    truncLen=c(220,190),
                    maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                    compress=TRUE, verbose=TRUE)
}

derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)
# Name the derep-class objects by the sample names
names(derepFs) <- sample.names
names(derepRs) <- sample.names

dadaFs.lrn <- dada(derepFs, err=NULL, selfConsist = TRUE,
    multithread=TRUE)
errF <- dadaFs.lrn[[1]]$err_out
```

```
dadaRs.lrn <- dada(derepRs, err=NULL, selfConsist = TRUE,
    multithread=TRUE)
errR <- dadaRs.lrn[[1]]$err_out

saveRDS(errF, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/
    DADA2/N1N2/errF.rds")
saveRDS(errR, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/
    DADA2/N1N2/errR.rds")

dadaFs <- dada(derepFs, err=errF, multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, multithread=TRUE)

saveRDS(dadaFs, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/
    DADA2/N1N2/dadaFs_N.rds")
saveRDS(dadaRs, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/
    DADA2/N1N2/dadaRs_N.rds")

mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=
    TRUE)

saveRDS(mergers, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES
    /DADA2/N1N2/mergers.rds")
seqtab <- makeSequenceTable(mergers[names(mergers) != "Mock"])
dim(seqtab)

# Inspect distribution of sequence lengths
table(nchar(getSequences(seqtab)))

seqtab.nochim <- removeBimeraDenovo(seqtab, verbose=TRUE)

sum(seqtab.nochim)/sum(seqtab)

dim(seqtab.nochim)

saveRDS(seqtab.nochim, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/
    NAAMES/DADA2/N1N2/seqtab.nochim.rds")

taxa <- assignTaxonomy(seqtab.nochim, "/nfs0/Giovannoni_Lab/
    workspaces/bolanosl/BIOS/ProcessSeqs/SEQ1pr/silva_nr_v123_
    train_set.fa")
unname(head(taxa))

saveRDS(taxa, "/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES/
    DADA2/N1N2/taxa.rds")
```

```
write.table(cbind(t(seqtab.nochim) , taxa), "/nfs0/Giovannoni_Lab
    /workspaces/bolanosl/NAAMES/DADA2/N1N2/seqtab-nochimtaxa.txt"
    , sep="\t", row.names=TRUE, col.names=NA, quote=FALSE)
write.table(taxa,"/nfs0/Giovannoni_Lab/workspaces/bolanosl/NAAMES
    /DADA2/N1N2/N1N2/taxa.txt", sep="\t", row.names=TRUE, col.
    names=NA, quote=FALSE )
```

Task 3.3 Parse Dada output

Dada output is an ASVtable named seqtab-nochimtaxa.txt" with the unique sequence as row name (identifier) and assigned taxa as the last columns (SILVA hierarchical format). We need to parse this file to generate an ASVtable and a fasta file for the "photosynthetic origin" fraction of the sequences and link them with different identifiers.

Listing 4: parseNames.pl

```
#This script (parseNames.pl) change the headers name
#!/usr/bin/perl

use warnings;
use strict;

my $seqtab = $ARGV[0];
my $filename = 'seqtab-par.txt';
my @printablehds;
my $headers;

open my $file, '<', $seqtab or die "cant open the file: $! \n"; #
    Open Fileders
$headers = <$file>; # Get the first line, in this case, the
    headers

open my $out_fh, '>', "$filename.tmp" #Open output file
  or die "Cannot open $filename.tmp for writing: $!";

my @newheaders= split(/\t/,$headers); #split by tab the headers
    saved in the new line
foreach my $loop_variable (@newheaders) {
       if ($loop_variable =~ /(.+)-(.+-.+_.+)/){ #if the header
           is of the form X-X-X-X get just the last significant
           part
       push @printablehds, $2;
       }
       else{
       push @printablehds, $loop_variable; # New first line with
           shortened headers
```

```
        }
}

print {$out_fh} join "\t", @printablehds; # Print the first new
    line
print {$out_fh} $_ while <$file>; # printe everything of the
    original file except the 1st line
close $out_fh;
close $file;
```

The generated file seqtab-par.txt.tmp will be use as input of the following perl script addcolnm.pl to add a column with a fix number which will help us to identify the ASVs and link them to the fasta file.

Listing 5: addcolnm.pl

```
#!/usr/bin/perl

use warnings;
use strict;

my $input = $ARGV[0];
my $filename = 'seqtab-par.txt.tmp.co';
my $line=1;

open my $fh, '<', $input or die $!;

open my $out_fh, '>', $filename or die "Cannot open $filename for
     writing: $!";

my $firstLine = 1;

while (<$fh>){
    if($firstLine){
        $firstLine = 0;
        s/^/\t/;
        print $out_fh $_;
}
    else{
        s/^/N1N2_SNV$line\t/;
        print $out_fh $_;
        $line++;
}
}

close $out_fh;
close $fh;
```

Listing 6: Extracting only Photosynthetic sequences using shell commands

```
###Remove non 16S rRNA eukaryotic sequences and create two files
    for photosynthetic and heterotrophic 16S sequences


grep -vw Eukaryota" seqtab-par.txt.tmp.co > seqtab-par_on16.txt

grep "Cyanobacteria" seqtab-par_on16.txt | cut -f 1 >
    Assignationparse.input.phot.list



ON16 split into Phytoplankton and het bacteria

grep -vwf Assignationparse.input.phot.list seqtab-par_on16.txt >
    seqtab-par_on16.photo.txt #From the total this is the
    photosynthetic fraction
grep -vwf Assignationparse.input.hete.list seqtab-par_on16.txt >
    seqtab-par_on16.hete.txt #From the total this is the
    heterotrophic bacteria fraction


###Create fasta file from seqtab-par_on16.photo.txt and annotate
    using phyloassigner along the curated datasets found on https
    ://www.mbari.org/resources-worden-lab/

cut -f 1,2 seqtab-par_on16.photo.txt | sed "s/N1N2/>N1N2/" | sed
    "s/\t/\n/" > seqtab-par_on16.photo.fa

perl /raid1/home/micro/bolanosl/local/source/phyloassigner-6.166/
    phyloassigner.pl --hmmerdir /raid1/home/micro/bolanosl/bin/
    --pplacerdir /raid1/home/micro/bolanosl/local/source/
    phyloassigner-6.166/binaries/ -o /nfs0/Giovannoni_Lab/
    workspaces/bolanosl/NAAMES/DADA2/N1N2/phyto_N1N2
    plastid_arb_691_30apr2015.phyloassignerdb /nfs0/
    Giovannoni_Lab/workspaces/bolanosl/NAAMES/DADA2/N1N2/seqtab-
    par_on16.photo.fa

cat phyto_N1N2_str/phyto_N1N2_str.fas.aln.jplace.tab
    phyto_N1N2_cya/phyto_N1N2_cya.fas.aln.jplace.tab
    phyto_N1N2_vir/phyto_N1N2_vir.fas.aln.jplace.tab >
    N1N_strvircya.tab

cut -f 1,2,3 N1N_strvircya.tab | sort | sed '/^#/ d'>
    N1N_strvircya1.tab
```

```
cut -f 1 N1N_strvircya1.tab > lsttogrep.lst
grep -vwf lsttogrep.lst phyto_N1N2_plastid/phyto_N1N2.fa.aln.
    jplace.tab | cut -f 1,2,3 > complofstrvircya.tab
cat N1N_strvircya1.tab complofstrvircya.tab | sed '/^#/ d' | sed
    "s/SNV/SNV\t/" | sort -k 2,2 -n | sed "s/SNV\t/SNV/" | sed 's
    /;$//'> tax_node_ass.txt

