plotting_taxa_abundance

Visualizing results from DESeq, MetagenomeSeq, and Ancom

Read in results from above tests

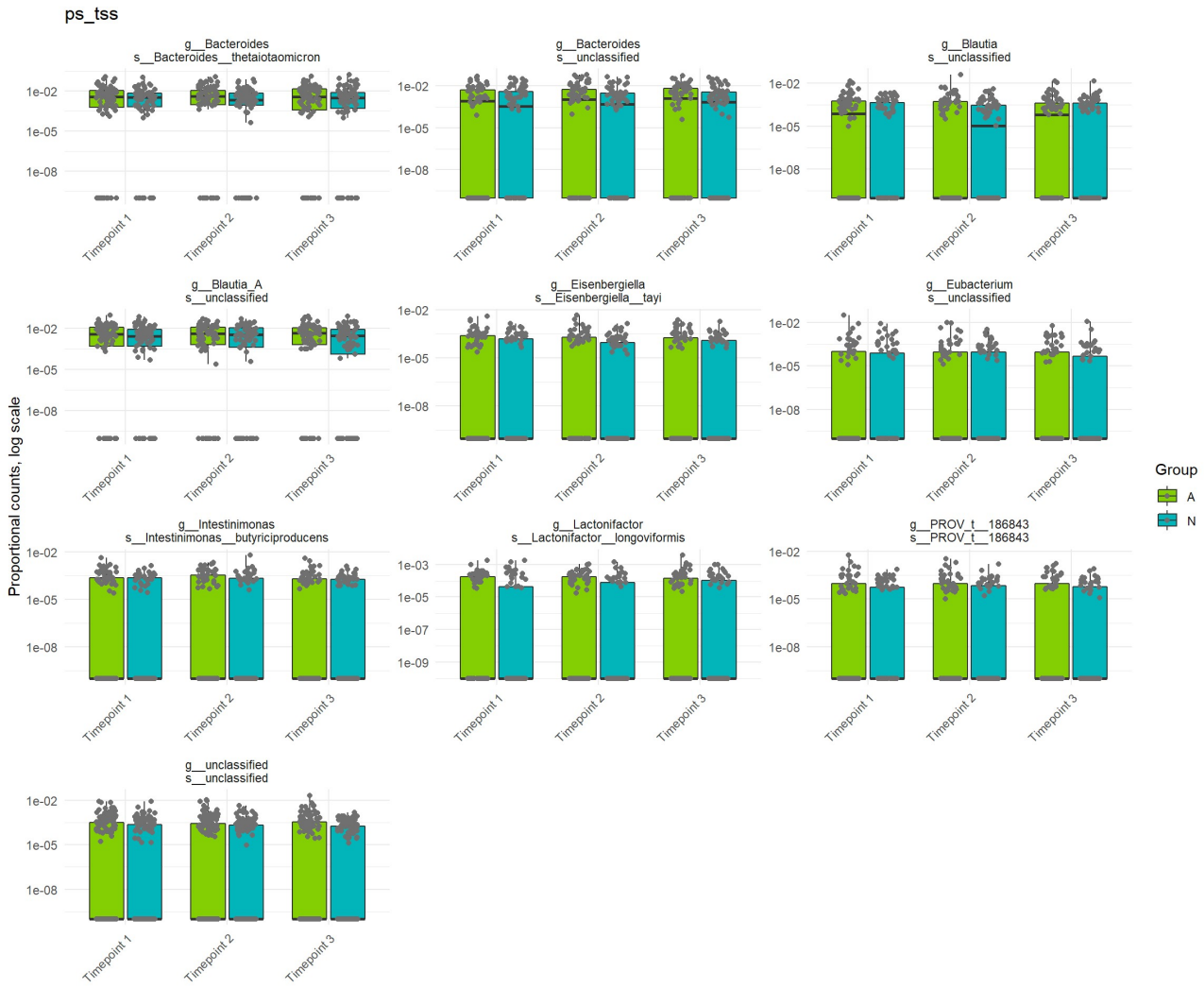
```
output_data <- "../results/"
sig_taxa <- read.table(paste0(output_data, "significant_taxa_timeseries.tsv"), sep =
"\t", header = T)

#Load for plotting purposes:
ps_tss <- readRDS(paste0(output_data, "Normalized/ps_tss_pass_min_postDD_min0.03.rd
s"))
```

