Relative abundance Spiny Lobster eDNA

Aiko Abo Dominguez

2025-08-07

Contents

Downloads

installing and loading packages

```
# Downloads

# Function to install CRAN packages if not already installed
install_if_missing <- function(pkgs) {
  installed <- installed.packages()[, "Package"]
  to_install <- setdiff(pkgs, installed)
  if (length(to_install) > 0) {
    install.packages(to_install)
  }
}

# CRAN packages
#install_if_missing(c("ggplot2", "ape", "vegan", "tidyverse", "sf", "mapview", "leaflet", "webshot2", "car", "

# devtools (for GitHub installs)
#install_if_missing("devtools")
library(devtools)

# GitHub packages
#if (!"pairwiseAdonis" %in% installed.packages()[, "Package"]) {
#  install_github("pmartinezarbizu/pairwiseAdonis/pairwiseAdonis")
#}

#if (!"qiime2R" %in% installed.packages()[, "Package"]) {
#  install_github("jbisanz/qiime2R")
#}

#devtools::install_github("kassambara/ggpubr")

# Bioconductor setup
# if (!"BiocManager" %in% installed.packages()[, "Package"]) {
#   install.packages("BiocManager")
# }
```

```r
# Install Bioconductor packages only if not present
# bioc_pkgs <- c("phyloseq", "limma", "microbiome", "ANCOMBC")
# for (pkg in bioc_pkgs) {
#   if (!requireNamespace(pkg, quietly = TRUE)) {
#     BiocManager::install(pkg)
#   }
# }

# Install microViz package from R Universe
# install.packages(
#   "microViz",
#   repos = c(davidbarnett = "https://david-barnett.r-universe.dev", getOption("repos"))
# )

# Loading required libraries
library(qiime2R)
library(sf)
library(mapview)
library(tidyverse)
library(phyloseq)
library(ggplot2)
library(ape)
library(vegan)
library(leaflet)
library(webshot2)
library(ggpmisc)
library(ggpubr)
library(RColorBrewer)
library(car)
library(rstatix)
library(limma)
library(microbiome)
library(pairwiseAdonis)
library(microViz)
```

