

Palmer_Station_Bacterial_Succesion_Analysis

Cat Luria

July 6, 2016

Contents

1	Preliminary analysis in Qiime	2
2	Getting started in R	2
3	Import main OTU table and ancillary data	2
4	Figure 1. Chlorophyll a, bacterial production, and OTU richness	2
5	Figure 2. Aggregated taxa plot for paper	5
6	Figure 3. OTU succession plot	5
7	Figure 4. NMDS	5
8	Supplemental Figure 1. Ancillary data	5
9	Supplemental Figure 2. Richness versus read number	10
10	Supplemental Figure 3 TMM and RLE normalizations of unrarefied data	10
10.1	Import data for edgeR	10
10.2	RLE normalization	10
10.3	TMM normalization	10
10.4	NMDS	10
10.5	Six panel ordination	12

1 Preliminary analysis in Qiime

OTU tables were generated in Qiime, using uclust open reference OTU picking, minimum cluster size of 2. Chloroplasts were removed and libraries were rarefied to smallest size.

The final dataset consists of 68 16S rRNA gene sequence libraries generated from samples collected from the seawater intake at Palmer Station, Antarctica from July 2013 through March 2014. There are triplicate libraries from all but 3 sampling dates. The following SWI samples failed during sequencing and were dropped from subsequent analyses:

```
LAZPALBv6.PAL05520130824F02: 851.0
LAZPALBv6.PAL01420130704F02: 967.0
LAZPALBv6.PAL02020130711F02: 968.0
LAZPALBv6.PAL01820130711F02: 1001.0
LAZPALBv6.PAL01920130711F02: 1058.0
```

The remaining libraries are summarized below:

```
Num samples: 68
Num observations: 28857
Total count: 15535915
Table density (fraction of non-zero values): 0.101
```

Counts/sample summary:

```
Min: 51210.0
Max: 548650.0
Median: 197541.000
Mean: 228469.338
Std. dev.: 98371.744
```

2 Getting started in R

3 Import main OTU table and ancillary data

4 Figure 1. Chlorophyll a, bacterial production, and OTU richness

```
## Warning: Removed 1 rows containing missing values (geom_errorbar).
```

```
## Warning: Removed 1 rows containing missing values (geom_errorbar).
```