From the Phyloassigner output, we changed to the following
    consensus taxonomy instead of abbreviations. SILVA
    hierarchical taxonomy (Domain Phylum Class Order Family Genus
     Species)

bac --> Eukaryota Stramenopiles Bacillariophyceae
Basal --> Bacteria Cyanobacteria
Basal;MarPico;MarSynA;SynVI --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynVI
Basal;MarPico;MarSynB;SynI --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynI
Basal;MarPico;MarSynC;SynEPC2 --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynEPC2
Basal;MarPico;MarSynC;SynII --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynII
Basal;MarPico;MarSynC;SynIII --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynIII
Basal;MarPico;MarSynC;SynIV --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynIV
Basal;MarPico;MarSynC;SynWPC1 --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynWPC1
Basal;MarPico;MarSynD;SynXVI --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus SynXVI
Basal;MarPico;Pro;ProHL;ProHLI --> Bacteria Cyanobacteria MarSynB
     Synechococcales Prochloraceae Prochlorococcus ProHLI
Basal;MarPico;Pro;ProHL;ProHLII --> Bacteria Cyanobacteria
    MarSynB Synechococcales Prochloraceae Prochlorococcus ProHLII
Basal;MarPico;Pro;ProHL;ProHNLCs;ProHNLC1 --> Bacteria
    Cyanobacteria MarSynB Synechococcales Prochloraceae
    Prochlorococcus ProHNLCs
Basal;MarPico;Pro;ProLLI --> Bacteria Cyanobacteria MarSynB
    Synechococcales Prochloraceae Prochlorococcus ProLLI
Basal;MarPico;Pro;ProLLIV --> Bacteria Cyanobacteria MarSynB
    Synechococcales Prochloraceae Prochlorococcus ProLLIV
Basal;MarPico;Syn5_3 --> Bacteria Cyanobacteria MarSynB
    Synechococcales Synechococcaceae Synechococcus Syn5_3
Basal;NonMarPico;Cmb2 --> Cmb2
Basal;NonMarPico;GroupB --> GroupB
```

```
bol --> Eukaryota Stramenopiles Bolidophyceae
Chlorophyta --> Eukaryota Chlorophyta
Chlorophyta;2ndaryPlastid --> Eukaryota Chlorophyta 2ndaryPlastid
Chlorophyta;2ndaryPlastid;Alveo_Green --> Eukaryota Chlorophyta 2
    ndaryPlastid Alveo_Green Chlorophyta;Chlo_Chlo --> Eukaryota
    Chlorophyta Chlorara
Chlorophyta;Chlo_mix --> Eukaryota Chlorophyta Chlorara
Chlorophyta;Chlorara --> Eukaryota Chlorophyta Chlorara
Chlorophyta;Chlo_Trebou --> Eukaryota Chlorophyta
    Trebouxiophyceae
Chlorophyta;PrasI --> Eukaryota Chlorophyta PrasinophyceaeI
Chlorophyta;PrasIII --> Eukaryota Chlorophyta PrasinophyceaeIII
Chlorophyta;PrasII;PrasII_Crusto --> Eukaryota Chlorophyta
    PrasinophyceaeII
Chlorophyta;PrasII;PrasII_Mamiello --> Eukaryota Chlorophyta
    PrasinophyceaeII Mamiellales
Chlorophyta;PrasII;PrasII_Mamiello;Bathy --> Eukaryota
    Chlorophyta PrasinophyceaeII Mamiellales Bathycoccaceae
    Bathycoccus
Chlorophyta;PrasII;PrasII_Mamiello;OstreoI --> Eukaryota
    Chlorophyta PrasinophyceaeII Mamiellales Bathycoccaceae
    Ostreococcus OstreococcusI
Chlorophyta;PrasII;PrasII_Mamiello;OstreoII --> Eukaryota
    Chlorophyta PrasinophyceaeII Mamiellales Bathycoccaceae
    Ostreococcus OstreococcusII
Chlorophyta;PrasII;PrasII_Mamiello;PrasII_Mam_Mant_Mic;
    PrasII_Mami --> Eukaryota Chlorophyta PrasinophyceaeII
Chlorophyta;PrasII;PrasII_Mamiello;PrasII_Mam_Mant_Mic;
    PrasII_Mant_Mic --> Eukaryota Chlorophyta PrasinophyceaeII
    Mamiellales Mamiellaceae Micromonas
Chlorophyta;PrasII;PrasII_Mamiello;PrasII_Mam_Mant_Mic;
    PrasII_Mant_Mic;PrasII_MicroABC --> Eukaryota Chlorophyta
    PrasinophyceaeII Mamiellales Mamiellaceae Micromonas
    MicromonasABC
Chlorophyta;PrasII;PrasII_Mamiello;PrasII_Mam_Mant_Mic;
    PrasII_Mant_Mic;PrasII_MicroABC;PrasII_MicroABC_C -->
    Eukaryota Chlorophyta PrasinophyceaeII Mamiellales
    Mamiellaceae Micromonas MicromonasABC_C
Chlorophyta;PrasII;PrasII_Mamiello;PrasII_Mam_Mant_Mic;
    PrasII_Mant_Mic;PrasII_MicroE2 --> Eukaryota Chlorophyta
    PrasinophyceaeII Mamiellales Mamiellaceae Micromonas
    MicromonasE2
Chlorophyta;PrasII;PrasII_Mono --> Eukaryota Chlorophyta
    PrasinophyceaeII
Chlorophyta;PrasIV --> Eukaryota Chlorophyta PrasinophyceaeIV
Chlorophyta;PrasV --> Eukaryota Chlorophyta PrasinophyceaeV
```

```
Chlorophyta;PrasVII_CCMP1205 --> Eukaryota Chlorophyta
    PrasinophyceaeVII_CCMP1205
Chlorophyta;PrasVIII --> Eukaryota Chlorophyta PrasinophyceaeVIII
Chlorophyta;PrasVI_Pcoccus --> Eukaryota Chlorophyta
    PrasinophyceaeVI_Pcoccus
Chlorophyta;PrasVI_Pderma --> Eukaryota Chlorophyta
    PrasinophyceaeVI_Pderma
chr --> Eukaryota Stramenopiles Chrysophyceae
cry --> Eukaryota Cryptophyta Cryptophyceae
dic --> Eukaryota Stramenopiles Dictyochophyceae
din --> Alveolata Dinophyceae
eus;eus_clade_B --> Eukaryota Eusiphoniidae
mel --> mel
NA --> NA
new_euk_A --> Eukaryota new_euk_A
new_euk_B --> Eukaryota new_euk_B
new_euk_C --> Eukaryota new_euk_C
pel --> Eukaryota Stramenopiles Pelagophyceae
pel;pel_clade_A --> Eukaryota Stramenopiles Pelagophyceae
    Pelagophyceae Pelagomonadales pelA
pel;pel_clade_B --> Eukaryota Stramenopiles Pelagophyceae
    Pelagophyceae Pelagomonadales pelB
pry --> Eukaryota Haptophyta Prymnesiophyceae
rap --> Eukaryota Rappemonad
rho --> Eukaryota Rhodophyta
str_env --> Eukaryota Stramenopiles
Strepto --> Strepto
xan --> xan

#We blasted all the sequences and those with a full alignment and
     identity >99% were annotated to genus.

blastall -p blastn -d nt -i seqtab-par_on16.photo.fa -e 0.001 -m
    8 -o seqtab-par_on16.bsn.tab -v 10 -b 10

Because of this blast result, we noticed that some sequences were
     mitochondrial segments and were removed manually. List of
    removed sequences:
SNV13881
SNV14167
SNV11730
SNV11084
SNV14641
SNV10225
SNV12585
SNV12882
```

```
SNV10859
SNV14617
SNV14128
SNV14174
SNV9399
SNV9469
SNV8073
SNV13811
SNV14301
SNV12559
SNV9677
SNV6157
SNV11829
SNV7863
SNV10876
SNV9746
SNV9037
SNV13550
SNV12778
SNV13496
SNV14468
SNV14599
SNV14470
SNV10819
SNV2961
SNV10774
SNV10432
SNV9541
SNV11167
SNV10664



This manually curated taxonomic file were named seqtab-par_on16.
    photo1351_taxa.txt

With the blast results we also noticed that all sequences
    annotated as "ML635J-21" (with the prefix Cyanobacteria)
    belonged to heterotrophic bacteria. So we removed them

grep -w "ML635J-21" seqtab-par_on16.photo1351_taxa.txt > ML635J
    -21.lst

cut -f 2 ML635J-21.lst > extractfrom1351.lst

grep -wf "890.lst" seqtab-par.txt.tmp.co | cut -f 1,2 | sed "s/
```

```
    N1N2/>N1N2/" | sed "s/\t/\n/" > on16_photo891.fasta
grep -vwf extractfrom1351.lst seqtab-par_on16.photo1351.tab >
    on16_photo891.tab
grep -vwf extractfrom1351.lst seqtab-par_on16.photo1351_taxa.txt
    > on16_photo891_taxa.txt
```

Task 3.4 Working files

Working files are located in github along this SOP. The following are the ones used to generate the figures:

- on16_photo891.tab = curated ASV table
- on16_photo891_taxa.txt = curated taxonomy table
- seqtab-par_on16.photo1351_sampledata.txt = file with the environmental data of the samples, including temperature and MLD (Dens_Thresh_MLD) used in figure 2a and 2b.

4 Figures

Figures created from amplicon datasets were done using R. Code is shown below.

Listing 7: N1N2figscripts