Importing data and metadata editing

##Editing metadata file

```r
#import file
metadata <- read.csv("/Users/aikoabodominguez/Documents/eDNA_spiny_lobster/sample_metadata_PAV1_lobster_eDNA.c

#change the name of the first column
metadata$sample_id_old <- metadata$sample.id
colnames(metadata)[1] <- "sample_id"

#change "don" column to "site ID". "don" is what the FWC labels these sites as.
colnames(metadata)[4] <- "site_id"

#make function to truncate ID name
truncate_before_second_underscore <- function(x) {
  sapply(strsplit(x, "_"), function(parts) paste(parts[1:2], collapse = "_"))
}

metadata$sample_id <- truncate_before_second_underscore(metadata$sample_id)
#View(metadata)

#set column 1 as row names
rownames(metadata) <- metadata$sample_id
```

```r
#make sure it is a dataframe
metadata <- as.data.frame(metadata)
#View(metadata)

#setting variables to be numeric, as appropriate
metadata$Lat <- as.numeric(metadata$Lat)
metadata$Long <- as.numeric(metadata$Long)
metadata$Total_Lobsters <- as.numeric(metadata$Total_Lobsters)
metadata$EstimatedBiomass <- as.numeric(metadata$EstimatedBiomass)

str(metadata)
```

```
## 'data.frame':    36 obs. of  17 variables:
##  $ sample_id                : chr  "P1546_36" "P1546_12" "P1546_29" "P1546_37" ...
##  $ forward.absolute.filepath: chr  "/scratch/16S-M/Data/fastp/P1546_36_RS062024_8_1_16S_M_S36_L001_R1_fastp
##  $ reverse.absolute.filepath: chr  "/scratch/16S-M/Data/fastp/P1546_36_RS062024_8_1_16S_M_S36_L001_R2_fastp
##  $ site_id                  : chr  "1" "1" "1" "2" ...
##  $ Replicate                : chr  "8_1" "8_3" "8_2" "9_3" ...
##  $ sample_date              : chr  "6_26_2024" "6_26_2024" "6_26_2024" "6_26_2024" ...
##  $ sample_time              : chr  "11:41" "11:41" "11:41" "12:54" ...
##  $ Total_Lobsters_ren       : int  1 1 1 25 25 25 27 27 27 10 ...
##  $ Total_Lobsters           : num  1 1 1 23 23 23 20 20 20 9 ...
##  $ PAV_count                : int  0 0 0 1 1 1 0 0 0 0 ...
##  $ EstimatedBiomass         : num  28.5 28.5 28.5 829.3 829.3 ...
##  $ Lat                      : num  24.8 24.8 24.8 24.8 24.8 ...
##  $ Long                     : num  -81 -81 -81 -81 -81 ...
##  $ site_type                : chr  "Block" "Block" "Block" "Block" ...
##  $ nBlocks                  : int  12 12 12 50 50 50 25 25 25 23 ...
##  $ group                    : chr  "north" "north" "north" "north" ...
##  $ sample_id_old            : chr  "P1546_36_RS062024_8_1_16S_M_S36_L001" "P1546_12_RS062024_8_3_16S_M_S12_
```

```r
# Add a  column sample_id_location by pasting site_id and Replicate_new
metadata$Replicate_new <- sub(".*_", "", metadata$Replicate)
metadata$sample_id_location <- paste(metadata$site_id, metadata$Replicate_new, sep = "_")

#add column for sample site ID based on latitude
sites <- unique(metadata$Long)

#sites
metadata <- metadata %>%
  mutate(site_letter = ifelse(Long==sites[1], "A",
                   ifelse(Long==sites[2], "B",
                   ifelse(Long==sites[3], "C",
                   ifelse(Long==sites[4], "D",
                   ifelse(Long==sites[5], "E",
                   ifelse(Long==sites[6], "F",
                   ifelse(Long==sites[7],"G",
                   ifelse(Long==sites[8], "H",
                   ifelse(Long==sites[9], "I",
                   ifelse(Long==sites[10], "J",
                   ifelse(Long==sites[11], "K",
                   ifelse(Long==sites[12], "L", NA)))))))))))))
unique(metadata$site_id)
```

```
##  [1] "1"    "2"    "5"    "7"    "8"    "10"   "11"   "14"   "17"   "20"
## [11] "21"   "4_24"
```

```r
#View(metadata)

#avg lobster mass
metadata$avg_lobster_mass <- metadata$EstimatedBiomass/metadata$Total_Lobsters
```

```r
mean(metadata$avg_lobster_mass)
```

```
## [1] 45.55232
```

```r
sd(metadata$avg_lobster_mass)
```

```
## [1] 11.5607
```

Exporting coordinates

```r
coordinates <- data.frame(
  site_id = unique(metadata$site_id),
  lat = unique(metadata$Lat),
  long = unique(metadata$Long)
)

#write.csv(coordinates,file="PAV1_eDNA_sample_site_coordinates.csv")
```