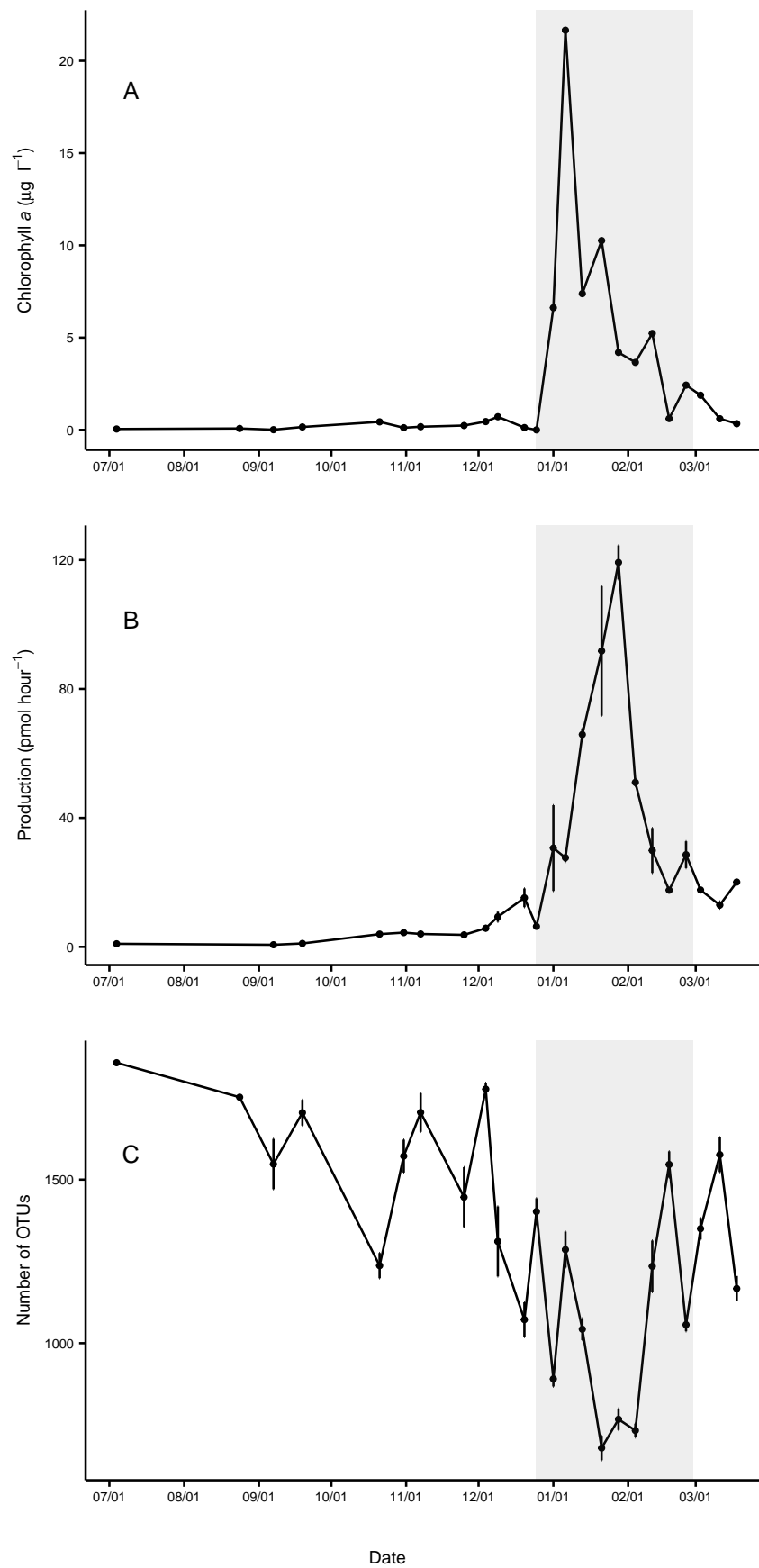


Figure 1:
3

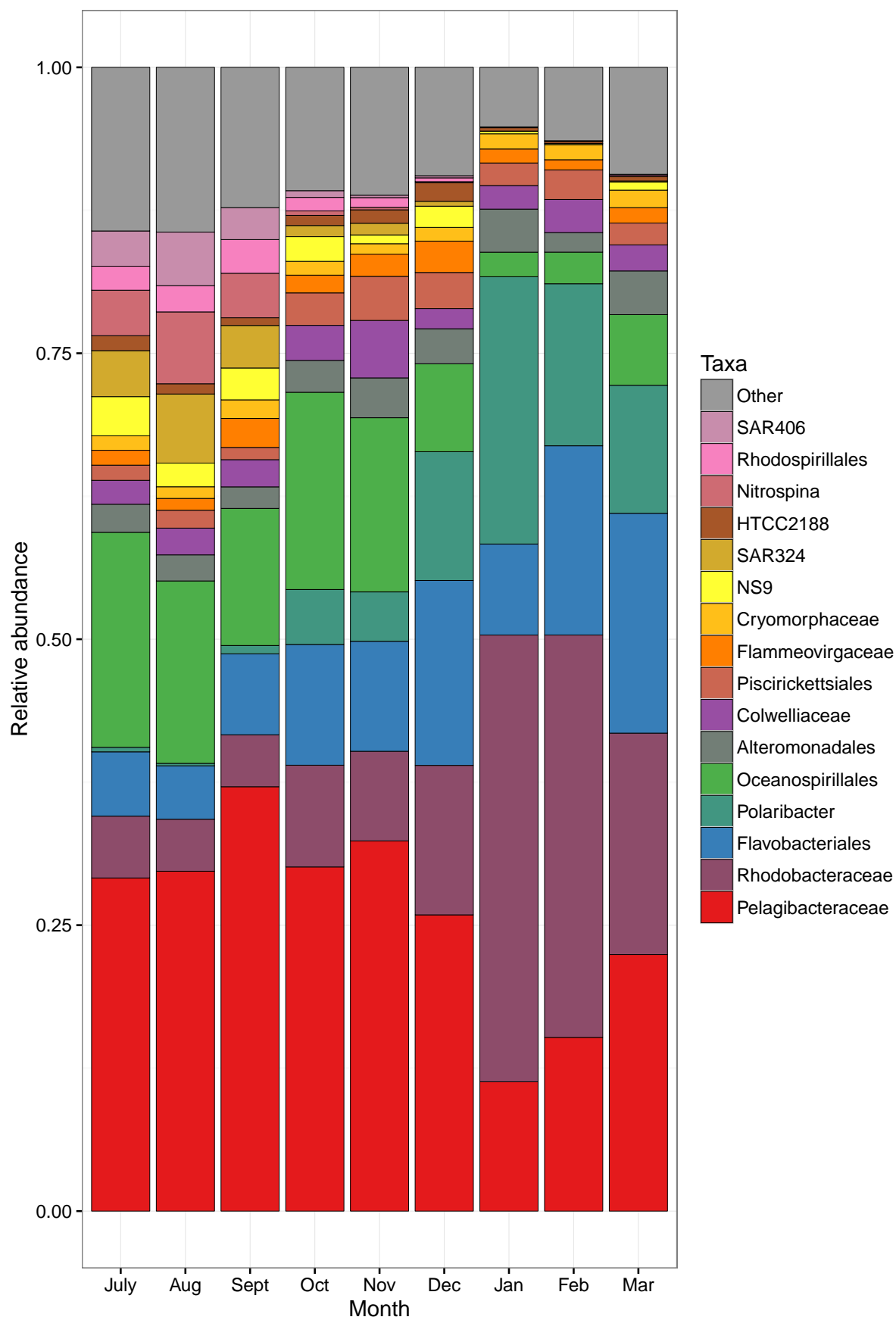


Figure 2: Relative abundance of select taxa, July 2013-March 2014.

5 Figure 2. Aggregated taxa plot for paper

6 Figure 3. OTU succession plot

```
## [1] 68 48
```

```
## [1] 68 48
```

```
## pdf
```

```
## 2
```

Out of 12K OTUs, only about 5K make up at least 1% of the community in at least three samples. Out of the 12K OTUs, about 8K had a very low coefficient of variation. 4061 OTUs meet both criteria. Of these, only 48 had an R-squared >0.8.

```
## Warning: Removed 3 rows containing missing values (geom_errorbar).
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbar).
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbar).
```

```
## Warning: Removed 3 rows containing missing values (geom_errorbar).
```

```
## Warning: Removed 2 rows containing missing values (geom_errorbar).
```

7 Figure 4. NMDS

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.0952765
## Run 1 stress 0.0952754
## ... New best solution
## ... procrustes: rmse 0.0002214993 max resid 0.0006486576
## *** Solution reached
```

8 Supplemental Figure 1. Ancillary data

```
panel.pho <- ggplot_gtable(ggplot_build(g.pho))
panel.sil <- ggplot_gtable(ggplot_build(g.sil))
panel.nit <- ggplot_gtable(ggplot_build(g.nit))
panel.pc <- ggplot_gtable(ggplot_build(g.pc))
```