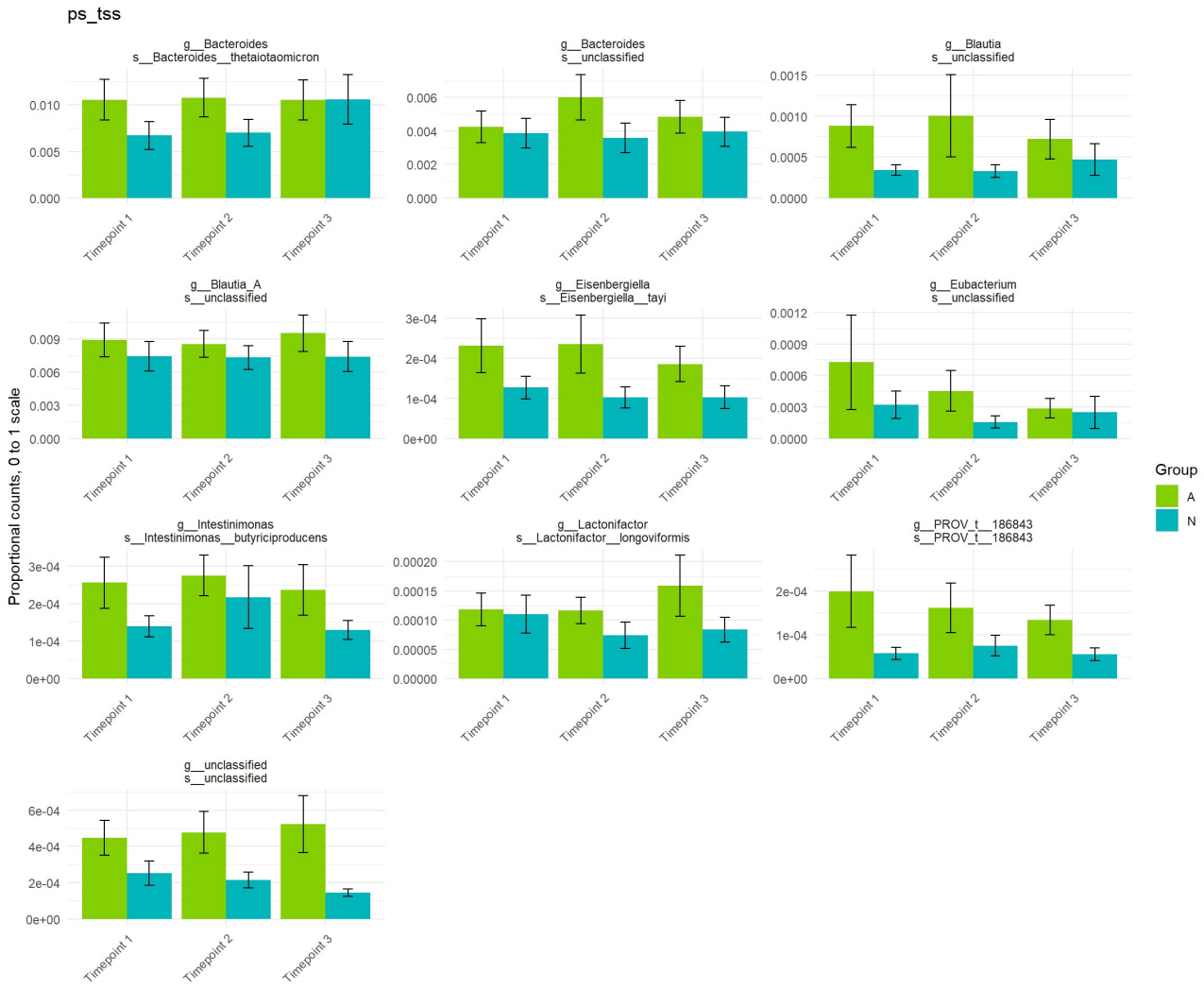
```
### functions to plot
make_vis_plots <- function(ps_norm, grouping, tax, plot_type=c('box', 'bar')){
  # ps should be a normalized (DESeq or CSS) phyloseq object
  # grouping should match the column name in the sample_data
  # tax is a taxonomical bin id (ASV) in the counts table to plot

  # subset phyloseq object to select ASV of interest
  ps_filt=prune_taxa(taxa_names(ps_norm) %in% tax, ps_norm)

  # get normalized counts
  plot_table<-data.table(otu_table(ps_filt), keep.rownames=TRUE)[rn %in% tax]
  # add very small value, min/100000 to 0
  plot_table <- melt(plot_table, id.vars='rn')
  plot_table$value <- plot_table$value+min(plot_table[value!=0]$value)/100000

  # add metadata
  groupDT=data.table(data.frame(sample_data(ps_filt)[, c(grouping, 'Within.study.sampl
ing.date')]), keep.rownames=TRUE)
  setnames(groupDT, 'rn', 'variable')
  plot_table <- merge(plot_table, groupDT, by='variable', all.x=TRUE)

  #taxa names for plot labels
  species_names <- as.character(ps_tss@tax_table[plot_table$rn, 7])
  genus_names <- as.character(ps_tss@tax_table[plot_table$rn, 6])
  taxa_names <- paste( genus_names,species_names, sep = "\n")
  unclassified <- species_names == "s__unclassified"
  species_names[unclassified] <- genus_names[unclassified]
  plot_table$rn <- taxa_names