##Importing ASV sequences, taxa table, and phylogenetic tree

```r
#sequence table from dada2 of forward reads, only eukaryotes
seq_qza <- read_qza("/Users/aikoabodominguez/Documents/eDNA_spiny_lobster/seq_ASVs_lobC16S_euks_fwd.qza")
seq_table <- as.data.frame(seq_qza$data)


#ASV table from dada2 of forward reads, only eukaryotes
ASV_qza <- read_qza("/Users/aikoabodominguez/Documents/eDNA_spiny_lobster/table_ASVs_lobC16S_euks_fwd.qza")
ASV_table <- as.matrix(ASV_qza$data)
#View(ASV_table)

#truncating ASV column names to match "metadata" truncated sample names
colnames(ASV_table) <- truncate_before_second_underscore(colnames(ASV_table))
colnames(ASV_table)
```

```
##  [1] "P1546_10" "P1546_11" "P1546_12" "P1546_13" "P1546_14" "P1546_15"
##  [7] "P1546_16" "P1546_17" "P1546_18" "P1546_19" "P1546_20" "P1546_21"
## [13] "P1546_22" "P1546_23" "P1546_24" "P1546_25" "P1546_26" "P1546_27"
## [19] "P1546_28" "P1546_29" "P1546_3"  "P1546_30" "P1546_31" "P1546_32"
## [25] "P1546_33" "P1546_34" "P1546_35" "P1546_36" "P1546_37" "P1546_38"
## [31] "P1546_4"  "P1546_5"  "P1546_6"  "P1546_7"  "P1546_8"  "P1546_9"
```

```r
metadata[,"sample_id"]
```

```
##  [1] "P1546_36" "P1546_12" "P1546_29" "P1546_37" "P1546_4"  "P1546_34"
##  [7] "P1546_8"  "P1546_10" "P1546_14" "P1546_38" "P1546_25" "P1546_22"
## [13] "P1546_5"  "P1546_35" "P1546_32" "P1546_9"  "P1546_6"  "P1546_31"
## [19] "P1546_15" "P1546_23" "P1546_26" "P1546_7"  "P1546_17" "P1546_3"
## [25] "P1546_21" "P1546_18" "P1546_28" "P1546_30" "P1546_33" "P1546_27"
## [31] "P1546_24" "P1546_19" "P1546_20" "P1546_13" "P1546_16" "P1546_11"
```

```r
#check that it worked
rownames(metadata) <- metadata$sample_id
all(rownames(metadata)%in%colnames(ASV_table))
```

```
## [1] TRUE
```

```r
#import taxa from dada2 of forward reads
taxa_qza <- read_qza("/Users/aikoabodominguez/Documents/eDNA_spiny_lobster/taxonomy_fwd.qza")
taxa_table <- taxa_qza$data %>% select(-Consensus)
taxa_table <- taxa_table %>%
  as_tibble() %>%
  separate(Taxon, sep=";", c("Superkingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species")) %>%
  arrange(Feature.ID) %>%
  mutate(ASV = paste('ASV',1:n(), sep="_")) %>%
```

```
  column_to_rownames("Feature.ID") %>%
  as.matrix()
```

```
## Warning: Expected 7 pieces. Missing pieces filled with `NA` in 9516 rows [1, 2, 3, 4, 5,
## 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, ...].
```

```
#View(taxa_table)

#Importing phylogenetic tree
tree_file_qza <-read_qza("/Users/aikoabodominguez/Documents/eDNA_spiny_lobster/rooted_masked_aligned_seq_ASVs_
tree_file<- tree_file_qza$data
#View(tree_file)
```

Phyloseq generation

Generating phyloseq

```
#generating object
ps <- phyloseq(otu_table(t(ASV_table), taxa_are_rows=FALSE),
               sample_data(metadata),
               tax_table(taxa_table),
               tree_file)
ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 8879 taxa and 36 samples ]
## sample_data() Sample Data:       [ 36 samples by 21 sample variables ]
## tax_table()   Taxonomy Table:    [ 8879 taxa by 8 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 8879 tips and 8705 internal nodes ]
```

Identifying for NAs in ps filtered by taxonomic rank and turned into a df

```
ps_filter = filter_taxa(ps, function(x) sum(x > 5) > (0.05*length(x)), TRUE)
ps_filter
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 2004 taxa and 36 samples ]
## sample_data() Sample Data:       [ 36 samples by 21 sample variables ]
## tax_table()   Taxonomy Table:    [ 2004 taxa by 8 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 2004 tips and 1993 internal nodes ]
```

```
ps_na = transform_sample_counts(ps_filter, function(x) x / sum(x)) %>%
  psmelt()

order_NA <- subset(ps_na, Order=="NA")
family_NA <- subset(ps_na, Family=="NA")
genus_NA <- subset(ps_na, Genus=="NA")
dim(order_NA)
```

```
## [1] 612  32
```

```
dim(family_NA)
```

```
## [1] 144  32
```

```
dim(genus_NA)
```

```
## [1] 396  32
```

```
#View(family_NA)

#ASVs that are NA/missing in order, but genus is labelled
na_ASVs <- setdiff(order_NA$OTU,genus_NA$OTU)
na_ASVs
```

```
##  [1] "95fa735e765c9f9bdb51c735dafb315f" "fc0975f0b26f430e8b6ab0b1c911ea0d"
##  [3] "bf8854d330eb0f55b1bc3fa4d4c52d97" "e3bf28554523da67b9203b94c4bd110b"
##  [5] "a9d4ae9f0286fe9bdb3881218ea6b4c1" "9abb66b0fa12366397e7b4e588d49d01"
##  [7] "8ab5f93e9ac490a760599fb46ab9f46a" "dd7c6dc42a585ae53a2cd701fb221627"
##  [9] "24ee461d377753b6015486f475ed3f59" "42752ae4baedfc05afd2e08db7dc9fcc"
## [11] "7f13ff9b780c6fe5f0a0b3e2f9249cbf" "61a8799f6e1942194a55c98c850de8dc"
## [13] "c1f814fb5cc3e9071bd97660d4a4dfa4" "bc9f5770c5a67b71d6499f4136875c69"
## [15] "725b8fc8ad20a95abd49b8b0fd060d66"
```

```
#this ASV does not have Order, but does have Genus listed
sum(order_NA$OTU=="a9d4ae9f0286fe9bdb3881218ea6b4c1")
```

```
## [1] 36
```