```
setwd("/Users/luisbolanos/Documents/OSU_postdoc/NAAMES/N1N2/
    analysisDada/Photo/FinalDatasets")

#Load libraries
library("phyloseq")
library("vegan")
library("DESeq2")
library("ggplot2")
library("dendextend")
library("tidyr")
library("viridis")
library("reshape")
library("dplyr")
library("phangorn")
library("data.table")
library("gplots")
library("VennDiagram")
library("UpSetR")

count_tab <- read.table("on16_photo891.tab", header=T, row.names
    =1, check.names=F)
```

```
sample_info_tab <- read.table("seqtab-par_on16.photo1351_
    sampledata.txt", header=T, row.names=1, check.names=F, sep ="
    \t")
tax_tab <- as.matrix(read.table("on16_photo891_taxa.txt", header=
    T, row.names=1, check.names=F, na.strings="", sep="\t"))
sample_info_tab$color[sample_info_tab$Type == "Subtropical-Spring
    "] <- "seagreen3"
sample_info_tab$color[sample_info_tab$Type == "Subtropical-Winter
    "] <- "darkgreen"
sample_info_tab$color[sample_info_tab$Type == "Subpolar-Spring"]
    <- "steelblue2"
sample_info_tab$color[sample_info_tab$Type == "Subpolar-Winter"]
    <- "royalblue"

sample_info_tab$colordepth[sample_info_tab$depth== "5"] <- "
    cadetblue1"
sample_info_tab$colordepth[sample_info_tab$depth== "25"] <- "
    cadetblue2"
sample_info_tab$colordepth[sample_info_tab$depth== "50"] <- "
    cadetblue3"
sample_info_tab$colordepth[sample_info_tab$depth== "75"] <- "
    aquamarine3"
sample_info_tab$colordepth[sample_info_tab$depth== "100"] <- "
    cadetblue4"
sample_info_tab$colordepth[sample_info_tab$depth== "150"] <- "
    cadetblue"
sample_info_tab$colordepth[sample_info_tab$depth== "200"] <- "
    gray47"
sample_info_tab$colordepth[sample_info_tab$depth== "300"] <- "
    gray27"

OTU = otu_table(count_tab, taxa_are_rows = TRUE)
TAX = tax_table(tax_tab)
SAM = sample_data(sample_info_tab)
physeqphot = phyloseq(OTU,TAX,SAM)

euph = get_variable(physeqphot, "depth") %in% c("5", "25", "50",
    "75", "100")
sample_data(physeqphot)$euph <- factor(euph)
phyeuph<-subset_samples(physeqphot, euph %in% TRUE)

phyeuphminV1 = prune_samples(sample_sums(phyeuph) > 1600, phyeuph
    )
phyeuphminV1filt= filter_taxa(phyeuphminV1, function(x) sum(x >
    2) > (0.015*length(x)), TRUE) #This is going to be considered
     the master PHYLOSEQ OBJECT, where we are going to derive
```

```
    most of the data
phyeuphminV1rel<-transform_sample_counts(phyeuphminV1filt,
    function(x){x / sum(x)})

####FIG 1 ####

#Hierarchical Clustering
deseq_counts <- phyloseq_to_deseq2(phyeuphminV1filt, ~Type)

deseq_counts_vst <- varianceStabilizingTransformation(deseq_
    counts)
vst_trans_count_tab <- assay(deseq_counts_vst)
euc_dist <- dist(t(vst_trans_count_tab))

euc_clust <- hclust(euc_dist, method="ward.D2")
euc_dend <- as.dendrogram(euc_clust, hang= -1,lwd = 3, lty = 3,
    sub = "")
dend_cols <- (sample_data(phyeuphminV1filt)$color)[order.
    dendrogram(euc_dend)]
labels_colors(euc_dend)<-dend_cols
namesdend<- (sample_data(phyeuphminV1filt)$name)[order.dendrogram
    (euc_dend)]
labels(euc_dend)<-namesdend

col_depth<-(sample_data(phyeuphminV1filt)$colordepth)

svg("euclplot.svg", height=11,width=12)
plot(euc_dend, xlab="", ylab="", main="", sub="", axes=FALSE)
colored_bars(col_depth,euc_dend, rowLabels = "Depth", cex.
    rowLabels=1, y_shift = -52)
par(cex=1)
title("NAAMES 1 and 2 dendogram", line=1)
par(cex=1)
title(ylab="VST Euclidean distance")
axis(2)
legend("topright", legend = c("Subtropical-Spring","Subtropical-
    Winter","Subpolar-Spring","Subpolar-Winter") , fill= c("
    seagreen3","darkgreen", "steelblue2", "royalblue"), bty="n",
    cex=1.00, title="Region-Season")
legend("topleft", legend = c("5m","25m","50m", "75m", "100m") ,
    fill= c("cadetblue1", "cadetblue2", "cadetblue3","aquamarine3
    ", "cadetblue4"), bty="n", cex=.8)
dev.off()

#MAP
library(ggmap)
```

```
library(maps)
library(mapdata)

##Fit Colors with clustering
##"Subtropical-Spring" "seagreen3"
##"Subtropical-Winter" "darkgreen"
##"Subpolar-Spring" "steelblue2"
##"Subpolar-Winter" "royalblue"


samps <- read.table("/Users/luisbolanos/Documents/OSU_postdoc/
    NAAMES/N1N2/Metadata/CoordsSt.txt", header=T,sep="\t")
image(x=-75:-15, y = 30:60, z = outer(0, 0), xlab = "lon", ylab =
     "lat")
map("world", add = TRUE, fill=TRUE,bg='light blue')

pdf("mapV1.pdf")
map("world", add = TRUE, fill=TRUE,bg='light blue')
points(samps[1:1,3], samps[1:1,2], pch=19, col="darkgreen", cex
    =1, type="o")
points(samps[2:3,3], samps[2:3,2], pch=19, col="royalblue", cex
    =1, type="o")
points(samps[4:7,3], samps[4:7,2], pch=19, col="darkgreen", cex
    =1, type="o")
points(samps[8:10,3], samps[8:10,2], pch=19, col="steelblue2",
    cex=1, type="o")
points(samps[11:12,3], samps[11:12,2], pch=19, col="seagreen3",
    cex=1, type="o")
dev.off()

#mapV1.pdf is overlapped and aligned in inkscape with the MDT map
     provided by Alice Della Penna and Peter Gaube. Colors were
    modified for better visualization when overlapped with the
    MDT.

####FIG 2 ####

#For Figure 2a and 2b we used a custom taxonomy file "on16_
    photo891_taxaNEW_V3"

##FIG 2a##

tax_tab <- as.matrix(read.table("on16_photo891_taxaNEW_V3.txt",
    header=T, row.names=1, check.names=F, na.strings="", sep="\t"
    ))
```

```
count_tab <- read.table("on16_photo891.tab", header=T, row.names
    =1, check.names=F)
sample_info_tab <- read.table("seqtab-par_on16.photo1351_
    sampledata.txt", header=T, row.names=1, check.names=F, sep ="
    \t")

OTU = otu_table(count_tab, taxa_are_rows = TRUE)
TAX = tax_table(tax_tab)
SAM = sample_data(sample_info_tab)
TREE = phy_tree(bs_inp)
physeqphot = phyloseq(OTU,TAX,SAM,TREE)

euph = get_variable(physeqphot, "depth") %in% c("5", "25", "50",
    "75", "100")
sample_data(physeqphot)$euph <- factor(euph)
phyeuph<-subset_samples(physeqphot, euph %in% TRUE)

phyeuphminN1N2 = prune_samples(sample_sums(phyeuph) > 1600,
    phyeuph)


phyeuphminV1<-tip_glom(phyeuphminN1N2, h = 0.02)#Disminuye de
    1351 a 886
phyeuphminV1filt= filter_taxa(phyeuphminV1, function(x) sum(x >
    2) > (0.015*length(x)), TRUE)

glomV1filt<-tax_glom(phyeuphminV1filt, taxrank="Taxa")
MeanStV2rel<-transform_sample_counts(glomV1filt, function(x){x /
    sum(x)})

taxaSubp<-as.data.frame(tax_table(MeanStV2rel)[,2])
ASV_frame<-as.data.frame(otu_table(MeanStV2rel))
ASV_frame[ "Taxa" ] <- taxaSubp[,1]

#dim(ASV_frame)
#[1] 22 72
ASV_frame2 <- ASV_frame[,-72]
rownames(ASV_frame2) <- ASV_frame[,72]