  # change variable to general name
  setnames(plot_table, grouping, 'Group')

  # boxplot
  if(plot_type=='box'){
    ggplot(data=plot_table, aes(x=Within.study.sampling.date, y = value, fill=Group))
+
      geom_boxplot(outlier.color=NA) +
      geom_jitter(position=position_jitterdodge(0.2), cex=1.5, color="gray44") +
      labs(title =deparse(substitute(ps_norm)), x='', y ='Proportional counts, log sca
le') +
      scale_y_log10() +
      scale_fill_manual(values=sgColorPalette)+
      theme_minimal() + facet_wrap(~rn, scales='free', ncol=3)+
      theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
  } else if (plot_type=='bar'){
    plot_table2 <- plot_table[, list(mean_ct=mean(value), sem=sd(value)/sqrt(.N)), by=
c('Group', 'Within.study.sampling.date', 'rn')]
    ggplot(data=plot_table2, aes(x=Within.study.sampling.date, y =mean_ct, fill=Grou
```

```
p)) +
      geom_bar(stat='identity', position=position_dodge()) +
      geom_errorbar(aes(ymin=mean_ct-sem, ymax=mean_ct+sem), width=0.2, position=posit
ion_dodge(0.9))+
      labs(title =deparse(substitute(ps_norm)), x='', y ='Proportional counts, 0 to 1
scale') +
      scale_fill_manual(values=sgColorPalette)+
      theme_minimal() + facet_wrap(~rn, scales='free', ncol=3)+
      theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
  }
}
```