```
## Warning: Removed 1 rows containing missing values (geom_errorbar).
```

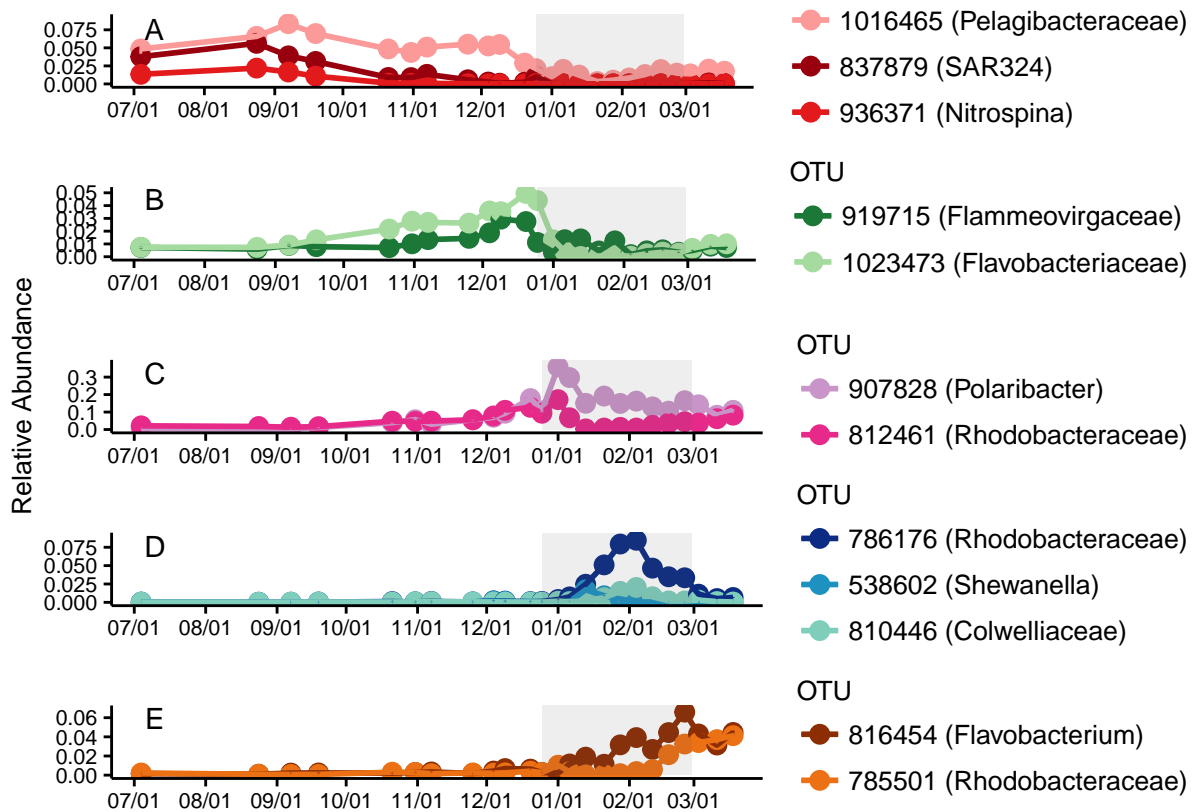


Figure 3:

```
panel.pn <- ggplot_gtable(ggplot_build(g.pn))

## Warning: Removed 1 rows containing missing values (geom_errorbar).

panel.ba <- ggplot_gtable(ggplot_build(g.ba))

maxWidth1 = grid::unit.pmax(panel.pho$widths[2:3], panel.sil$widths[2:3],
                             panel.nit$widths[2:3])

maxWidth2 = grid::unit.pmax (panel.pc$widths[2:3],
                              panel.pn$widths[2:3], panel.ba$widths[2:3])

panel.pho$widths[2:3] <- maxWidth1
panel.sil$widths[2:3] <- maxWidth1
panel.nit$widths[2:3] <- maxWidth1
panel.pc$widths[2:3] <- maxWidth2
panel.pn$widths[2:3] <- maxWidth2
panel.ba$widths[2:3] <- maxWidth2

g <- grid.arrange(panel.pho, panel.sil, panel.nit)