ASV_frw<-t(ASV_frame2)
md_to_add<-as.data.frame(sample_data(MeanStV2rel))[,c
    (1,2,12,14,33,36)]

final_2a<-cbind(ASV_frw,md_to_add)
```

```
fn2a_melt<-melt(final_2a,id.vars=c("name","depth","Station","TEMP
    ","Type","position"), measure.vars = c("Diatoms","
    Bolidophyceae","Dictyochophyceae","Pelagophyceae","
    Chrysophyceae","Micromonas","Bathycoccus","OstreococcusI","
    OstreococcusII","Cryptophyceae","Prymnesiophyceae","
    Rappemonad", "ASV357","Other plastid","ProchlorococcusHLI","
    ProchlorococcusHLII","ProchlorococcusLLI","SynechococcusI","
    SynechococcusIV","SynechococcusII","Other Cyanobacteria","Not
     assigned"))

coloresbarplot = c("Diatoms"="blue","Bolidophyceae"="cadetblue","
    Dictyochophyceae"="lightskyblue","Pelagophyceae"="aquamarine"
    ,"Chrysophyceae"="turquoise","Prymnesiophyceae"="
    darkgoldenrod3 ","Rappemonad"= "gold2","Cryptophyceae"="
    coral3" ,"Micromonas"="forestgreen","Bathycoccus"="limegreen"
    ,"OstreococcusII"="olivedrab","OstreococcusI"="palegreen4","
    PrasinophyceaeI"="greenyellow","Other plastid"="lightgreen","
    ASV357"="lemonchiffon3","ProchlorococcusHLI"="lightcoral","
    ProchlorococcusHLII"="hotpink","ProchlorococcusLLI"="maroon1"
    ,"SynechococcusI"="blueviolet","SynechococcusII"="
    mediumpurple","SynechococcusIV"="plum3","Other Cyanobacteria"
    ="mediumvioletred","others"="cornsilk4","Not assigned"="
    gray34")

Fig2a<-ggplot(fn2a_melt, aes(x = position, y = value, fill =
    variable)) + geom_bar(stat = "identity",width=.85)+ scale_
    fill_manual(values = coloresbarplot) + theme_bw()+ ylab("
    Relative contribution [%]") +theme(strip.background = element
    _blank(),strip.text.x = element_text(size=18),axis.text.y=
    element_text(size=16), axis.text.x=element_text(size=12,angle
     = 90, hjust = 1, vjust=.5), text = element_text(size=21),
    strip.text = element_text(size=22),axis.title.y=element_text(
    size=18),legend.text=element_text(size=14)) +facet_grid(~
    Station,scales = "free_x",space = "free_x")

svg("newfig2a_July.svg", width =13, height=10)
Fig2a
dev.off()

##Temperature heat map to be added to one side of the barplots

ggplot(fn2a_melt, aes(Station,depth)) + geom_tile(aes(fill = TEMP
    )) + scale_fill_gradientn(colours = topo.colors(2))

##FIG 2b##
```

```
##Winter pie charts##
pies<- read.table("piedwinter.txt", header=T, row.names=1, check.
    names=F)

pies_molten <- melt( pies, id.vars="Taxa", value.name="RelAb",
    variable.name="Sample" )
pies_molten$Taxa<-factor(pies_molten$Taxa, levels = c("
    Cyanobacteria", "ASV357", "Prasinophyta", "Cryptophyceae", "
    Rappemonad", "Prymnesiophyceae", "Stramenopiles:Chrysophyceae
    ", "Stramenopiles:Pelagophyceae", "Stramenopiles:
    Dictyochophyceae", "Stramenopiles:Bolidophyceae", "
    Stramenopiles:Diatoms", "others"))

coloresbarplot = c("Stramenopiles:Diatoms"="blue","Stramenopiles:
    Bolidophyceae"="cadetblue","Stramenopiles:Dictyochophyceae"="
    lightskyblue","Stramenopiles:Pelagophyceae"="aquamarine","
    Stramenopiles:Chrysophyceae"="turquoise","Prymnesiophyceae"="
    darkgoldenrod3","Rappemonad"="gold2","Cryptophyceae"="coral3"
    ,"Prasinophyta"="forestgreen","ASV357"="lemonchiffon3","
    Cyanobacteria"="purple","others"="cornsilk4")

svg("figwinterpied2.svg", height=12,width=10)
ggplot(pies_molten, aes(x = "", y = value, fill = Taxa)) +
geom_bar(stat = "identity", width = 1, position = position_fill()
    ) +
labs(x = NULL, y = NULL, fill = NULL ) +
coord_polar(theta = "y") +
facet_wrap( ~ variable)+ theme_bw()+ theme(axis.text.x=element_
    blank())+ scale_fill_manual(values = coloresbarplot)
dev.off()

##Spring pie charts##
piesSpring<- read.table("piedspring.txt", header=T, row.names=1,
    check.names=F)

pies_moltenSpring <- melt( piesSpring, id.vars="Taxa", value.name
    ="RelAb", variable.name="Sample" )
pies_moltenSpring$variable<- factor(pies_moltenSpring$variable,
    levels = c("NAAMES2-1_S21", "NAAMES2-2_S22", "NAAMES2-4_S37",
     "NAAMES2-5_S23", "NAAMES2-9_S24","NAAMES2-10_S39", "NAAMES2
    -11_S40", "NAAMES2-13_S41", "NAAMES2-14_S42", "NAAMES2-17_S25
    " ,"NAAMES2-18_S26", "NAAMES2-19_S44", "NAAMES2-21_S27"))

pies_moltenSpring$Taxa<-factor(pies_moltenSpring$Taxa, levels = c
    ("Cyanobacteria", "ASV357", "Prasinophyta", "Cryptophyceae",
    "Rappemonad", "Prymnesiophyceae", "Stramenopiles:
```

```
    Chrysophyceae", "Stramenopiles:Pelagophyceae", "Stramenopiles
    :Dictyochophyceae", "Stramenopiles:Bolidophyceae", "
    Stramenopiles:Diatoms", "others"))
svg("figspringpied2.svg", height=12,width=10)
ggplot(pies_moltenSpring, aes(x = "", y = value, fill = Taxa)) +
geom_bar(stat = "identity", width = 1, position = position_fill()
    ) +
labs(x = NULL, y = NULL, fill = NULL ) +
coord_polar(theta = "y") +
facet_wrap( ~ variable)+ theme_bw()+ theme(axis.text.x=element_
    blank())+ scale_fill_manual(values = coloresbarplot)
dev.off()

#Chlorophyl Background #

subpolar = get_variable(phyeuphminV1rel, "region") %in% "Subpolar
    "

sample_data(phyeuphminV1rel)$Subpolar <- factor(Subpolar)
physubp<-subset_samples(phyeuphminV1rel, subpolar %in% TRUE)

sp<-data.frame(sample_data(physubp))
hsp<-sp[,c(2,12,32)]

hsp[is.na(hsp)] <- 0

hspm<-melt( hsp, id.vars=c("Station", "depth"))

write.table(hspm,file= "hspm_CHLA.txt", quote=FALSE, sep = "\t")

hspm1<-read.table("hspm_CHLA.txt",sep = "\t", header=TRUE)

ggplot(hspm1, aes(Station, depth)) + geom_tile(aes(fill = value))
    +scale_fill_gradient(low = "white",high = "forestgreen",
    breaks=c(0,0.5,1,1.5,2,2.5,3,3.5,4,4.5,5), limits=c(0,5))

####FIG 3a Flow Cytometry STALCKED BARPLOT FCM ####

library("ggplot2")
library("reshape2")
library("dplyr")

dflow<-read.table("/Users/luisbolanos/Documents/OSU_postdoc/
    NAAMES/Metadata/Jason_Data/Rorganized/flowcytmetadata.txt",
    header=T, sep ="\t")
```

```
coloresFCM<-c(Prochlorococcus="lightcoral", Synechococcus="plum3"
    , Picoeukaryotes="forestgreen", Nanoeukaryotes="blue")

d5<-dflow[dflow$depth==5,]

inp<-d5 %>% select(Sample,Prochlorococcus, Synechococcus,
    Picoeukaryotes, Nanoeukaryotes, CHLA,cruise)

meltedDat<-melt(inp, id.vars = c("Sample","CHLA","cruise"))

stackedplot<-ggplot(data=meltedDat, aes(x=Sample, y=value, fill=
    factor(variable))) + geom_bar(stat="identity")+ scale_fill_
    manual(values =coloresFCM) +geom_line(aes(x=Sample, y=CHLA*
    10000000), stat="identity",color="red",group = 1, size=1.5)+
    scale_y_continuous(sec.axis = sec_axis(~./10000000,name = "
    Chlorophyll a [mg/m^3]"))+theme_bw()+facet_grid( . ~ cruise,
    scales = "free") + ylab("Cell counts [cells/mL]")+theme(axis.
    text.x = element_text(angle=30, size=6),axis.text.y= element_
    text(size=12),legend.text=element_text(size=13) )

svg("FCM_barplot.svg", width=12)
stackedplot
dev.off()