Both DESeq and metagenomeseq timeseries

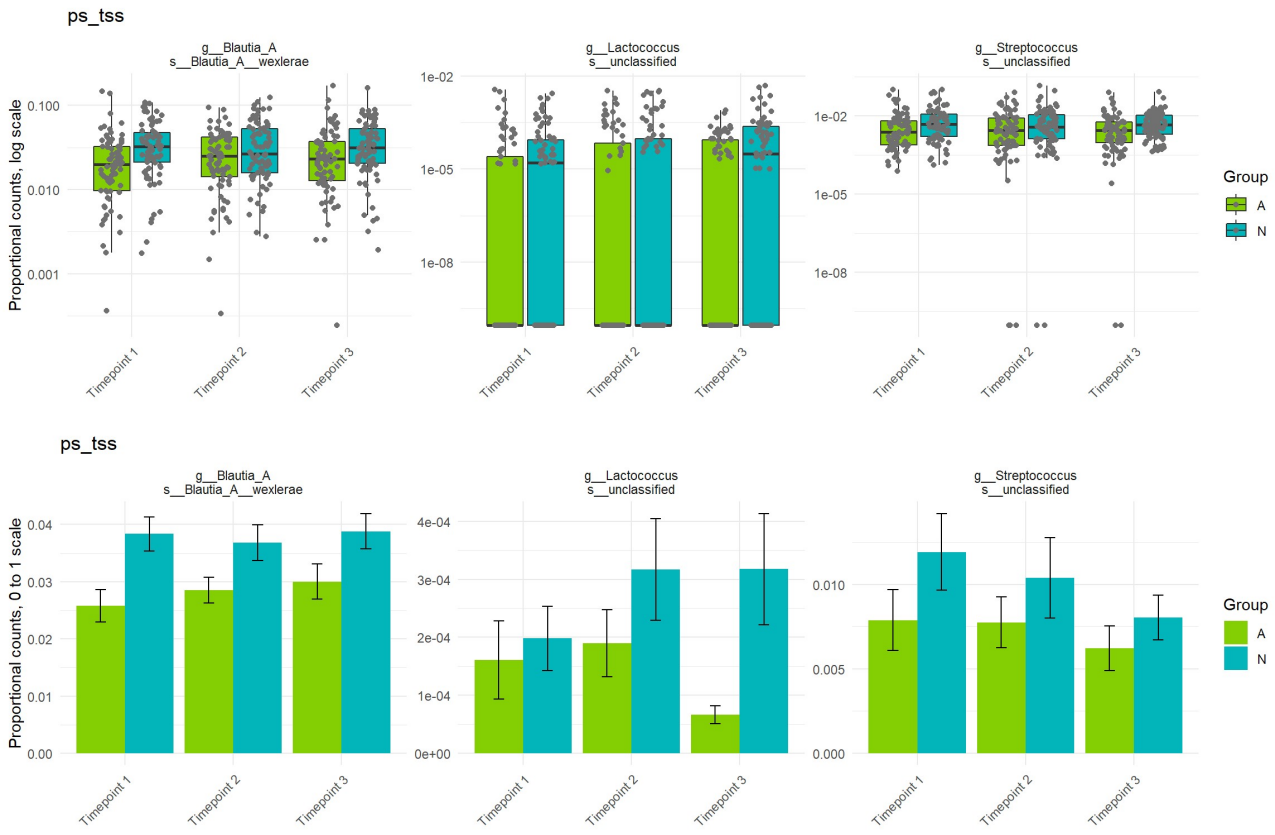
```
keep <- grepl("DESEQ", sig_taxa$method) & grepl("ZIG", sig_taxa$method)
if(sum(keep)>0){
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'box'))
  # plot bar as well
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'bar'))
} else {
  print('no combined timeseries significant taxa')
}
```


ps_tss
g__Bacteroides
s__Bacteroides__thetaiotaomicron
g__Bacteroides
s__unclassified
g__Blautia
s__unclassified
g__Blautia_A
s__unclassified
g__Eisenbergiella
s__Eisenbergiella__tayi
g__Eubacterium
s__unclassified
g__Intestinimonas
s__Intestinimonas__butyriciproducens
g__Lactonifactor
s__Lactonifactor__longoviformis
g__PROV_t__186843
s__PROV_t__186843
g__unclassified
s__unclassified
Proportional counts, log scale
Timepoint 1
Timepoint 2
Timepoint 3
Group
A
N



Both ZIG and ANCOM timeseries

```
keep <- grepl("ZIG", sig_taxa$method) & grepl("ANCOM", sig_taxa$method)
if(sum(keep)>0){
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'box'))
  # plot bar as well
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'bar'))
} else {
  print('no combined timeseries significant taxa')
}
```



Both DESeq and ANCOM timeseries

```
keep <- grepl("DESeq2", sig_taxa$method) & grepl("ANCOM", sig_taxa$method)
if(sum(keep)>0){
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'box'))
  # plot bar as well
  print(make_vis_plots(ps_tss, 'phenotype', sig_taxa[keep, ]$asv, 'bar'))
} else {
  print('no combined timeseries significant taxa')
}
```

```
## [1] "no combined timeseries significant taxa"
```

At least 1 timepoint in both DESeq and metagenomeSeq

DESeq all 3 timepoints

metagenomeSeq all 3 timepoints

DESeq at least 2 timepoints

Meta_genome at least 2 timepoints