Edit "NAs" from taxa table

```
#adding back Order to taxa missing it, had to look up by hand
taxa_table["a9d4ae9f0286fe9bdb3881218ea6b4c1", "Order"] <- "Neotaenioglossa"
taxa_table["8ab5f93e9ac490a760599fb46ab9f46a", "Order"] <- "Hubrechtiiformes"
```

Relative abundance

##Filter ps

Filter all ps taxa for abundance

```
#filtering taxa. Keeps ASVs where its abundance is greater than 10 in more than 10% of samples. Adjust to see
ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 8879 taxa and 36 samples ]
## sample_data() Sample Data:       [ 36 samples by 21 sample variables ]
## tax_table()   Taxonomy Table:    [ 8879 taxa by 8 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 8879 tips and 8705 internal nodes ]
```

```
#SELECT FILTER
ps_filter = filter_taxa(ps, function(x) sum(x > 10) > (0.10*length(x)), TRUE) %>% tax_fix()
ps_filter
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 944 taxa and 36 samples ]
## sample_data() Sample Data:       [ 36 samples by 21 sample variables ]
## tax_table()   Taxonomy Table:    [ 944 taxa by 8 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 944 tips and 937 internal nodes ]
```

```
ps_filter = filter_taxa(ps, function(x) sum(x > 5) > (0.05*length(x)), TRUE) %>% tax_fix()
ps_filter
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 2004 taxa and 36 samples ]
## sample_data() Sample Data:       [ 36 samples by 21 sample variables ]
## tax_table()   Taxonomy Table:    [ 2004 taxa by 8 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 2004 tips and 1993 internal nodes ]
```

##Calculate relative abundance before plotting

Relative abundance computed using custom function, sample-wise (compositional data) for filtered ps (phyloseq object)

```
#convert ps absolute ASV count into relative abundances by classification level
ps_ra = transform_sample_counts(ps_filter, function(x) x / sum(x)) %>%
tax_glom("Phylum") %>% #select taxonomic rank
  psmelt()

#View(ps_ra)
dim(ps_ra)
```

```
## [1] 1008   26
```