####FIG S2-S5####
PROFILES, we added CHAO1 index to the metadata file
(seqtab-par_on16.photo1351_envdataV2.txt)

#####MULTIPLOT FUNCTION#####
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=
    NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }
```

```
 if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(
        layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain
          this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE)
          )

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$
          row,
                                      layout.pos.col = matchidx$col))
    }
  }
}

set.seed(717)
phyeuphminrartodos = rarefy_even_depth(phyeuphminV1filt, sample.
    size = 1594)
dfchao<- estimate_richness(phyeuphminrartodos, measures="Chao1")

write.table(dfchao,file= "dfchao.txt", quote=FALSE, sep = "\t")

metadata <- read.table("/Users/luisbolanos/Documents/MisDrafts/
    InProgress/PhytoNAAMES/V5/Figures/seqtab-par_on16.photo1351_
    envdataV2.txt", header=T, row.names=1, sep="\t")

newdata <- subset(metadata, depth <= 100) #Only 0-100

sp_chao1<-ggplot(newdata, aes(x = depth, y = Chao1_phyto)) + geom
    _point(aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef), ) +scale_x_reverse() +scale_color_manual(values
    =c("darkcyan","darkblue","cyan2","chartreuse4","chartreuse3",
    "chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2"
    ,"brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type))+ ylab("Species richness [Chao1]") + theme(strip.
    background = element_blank(),strip.text.x = element_blank(),
    axis.text.y=element_blank(), axis.text.x=element_text(size
```

```
    =16), text = element_text(size=18),strip.text = element_text(
    size=20),axis.title.y=element_blank(),legend.text=element_
    text(size=18))

CHLAprof<-ggplot(newdata, aes(x = depth, y = CHLA)) + geom_point(
    aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type))+ ylab("Chlorophyll a [mg/m^3]")+ xlab("depth [m]"
    )+ theme(strip.background = element_blank(), axis.text.x=
    element_text(size=16), text = element_text(size=18), strip.
    text.y = element_blank(),legend.position = "none")

Photoperc<-ggplot(newdata, aes(x = depth, y = Chl.perce)) + geom_
    point(aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type)) + ylab("Phytoplankton sequences [%]")+xlab("depth
     [m]")+ theme(strip.background = element_blank(),strip.text.x
     = element_blank(),axis.text.y=element_blank(), axis.text.x=
    element_text(size=16), text = element_text(size=18),strip.
    text = element_text(size=20),strip.text.y = element_blank(),
    axis.title.y=element_blank(),legend.position = "none")

svg("/Users/luisbolanos/Documents/MisDrafts/InProgress/
    PhytoNAAMES/V5/Figures/bio_profile.svg", width=16, height=12)
multiplot(CHLAprof, Photoperc, sp_chao1, cols=3)
dev.off()



O2<-ggplot(newdata, aes(x = depth, y = O2)) + geom_point(aes(
    color=statioNodef), size=2) + geom_line(aes(color=statioNodef
    ))+ geom_errorbar(aes(ymin=O2-O2sd, ymax=O2+O2sd,color=
    statioNodef), width=.2) +scale_x_reverse() +scale_color_
    manual(values=c("darkcyan","darkblue","cyan2","chartreuse4","
    chartreuse3","chartreuse","darkolivegreen2","darkorange2","
    darkgoldenrod2","brown3")) + coord_flip()+ theme_bw() + facet
    _grid(rows = vars(Type))+ ylab("Dissolved Oxygen [mg/L]") +
    theme(strip.background = element_blank(),strip.text.x =
    element_blank(),axis.text.y=element_blank(), axis.text.x=
    element_text(size=16), text = element_text(size=18),strip.
```

```
    text = element_text(size=20),axis.title.y=element_blank(),
    legend.text=element_text(size=18),legend.position = "none")

Temp<-ggplot(newdata, aes(x = depth, y = TEMP)) + geom_point(aes(
    color=statioNodef), size=2) + geom_line(aes(color=statioNodef
    )) +scale_x_reverse()+scale_color_manual(values=c("darkcyan",
    "darkblue","cyan2","chartreuse4","chartreuse3","chartreuse","
    darkolivegreen2","darkorange2","darkgoldenrod2","brown3")) +
    coord_flip()+ theme_bw() + facet_grid(rows = vars(Type))+
    ylab("Temperature [C]")+ xlab("depth [m]")+ theme(strip.
    background = element_blank(), axis.text.x=element_text(size
    =16), text = element_text(size=18), strip.text.y = element_
    blank(),legend.position = "none")

Sal<-ggplot(newdata, aes(x = depth, y = Salinity)) + geom_point(
    aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type)) + ylab("Salinity [PSU]")+xlab("depth [m]")+ theme
    (strip.background = element_blank(),strip.text.x = element_
    blank(),axis.text.y=element_blank(), axis.text.x=element_text
    (size=16), text = element_text(size=18),strip.text = element_
    text(size=20),strip.text.y = element_blank(),axis.title.y=
    element_blank(),legend.position = "none")

svg("/Users/luisbolanos/Documents/MisDrafts/InProgress/
    PhytoNAAMES/V5/Figures/phy_profile.svg", width=16, height=12)
multiplot(Temp, Sal, O2, cols=3)
dev.off()



SiO4<-ggplot(newdata, aes(x = depth, y = SiO4)) + geom_point(aes(
    color=statioNodef), size=2) + geom_line(aes(color=statioNodef
    ))+scale_x_reverse() +scale_color_manual(values=c("darkcyan",
    "darkblue","cyan2","chartreuse4","chartreuse3","chartreuse","
    darkolivegreen2","darkorange2","darkgoldenrod2","brown3")) +
    coord_flip()+ theme_bw() + facet_grid(rows = vars(Type))+
    ylab("Silicate SiO4 [umol]") + theme(strip.background =
    element_blank(),strip.text.x = element_blank(),axis.text.y=
    element_blank(), axis.text.x=element_text(size=16), text =
    element_text(size=18),strip.text = element_text(size=20),axis
    .title.y=element_blank(),legend.text=element_text(size=18),
```

```
    legend.position = "none")

Nitrate<-ggplot(newdata, aes(x = depth, y =NO3)) + geom_point(aes
    (color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type))+ ylab("Nitrate NO3 [umol]")+ xlab("depth [m]")+
    theme(strip.background = element_blank(), axis.text.x=element
    _text(size=16), text = element_text(size=18), strip.text.y =
    element_blank(),legend.position = "none")

Ammonia<-ggplot(newdata, aes(x = depth, y = NH4)) + geom_point(
    aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type)) + ylab("Ammonia NH4 [umol]")+xlab("depth [m]")+
    theme(strip.background = element_blank(),strip.text.x =
    element_blank(),axis.text.y=element_blank(), axis.text.x=
    element_text(size=16), text = element_text(size=18),strip.
    text = element_text(size=20),strip.text.y = element_blank(),
    axis.title.y=element_blank(),legend.position = "none")

Phosphate<-ggplot(newdata, aes(x = depth, y = PO43)) + geom_point
    (aes(color=statioNodef), size=2) + geom_line(aes(color=
    statioNodef)) +scale_x_reverse()+scale_color_manual(values=c(
    "darkcyan","darkblue","cyan2","chartreuse4","chartreuse3","
    chartreuse","darkolivegreen2","darkorange2","darkgoldenrod2",
    "brown3")) + coord_flip()+ theme_bw() + facet_grid(rows =
    vars(Type)) + ylab("Phosphate PO4 [umol]")+xlab("depth [m]")+
     theme(strip.background = element_blank(),strip.text.x =
    element_blank(),axis.text.y=element_blank(), axis.text.x=
    element_text(size=16), text = element_text(size=18),strip.
    text = element_text(size=20),strip.text.y = element_blank(),
    axis.title.y=element_blank(),legend.position = "none")

svg("/Users/luisbolanos/Documents/MisDrafts/InProgress/
    PhytoNAAMES/V5/Figures/nutr_profile.svg", width=16, height
    =12)
multiplot(Nitrate, Ammonia, Phosphate,SiO4, cols=4)
dev.off()