Plotting

Relative abundance of all samples, by site id

Set colors for graph

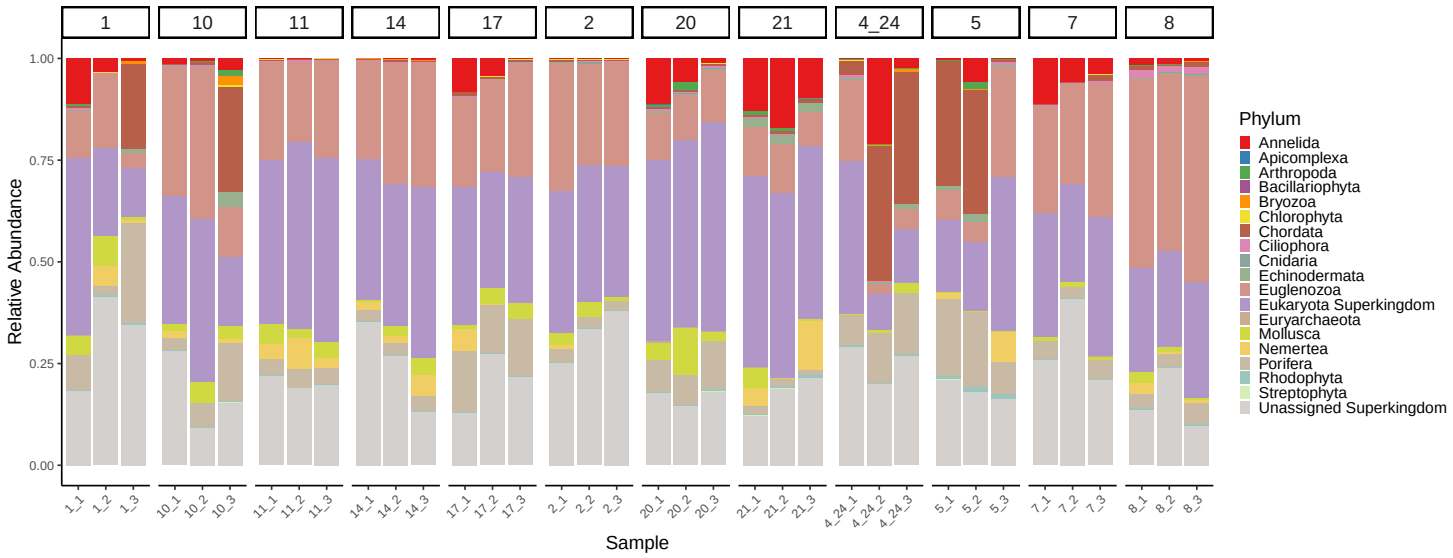
```
#choose which ps to visualize
plot_ps <- ps_ra


#calculate number of colors needed in your color palette, based on count for that taxonomic rank, eg # of uniq
nb.cols <- length(unique(plot_ps$Phylum)) #set taxonomic rank
nb.cols
```

```
## [1] 28
```

```
#set color palette
mycolors <- colorRampPalette(c(brewer.pal(9, "Set1"),
                               brewer.pal(8, "Set2"),
                               brewer.pal(12, "Set3")))(nb.cols)
```

Making relative abundance graph of orders present

```
RA_sup = plot_ps %>%
  subset(Abundance >0.001) %>% #filter for abundance

  #create plot
  ggplot(
         aes(x = sample_id_location, y=Abundance,
             fill=Phylum)) + #fill is  taxonomic rank

  geom_bar(stat="identity", position="fill") +
  scale_fill_manual(values = mycolors) +
  guides(fill = guide_legend(keywidth = 0.5, , keyheight =.60, ncol=1)) +
  theme_classic() + #theme
  facet_grid(.~site_id, scales="free") +

  #text sizes
  theme(legend.text =element_text(size=10)) +  #size of legend items
      theme(legend.title =element_text(size=12)) + #size of legend title
      theme(axis.title.x = element_text(size = 12)) + #size of x axis title
      theme(axis.title.y = element_text(size = 12)) +
      theme(axis.text.y = element_text(size =8)) +
      theme(axis.text.x = element_text(size = 8, angle = 45, hjust = 1)) +
      theme(strip.text = element_text(size = 12)) +
  ylab("Relative Abundance") +
  xlab("Sample")

RA_sup
```




```
# ggsave('Phylum_ra_by_siteID_lobster_eDNA.png', plot=RA_sup,
#        width = 14,
#        height = 7,
#        path = "/Users/aiko.abo.dominguez/Desktop/lobster_eDNA_project/R_code/figures/relative_abundance")
```

###Relative abundance of all samples, by Latitude

```
#choose which ps to visualize
plot_ps <- ps_ra


#calculate number of colors needed in your color palette, based on count for that taxonomic rank, eg # of uniq
nb.cols <- length(unique(plot_ps$Phylum)) #set taxonomic rank
nb.cols
```

```
## [1] 28
```

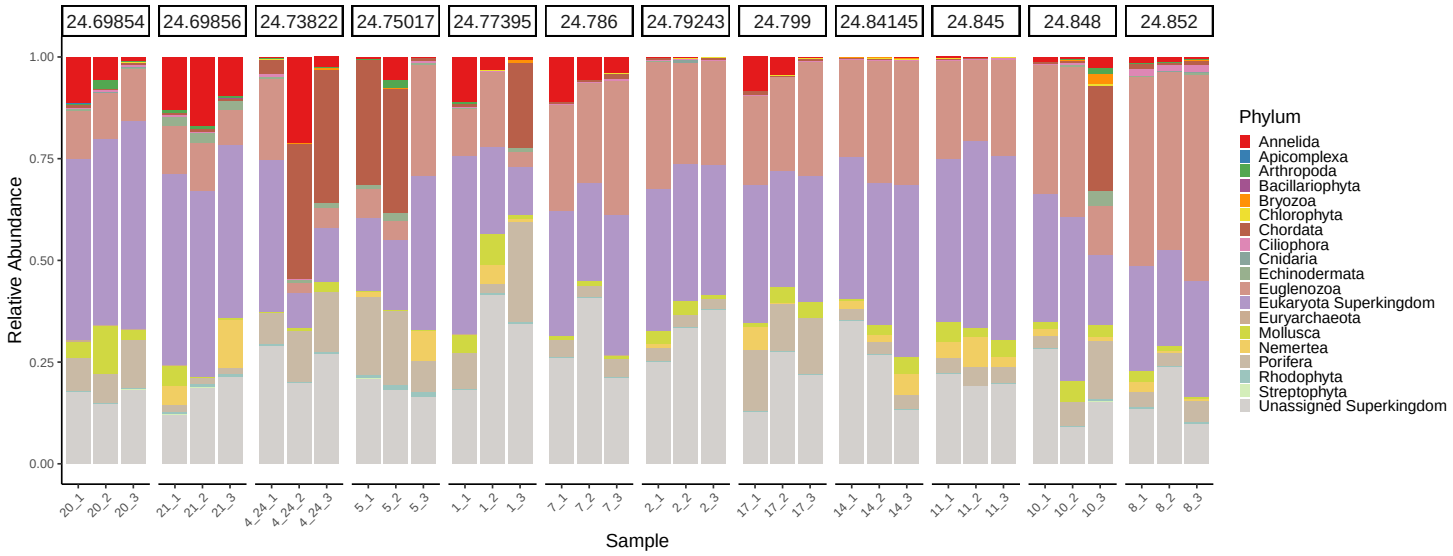
```
#set color palette
mycolors <- colorRampPalette(c(brewer.pal(9, "Set1"),
                               brewer.pal(8, "Set2"),
                               brewer.pal(12, "Set3")))(nb.cols)
```

Making relative abundance graph of taxa present

```
RA_sup = plot_ps %>%
  subset(Abundance >0.001) %>% #filter for abundance

  #create plot
  ggplot(
         aes(x = sample_id_location, y=Abundance,
             fill=Phylum)) + #fill is  taxonomic rank

  geom_bar(stat="identity", position="fill") +
  scale_fill_manual(values = mycolors) +
  guides(fill = guide_legend(keywidth = 0.5, , keyheight =.60, ncol=1)) +
  theme_classic() + #theme
  facet_grid(.~Lat, scales="free") + #separate by lattidue ("Lat")

  #text sizes
  theme(legend.text =element_text(size=10)) +  #size of legend items
      theme(legend.title =element_text(size=12)) + #size of legend title
      theme(axis.title.x = element_text(size = 12)) + #size of x axis title
      theme(axis.title.y = element_text(size = 12)) +
      theme(axis.text.y = element_text(size =8)) +
```

```
      theme(axis.text.x = element_text(size = 8, angle = 45, hjust = 1)) +
      theme(strip.text = element_text(size = 12)) +



  ylab("Relative Abundance") +
  xlab("Sample")
```

```
RA_sup
```




```
# ggsave('Phylum_ra_by_Lattitude_lobster_eDNA.png', plot=RA_sup,
#        width = 14,
#        height = 7,
#        path = "/Users/aiko.abo.dominguez/Desktop/lobster_eDNA_project/R_code/figures/relative_abundance")
```

Relative abundance of all samples, by site type

Set colors for graph

```
#choose which ps to visualize
plot_ps <- ps_ra



#calculate number of colors needed in your color palette, based on count for that taxonomic rank, eg # of uniq
nb.cols <- length(unique(plot_ps$Phylum)) #set taxonomic rank
nb.cols
```

```
## [1] 28
```