####FIG S6####
```

PCoA

```
set.seed(717)
phyeuphminrartodos = rarefy_even_depth(phyeuphminV1filt, sample.
    size = 1594)


ordtodos = ordinate(phyeuphminrartodos, "PCoA", "bray")


#ord
p = plot_ordination(phyeuphminrartodos, ordtodos, color = "lat",
    shape = "cruise")
 p = p + geom_point(size = 3, alpha = 0.7) + scale_color_gradient
     (low = "#132B43", high ="red") + geom_text(aes(label =
     Station), size = 2.5, vjust = 2) + theme_bw()
p
```

Color Lines were added arbitrarily in inkscape to highlight the order and position of certain samples

```
####FIG S7####
#PLOT SHOWING COMMON

wint_pol= subset_samples(phyeuphminV1filt, Type=="Subpolar-Winter
    ")
wintpol<-unname(unlist(as.vector(rownames(otu_table(prune_taxa(
    taxa_sums(wint_pol) > 0, wint_pol))))))

spr_pol= subset_samples(phyeuphminV1filt, Type=="Subpolar-Spring"
    )
sprpol<-unname(unlist(as.vector(rownames(otu_table(prune_taxa(
    taxa_sums(spr_pol) > 0, spr_pol))))))

wint_trop= subset_samples(phyeuphminV1filt, Type =="Subtropical-
    Winter")
winttrop<-unname(unlist(as.vector(rownames(otu_table(prune_taxa(
    taxa_sums(wint_trop) > 0, wint_trop))))))

spr_trop= subset_samples(phyeuphminV1filt, Type =="Subtropical-
    Spring")
sprtrop<-unname(unlist(as.vector(rownames(otu_table(prune_taxa(
    taxa_sums(spr_trop) > 0, spr_trop))))))

listinput<-list(Subpolar_winter=wintpol,Subpolar_spring=sprpol,
    Subtropical_winter=winttrop,Subtropical_spring=sprtrop)
```

```
venn(list(wintpol, sprpol,winttrop,sprtrop))

upset(fromList(listinput), order.by = "freq", sets.bar.color = "
    #56B4E9")



 GET THE LIST of intersections
#FUNCTION from list 1#

fromList1 <- function (input) {
  # Same as original fromList()...
  elements <- unique(unlist(input))
  data <- unlist(lapply(input, function(x) {
      x <- as.vector(match(elements, x))
      }))
  data[is.na(data)] <- as.integer(0)
  data[data != 0] <- as.integer(1)
  data <- data.frame(matrix(data, ncol = length(input), byrow = F
      ))
  data <- data[which(rowSums(data) != 0), ]
  names(data) <- names(input)
  # ... Except now it conserves your original value names!
  row.names(data) <- elements
  return(data)
  }

#Example list:

getlist<-fromList1(listinput) #getlist is going to be a dataframe
     with 1 and 0s

#Then use getintersect to generate the different groups seen in
    the upsetR plot

get_intersect_members <- function (x, ...){
  require(dplyr)
  require(tibble)
  x <- x[,sapply(x, is.numeric)][,0<=colMeans(x[,sapply(x, is.
      numeric)],na.rm=T) & colMeans(x[,sapply(x, is.numeric)],na.
      rm=T)<=1]
  n <- names(x)
  x %>% rownames_to_column() -> x
  l <- c(...)
  a <- intersect(names(x), l)
```

```
  ar <- vector('list',length(n)+1)
  ar[[1]] <- x
  i=2
  for (item in n) {
    if (item %in% a){
      if (class(x[[item]])=='integer'){
        ar[[i]] <- paste(item, '>= 1')
        i <- i + 1
      }
    } else {
      if (class(x[[item]])=='integer'){
        ar[[i]] <- paste(item, '== 0')
        i <- i + 1
      }
    }
  }
  do.call(filter_, ar) %>% column_to_rownames() -> x
  return(x)
}


######Now get intersect for all the combinations (4 groups = 15
    combinations)########

uniqsubpwint<-rownames(get_intersect_members(getlist, "Subpolar_
    winter"))
Uniqsubpspr<-rownames(get_intersect_members(getlist, "Subpolar_
    spring"))
Uniqsubtrspr<-rownames(get_intersect_members(getlist, "
    Subtropical_spring"))
Uniqsubtrwint<-rownames(get_intersect_members(getlist, "
    Subtropical_winter"))
All4<-rownames(get_intersect_members(getlist, "Subpolar_winter",
    "Subpolar_spring", "Subtropical_spring", "Subtropical_winter"
    ))
polar2<-rownames(get_intersect_members(getlist, "Subpolar_winter"
    ,"Subpolar_spring"))
Trop2<-rownames(get_intersect_members(getlist, "Subtropical_
    winter","Subtropical_spring"))
Spr2<-rownames(get_intersect_members(getlist, "Subpolar_spring","
    Subtropical_spring"))
wint2<-rownames(get_intersect_members(getlist, "Subpolar_winter",
    "Subtropical_winter"))
cruz1<-rownames(get_intersect_members(getlist, "Subpolar_winter",
    "Subtropical_spring"))
```

```
cruz2<-rownames(get_intersect_members(getlist,"Subpolar_spring","
    Subtropical_winter"))
Tres1<-rownames(get_intersect_members(getlist, "Subpolar_winter",
    "Subpolar_spring", "Subtropical_winter"))
Tres2<-rownames(get_intersect_members(getlist, "Subpolar_winter",
    "Subpolar_spring", "Subtropical_spring"))
Tres3<-rownames(get_intersect_members(getlist, "Subtropical_
    winter","Subtropical_spring","Subpolar_winter"))
Tres4<-rownames(get_intersect_members(getlist, "Subtropical_
    winter","Subtropical_spring","Subpolar_spring"))

#Sacar del total de reads en los 4 grupos el porcentaje que hace
    los subgroups)
sum(sample_sums(phyeuphminV1filt)) #TOTAL para usar como
    complemento
[1] 1642543

uniqsubpwintcount<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% uniqsubpwint)))
Uniqsubpsprcount<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Uniqsubpspr)))
Uniqsubtrsprcount<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Uniqsubtrspr)))
Uniqsubtrwintcount<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
     rownames(tax_table(physeq)) %in% Uniqsubtrwint)))
All4count<-sum(sample_sums(subset_taxa(phyeuphminV1filt, rownames
    (tax_table(physeq)) %in% All4)))
polar2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% polar2)))
Trop2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Trop2)))
Spr2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt, rownames
    (tax_table(physeq)) %in% Spr2)))
wint2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% wint2)))
cruz1count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% cruz1)))
cruz2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% cruz2)))
Tres1count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres1)))
Tres2count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres2)))
Tres3count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres3)))
```

```
Tres4count<-sum(sample_sums(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres4)))

#Counts were organized in a tsv file.

upsetR bars use use the counts of Phylum

Etc
data.frame(tax_table(subset_taxa(phyeuphminV1filt, rownames(tax_
    table(physeq)) %in% Tres1))))[,1:3])
(data.frame(tax_table(subset_taxa(phyeuphminV1filt, rownames(tax_
    table(physeq)) %in% Tres2))))
(data.frame(tax_table(subset_taxa(phyeuphminV1filt, rownames(tax_
    table(physeq)) %in% Tres3))))
(data.frame(tax_table(subset_taxa(phyeuphminV1filt, rownames(tax_
    table(physeq)) %in% Tres4))))

#Counts were organized in a tsv file.

#Read phylotypes distribution


#Create upsetRinputbarsV1.txt

PS_PW_TW1<-data.frame(tax_table(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres1))[,1:3]) #foreachone
PS_PW_TS2<-data.frame(tax_table(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres2))[,1:3])
TS_TW_PS4<-data.frame(tax_table(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres4))[,1:3])
TS_TW_PW3<-data.frame(tax_table(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% Tres3))[,1:3])
Unique_Subtr_wint<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in%
    Uniqsubtrwint))[,1:3])
Unique_Subtr_spr<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in%
    Uniqsubtrspr))[,1:3])
Unique_Subpol_spr<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in%
    Uniqsubpspr))[,1:3])
Unique_Subpol_wint<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in%
    uniqsubpwint))[,1:3])
All4_all<-data.frame(tax_table(subset_taxa(phyeuphminV1filt,
    rownames(tax_table(physeq)) %in% All4))[,1:3])
```

```
Polar_2seasons<-data.frame(tax_table(subset_taxa(phyeuphminV1filt
    , rownames(tax_table(physeq)) %in% polar2))[,1:3])
Subtropical_2seasons<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in% Trop2))
    [,1:3])
Spring_2regions<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in% Spr2))
    [,1:3])
Winter_2regions<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in% wint2))
    [,1:3])
PolWint_TropSpring<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in% cruz1))
    [,1:3])
PolSpr_Tropwint<-data.frame(tax_table(subset_taxa(
    phyeuphminV1filt, rownames(tax_table(physeq)) %in% cruz2))
    [,1:3])


Unique_Subtr_wint %>% group_by_all() %>% summarise(COUNT = n())
Unique_Subtr_spr %>% group_by_all() %>% summarise(COUNT = n())
Unique_Subpol_spr %>% group_by_all() %>% summarise(COUNT = n())
Unique_Subpol_wint %>% group_by_all() %>% summarise(COUNT = n())
All4_all%>% group_by_all() %>% summarise(COUNT = n())
Polar_2seasons %>% group_by_all() %>% summarise(COUNT = n())
Subtropical_2seasons %>% group_by_all() %>% summarise(COUNT = n()
    )
Spring_2regions %>% group_by_all() %>% summarise(COUNT = n())
Winter_2regions %>% group_by_all() %>% summarise(COUNT = n())
PolWint_TropSpring %>% group_by_all() %>% summarise(COUNT = n())
PolSpr_Tropwint %>% group_by_all() %>% summarise(COUNT = n())
PS_PW_TW1 %>% group_by_all() %>% summarise(COUNT = n())
PS_PW_TS2 %>% group_by_all() %>% summarise(COUNT = n())
TS_TW_PS4 %>% group_by_all() %>% summarise(COUNT = n())
TS_TW_PW3 %>% group_by_all() %>% summarise(COUNT = n())

#barsSNVs<-read.table("upsetRinputbars.txt", header=T, row.names
    =1, check.names=F)#OLD

coloresbarplot = c("Stramenopiles:Diatoms"="blue","Stramenopiles:
    Bolidophyceae"="cadetblue","Stramenopiles:Dictyochophyceae"="
    lightskyblue","Stramenopiles:Pelagophyceae"="aquamarine","
    Stramenopiles:Chrysophyceae"="turquoise","Prymnesiophyceae"="
    darkgoldenrod3","Rappemonad"="gold2","Cryptophyceae"="coral3"
    ,"Prasinophyta"="forestgreen","Eusiphoniidae"="lemonchiffon3"
    ,"Cyanobacteria"="purple","Unassigned"="cornsilk4", Alveolata
```

```
    ="rosybrown3", Rhodophyta="tan1")


barsSNVs<-read.table("upsetRinputbarsV1.txt", header=T, row.names
    =1, check.names=F)#NEW
melted <- melt(barsSNVs, id="Position")
ggplot(melted, aes(x = Position, y = value, fill = variable)) +
    geom_bar(stat = "identity",width=.5) + theme_bw()+ scale_fill
    _manual(values = coloresbarplot)

upset(fromList(listinput), order.by = "freq", sets.bar.color = "
    #56B4E9")

###Now with total reads

barsVenn<-read.table("venn_counts.txt", header=T, row.names=1,
    check.names=F)

meltedTot <- melt(barsVenn, id="Position")
ggplot(meltedTot, aes(x = Position, y = value, fill = variable))
    + geom_bar(stat = "identity",width=.5) + theme_bw()

svg(barplotforupster.svg)
ggplot(melted, aes(x = Position, y = value, fill = variable)) +
    geom_bar(stat = "identity",width=.5) + theme_bw()+ scale_fill
    _manual(values = coloresbarplot)

####FIG S8####

#DESEQ2 DIFFERENTIAL ABUNDANCE


pol_wintspr= subset_samples(phyeuphminV1filt, Type =="Subpolar-
    Winter" | Type=="Subpolar-Spring")
trop_wintspr= subset_samples(phyeuphminV1filt, Type ==""
    Subtropical-Winter" | Type=="Subtropical-Spring")
spr_poltrop= subset_samples(phyeuphminV1filt, Type =="Subtropical
    -Spring" | Type=="Subpolar-Spring")
wint_poltrop= subset_samples(phyeuphminV1filt, Type ==""
    Subtropical-Winter" | Type=="Subpolar-Winter")

diagdds_pol = phyloseq_to_deseq2(pol_wintspr, ~ Type)
diagdds_pol = DESeq(diagdds_pol, test="Wald", fitType="parametric
    ")

res = results(diagdds_pol, cooksCutoff = FALSE)
```

```
alpha = 0.01
sigtab = res[which(res$padj < alpha), ]
sigtab = cbind(as(sigtab, "data.frame"), as(tax_table(
    phyeuphminV1filt)[rownames(sigtab), ], "matrix"))
head(sigtab)

diagdds_trop = phyloseq_to_deseq2(trop_wintspr, ~ Type)
diagdds_trop = DESeq(diagdds_trop, test="Wald", fitType="
    parametric")

restrop = results(diagdds_trop, cooksCutoff = FALSE)
alpha = 0.01
sigtabtrop = restrop[which(restrop$padj < alpha), ]
sigtabtrop = cbind(as(sigtabtrop, "data.frame"), as(tax_table(
    phyeuphminV1filt)[rownames(sigtabtrop), ], "matrix"))
head(sigtabtrop)

diagdds_spr = phyloseq_to_deseq2(spr_poltrop, ~ Type)
diagdds_spr = DESeq(diagdds_spr, test="Wald", fitType="parametric
    ")

resspr = results(diagdds_spr, cooksCutoff = FALSE)
alpha = 0.01
sigtabspr = resspr[which(resspr$padj < alpha), ]
sigtabspr = cbind(as(sigtabspr, "data.frame"), as(tax_table(
    phyeuphminV1filt)[rownames(sigtabspr), ], "matrix"))
head(sigtabspr)

diagdds_wint = phyloseq_to_deseq2(wint_poltrop, ~ Type)
diagdds_wint = DESeq(diagdds_wint, test="Wald", fitType="
    parametric")

reswint = results(diagdds_wint, cooksCutoff = FALSE)
alpha = 0.01
sigtabwint = reswint[which(reswint$padj < alpha), ]
sigtabwint = cbind(as(sigtabwint, "data.frame"), as(tax_table(
    phyeuphminV1filt)[rownames(sigtabwint), ], "matrix"))
head(sigtabwint)

#PLOT CIRCLES PHYLA_CLASS#

phylumcolors<-c("new_euk_C" = "black","Chlorophyta" = "
    chartreuse4", "Cryptophyta" = "deeppink","Cyanobacteria"="
    purple", "Alveolata"="brown", "new_euk_A"="darkseagreen","
    Haptophyta"="orange","Eusiphoniidae"="gray", "NA" = "
    firebrick", "Stramenopiles"="darkturquoise"," Rhodophyta"="
```

```
    antiquewhite3")

theme_set(theme_bw())
scale_fill_discrete <- function(palname = "Set1", ...) {
    scale_fill_brewer(palette = palname, ...)
}

_____________________________________Differential
#AFTER PLOTTING, We want to know from this differential
    abundances, which ones didnt have a representative in winter.
     Overwintering/NonOverwintering Add the Overwintering/No OV
    condition

#1st get names of the differential phylotypes from the sigtab
    files

sigtab_names<-row.names(sigtab)
sigtabwint_names<-row.names(sigtabwint)
sigtabspr_names<-row.names(sigtabspr)
sigtabtrop_names<-row.names(sigtabtrop)

#Extract them from the main Phyloseq object

differentPolar <- subset(otu_table(phyeuphminV1filt), rownames(
    otu_table(phyeuphminV1filt)) %in% sigtab_names)

differentTrop <- subset(otu_table(phyeuphminV1filt), rownames(otu
    _table(phyeuphminV1filt)) %in% sigtabtrop_names)

differentwint<- subset(otu_table(phyeuphminV1filt), rownames(otu_
    table(phyeuphminV1filt)) %in% sigtabwint_names)

differentspr<- subset(otu_table(phyeuphminV1filt), rownames(otu_
    table(phyeuphminV1filt)) %in% sigtabspr_names)

#create a phyloseq for each one (collapse info and makes easier
    to work with them) Just for the moment only seasonal
    comparison POLAR and TROPCIAL

diffpolar_physeq <- merge_phyloseq(differentPolar, tax_table(
    phyeuphminV1filt), sample_data(phyeuphminV1filt), phy_tree(
    phyeuphminV1filt))

difftrop_physeq <- merge_phyloseq(differentTrop, tax_table(
    phyeuphminV1filt), sample_data(phyeuphminV1filt), phy_tree(
    phyeuphminV1filt))
```