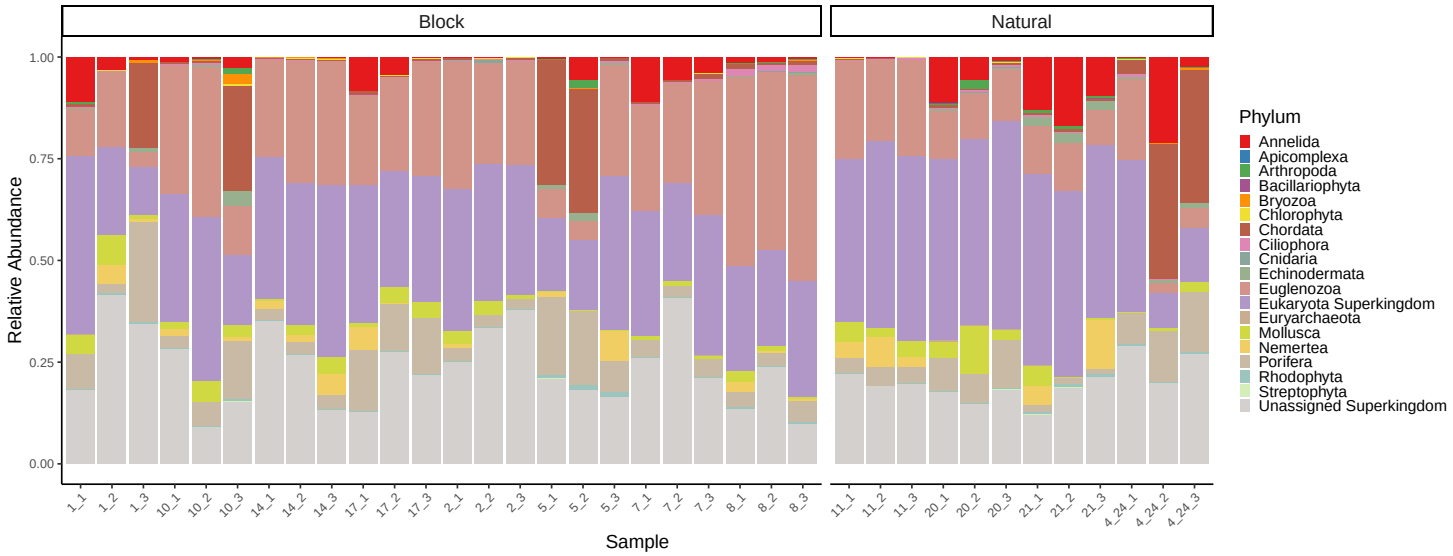
```
#set color palette
mycolors <- colorRampPalette(c(brewer.pal(9, "Set1"),
                               brewer.pal(8, "Set2"),
                               brewer.pal(12, "Set3")))
```

Making relative abundance graph of taxa present

```
    RA_sup = plot_ps %>%

      subset(Abundance >0.001) %>% #filter for abundance
      ggplot(
             aes(x = sample_id_location, y=Abundance,
                 fill=Phylum)) + #fill is  taxonomic rank

      geom_bar(stat="identity", position="fill") +
      scale_fill_manual(values = mycolors(nb.cols)) +
```

```
      guides(fill = guide_legend(keywidth = 0.5, , keyheight =.60, ncol=1)) +
      theme_classic() + #theme
      facet_grid(. ~ site_type, scales="free_x",space="free_x") +

      #text sizes
      theme(legend.text =element_text(size=10)) +  #size of legend items
      theme(legend.title =element_text(size=12)) + #size of legend title
      theme(axis.title.x = element_text(size = 12)) + #size of x axis title
      theme(axis.title.y = element_text(size = 12)) +
      theme(axis.text.y = element_text(size =8)) +
      theme(axis.text.x = element_text(size = 8, angle = 45, hjust = 1)) +
      theme(strip.text = element_text(size = 12)) +

      ylab("Relative Abundance") +
      xlab("Sample")
```

RA_sup




```
# ggsave('Family_ra_by_site_PAV1_eDNA.png', plot=RA_sup,
#        width = 14,
#        height = 7,
#        path = "/Users/aiko.abo.dominguez/Desktop/lobster_eDNA_project/R_code/figures/relative_abundance")
```