```
sumsPol<-rowSums(otu_table(subset_samples(diffpolar_physeq,
    cruise=="N1")))==0 #if it is == 0 means that we didnt detect
    in any samples above 100m

 sumstrop<-rowSums(otu_table(subset_samples(difftrop_physeq,
     cruise=="N1"))) ==0 #if it is == 0 means that we didnt detect
      in any samples above 100m

dfPol <- data.frame(names(sumsPol),as.vector(sumsPol))
dftrop <-data.frame(names(sumstrop),as.vector(sumstrop))

 dfPol$condition[dfPol$as.vector.sums== "TRUE"] = "non-
     overwintering"
 dfPol$condition[dfPol$as.vector.sums== "FALSE"] = "overwintering
     "

 dftrop$condition[dftrop$as.vector.sums== "TRUE"] = "non-
     overwintering"
 dftrop$condition[dftrop$as.vector.sums== "FALSE"] = "
     overwintering"

#Add "condition" column to sigtab and sigtabtrop

sampPol <- dfPol[,-1]
rownames(sampPol) <- dfPol[,1]

sampTrop <- dftrop[,-1]
 rownames(sampTrop) <- dftrop[,1]

Sigtab1<-cbind(sigtab, sampPol[, "condition"][match(rownames(
    sigtab), rownames(sampPol))])
colnames(Sigtab1)[14] <- "condition" #WORK with Sigtab1 for the
    plots

Sigtab2<-cbind(sigtabtrop, sampTrop[, "condition"][match(rownames
    (sigtabtrop), rownames(sampTrop))])
colnames(Sigtab2)[14] <- "condition" #WORK with Sigtab2 for the
    plots

# Phylum order POLAR
x = tapply(Sigtab1$log2FoldChange, Sigtab1$Phylum, function(x)
    max(x))
x = sort(x, TRUE)
Sigtab1$Phylum = factor(as.character(Sigtab1$Phylum), levels=
    names(x))
```

```
# Genus order POLAR
x = tapply(Sigtab1$log2FoldChange, Sigtab1$Class, function(x) max
    (x))
x = sort(x, TRUE)
Sigtab1$Class = factor(as.character(Sigtab1$Class), levels=names(
    x))

write.table(Sigtab1,file= "DeseqSub_Sigtab1.txt", quote=FALSE,
    sep = "\t")

jitter <- position_jitter(width = 0.2, height = 0.0) #just to
    shake the symbols a little bit

polar<-ggplot(Sigtab1, aes(x=Class, y=log2FoldChange, color=
    Phylum, label=rownames(Sigtab1), shape=condition)) + geom_
    point(size=3, position=jitter) + theme(axis.text.x = element_
    text(angle = -45, hjust = 0, vjust=0.5))+ geom_text(vjust="
    inward",hjust="inward", size = 2,position=jitter) +geom_hline
    (yintercept = 0, color="black")+geom_hline(yintercept = c
    (-5,5), color="red")+ggtitle("subpolar region")+scale_color_
    manual(name = "Phylum",values=phylumcolors) + coord_flip()


# Phylum order TROPICAL

x1 = tapply(Sigtab2$log2FoldChange, Sigtab2$Phylum, function(x1)
    max(x1))
x1 = sort(x1, TRUE)
Sigtab2$Phylum = factor(as.character(Sigtab2$Phylum), levels=
    names(x1))
# Genus order POLAR
x1 = tapply(Sigtab2$log2FoldChange, Sigtab2$Class, function(x1)
    max(x1))
x1 = sort(x1, TRUE)
Sigtab2$Class = factor(as.character(Sigtab2$Class), levels=names(
    x1))

write.table(Sigtab2,file= "DeseqSubtr_Sigtab2.txt", quote=FALSE,
    sep = "\t")

tropical<-ggplot(Sigtab2, aes(x=Class, y=log2FoldChange, color=
    Phylum, label=rownames(Sigtab2), shape=condition)) + geom_
    text(vjust="inward",hjust="inward", size = 2,position=jitter)
    +geom_point(size=3,position=jitter) + theme(axis.text.x =
    element_text(angle = -45, hjust = 0, vjust=0.5)) +geom_hline(
    yintercept = 0, color="black")+geom_hline(yintercept = c
```

```
    (-5,5), color="red")+ggtitle("subtropical region")+scale_
    color_manual(name = "Phylum",values=phylumcolors) + coord_
    flip()

# Phylum order SPRING

x2 = tapply(sigtabspr$log2FoldChange, sigtabspr$Phylum, function(
    x2) max(x2))
x2 = sort(x2, TRUE)
sigtabspr$Phylum = factor(as.character(sigtabspr$Phylum), levels=
    names(x2))
# Genus order POLAR
x2 = tapply(sigtabspr$log2FoldChange, sigtabspr$Class, function(
    x2) max(x2))
x2 = sort(x2, TRUE)
sigtabspr$Class = factor(as.character(sigtabspr$Class), levels=
    names(x2))

spring<-ggplot(sigtabspr, aes(x=Class, y=log2FoldChange, color=
    Phylum, label=rownames(sigtabspr))) + geom_point(size=3)+
    geom_text(vjust="inward",hjust="inward", size = 2) + theme(
    axis.text.x = element_text(angle = -45, hjust = 0, vjust=0.5)
    ) +geom_hline(yintercept = 0, color="black")+geom_hline(
    yintercept = c(-5,5), color="red")+ggtitle("Significant
    differential abundances between subpolar and subtropical
    North Atlantic regions in spring")+scale_color_manual(name =
    "Phylum",values=phylumcolors)

# Phylum order Winter

x3 = tapply(sigtabwint$log2FoldChange, sigtabwint$Phylum,
    function(x3) max(x3))
x3 = sort(x3, TRUE)
sigtabwint$Phylum = factor(as.character(sigtabwint$Phylum),
    levels=names(x3))
# Genus order POLAR
x3 = tapply(sigtabwint$log2FoldChange, sigtabwint$Class, function
    (x3) max(x3))
x3 = sort(x3, TRUE)
sigtabwint$Class = factor(as.character(sigtabwint$Class), levels=
    names(x3))

winter<-ggplot(sigtabwint, aes(x=Class, y=log2FoldChange, color=
    Phylum, label=rownames(sigtabwint))) + geom_point(size=3)+
    geom_text(vjust="inward",hjust="inward", size = 2) + theme(
    axis.text.x = element_text(angle = -45, hjust = 0, vjust=0.5)
```

```
    ) +geom_hline(yintercept = 0, color="black")+geom_hline(
    yintercept = c(-5,5), color="red")+ggtitle("Significant
    differential abundances between subpolar and subtropical
    North Atlantic regions in winter")+scale_color_manual(name =
    "Phylum",values=phylumcolors)

svg("regiondeseq.svg", width=15, height=9)
multiplot(polar,tropical, cols=2)
dev.off()

svg("cruisedeseq.svg", width=12,height=12)
multiplot(winter,spring, cols=1)
dev.off()

####2nd part deseq ##### Corroborate condition in the winter
    100-300m

#GetNOW
NOW_polar<-row.names(subset(Sigtab1 , condition == "non-
    overwintering"))
NOW_trop<-row.names(subset(Sigtab2 , condition == "non-
    overwintering"))

#SacarPhyseq con las NOW
NOW_polPHYSEQ <- subset(otu_table(phyeuphminV1filt), rownames(otu
    _table(phyeuphminV1filt)) %in% NOW_polar)
NOW_tropPHYSEQ <- subset(otu_table(phyeuphminV1filt), rownames(
    otu_table(phyeuphminV1filt)) %in% NOW_trop)

#Get 100 to 300
deep = get_variable(physeqphot, "depth") %in% c("150", "200", "
    300")
sample_data(physeqphot)$deep <- factor(deep)
phydeep<-subset_samples(physeqphot, deep %in% TRUE)

phydeepminN1N2 = prune_samples(sample_sums(phydeep) > 1600,
    phydeep)


phydeepminV1<-tip_glom(phydeepminN1N2, h = 0.02)#Disminuye de
    1351 a 886
phydeepminV1filt= (filter_taxa(phydeepminV1, function(x) sum(x >
    2) > (0.015*length(x)), TRUE))
```

```
NOW_polPHYSEQ300 <- subset(otu_table(phydeepminV1filt), rownames(
    otu_table(phydeepminV1filt)) %in% NOW_polar)
NOW_tropPHYSEQ300 <- subset(otu_table(phydeepminV1filt), rownames
    (otu_table(phydeepminV1filt)) %in% NOW_trop)
```