g <- grid.arrange(arrangeGrob(panel.nit, panel.pho, panel.sil),
                  arrangeGrob(panel.pc, panel.pn, panel.ba), ncol = 2,
                  bottom=textGrob("Date", gp=gpar(fontsize=12)))
```

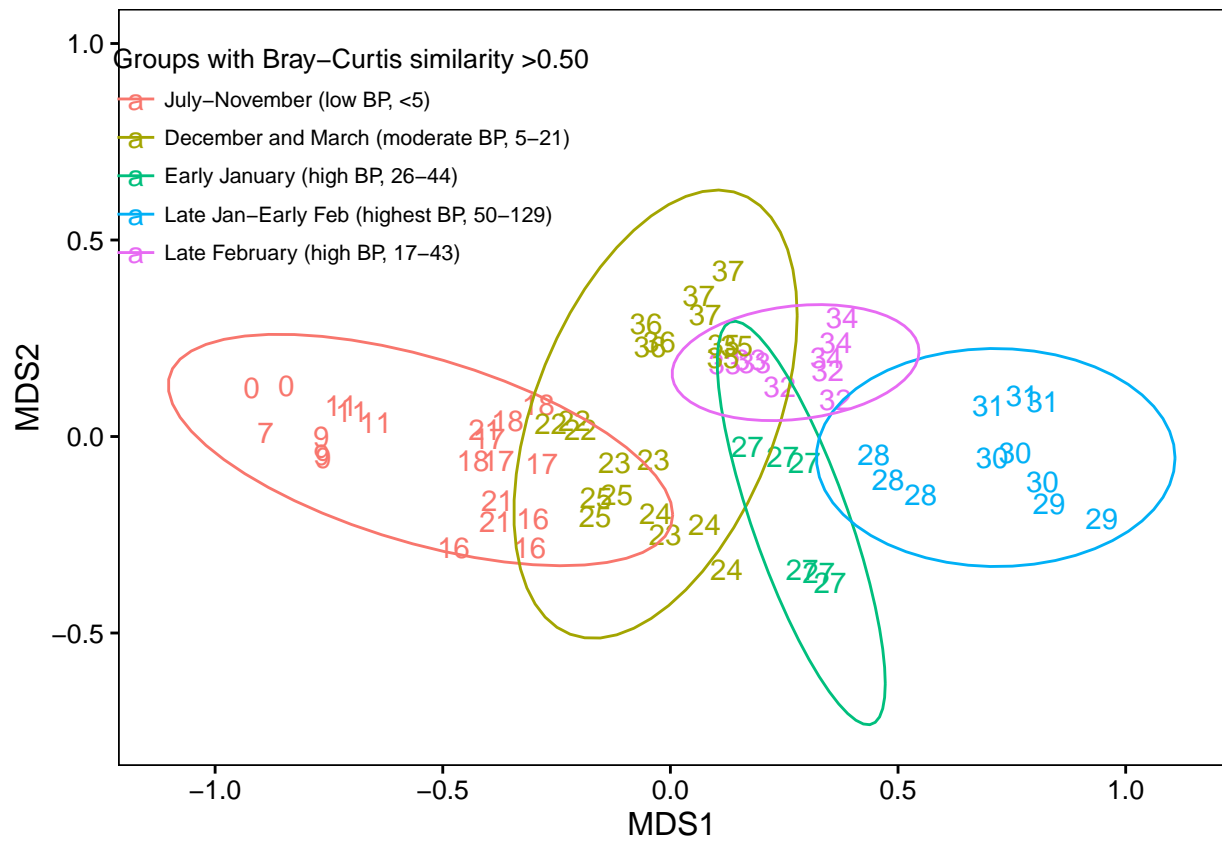


Figure 4: NMDS of SWI samples, rarefied data, July 2013–March 2014.

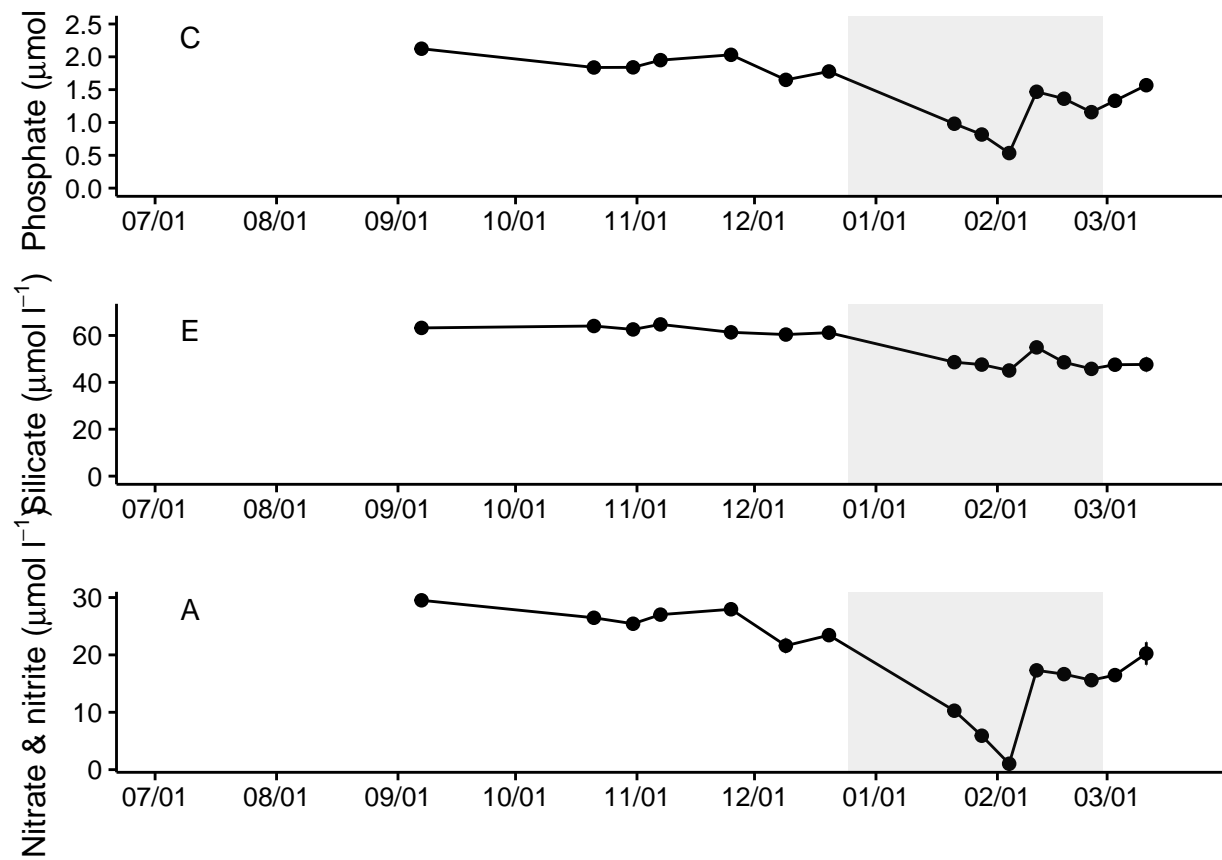


Figure 5:

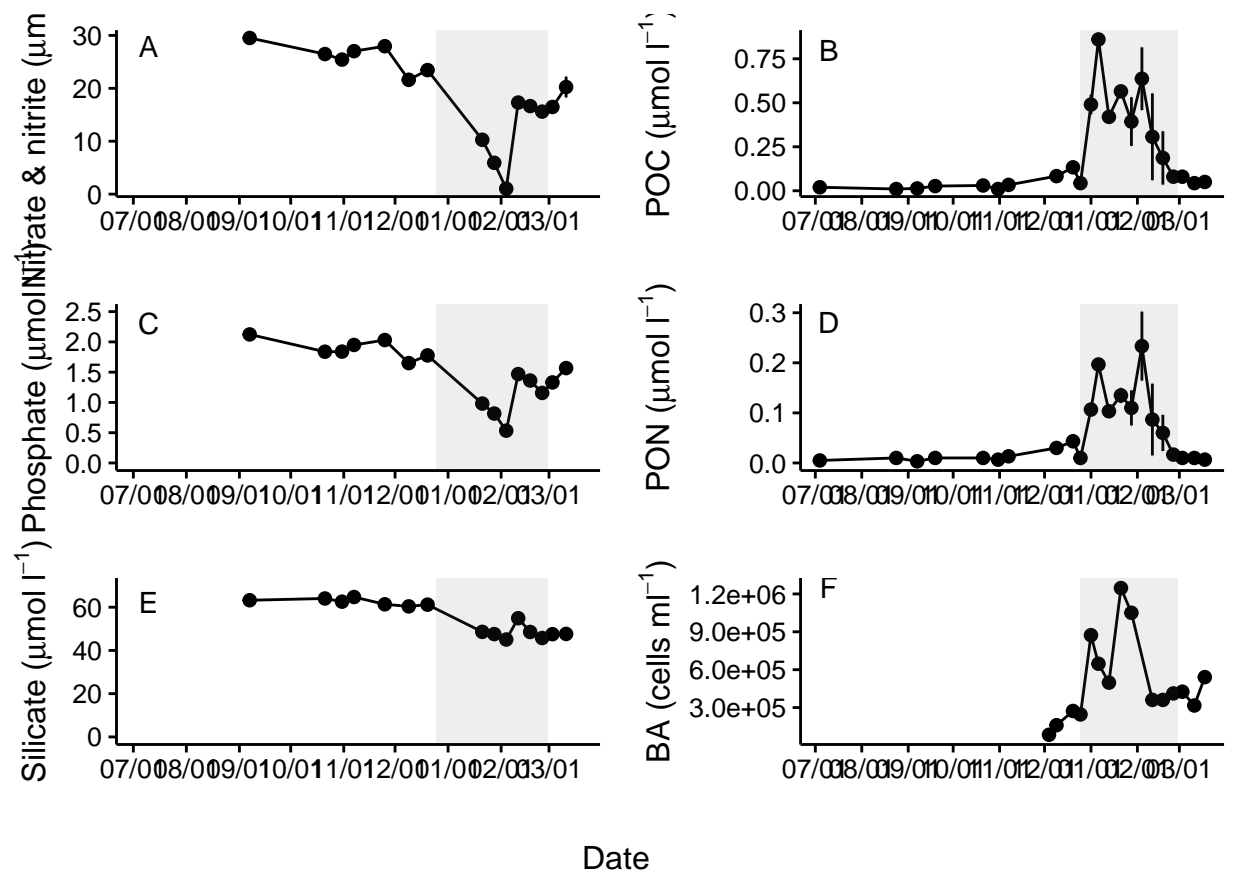


Figure 6:

```
#grid.draw(cbind(ggplotGrob(g.pho), ggplotGrob(g.sil), ggplotGrob(g.nit), size="last"), ncol=2)
```

9 Supplemental Figure 2. Richness versus read number

```
## Using 'Freq' as value column. Use 'value.var' to override
```

10 Supplemental Figure 3 TMM and RLE normalizations of unrarefied data

10.1 Import data for edgeR

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 27198 taxa and 68 samples ]
## sample_data() Sample Data: [ 68 samples by 54 sample variables ]
## tax_table() Taxonomy Table: [ 27198 taxa by 8 taxonomic ranks ]
## phy_tree() Phylogenetic Tree: [ 27198 tips and 4140 internal nodes ]
```

10.2 RLE normalization

10.3 TMM normalization

10.4 NMDS

10.4.1 Relative abundance

```
## Run 0 stress 0.07250697
## Run 1 stress 0.0725081
## ... procrustes: rmse 0.03336008 max resid 0.1452701
## Run 2 stress 0.0682625
## ... New best solution
## ... procrustes: rmse 0.02621909 max resid 0.1434732
## Run 3 stress 0.08848803
## Run 4 stress 0.07323105
## Run 5 stress 0.09006151
## Run 6 stress 0.07309428
## Run 7 stress 0.08625416
## Run 8 stress 0.08978789
## Run 9 stress 0.09332683
## Run 10 stress 0.07250703
## Run 11 stress 0.09297142
## Run 12 stress 0.07093793
## Run 13 stress 0.08563122
## Run 14 stress 0.07250839
## Run 15 stress 0.07281226
## Run 16 stress 0.09455371
## Run 17 stress 0.06952867
## Run 18 stress 0.08635099
## Run 19 stress 0.08765036
## Run 20 stress 0.09273026
```

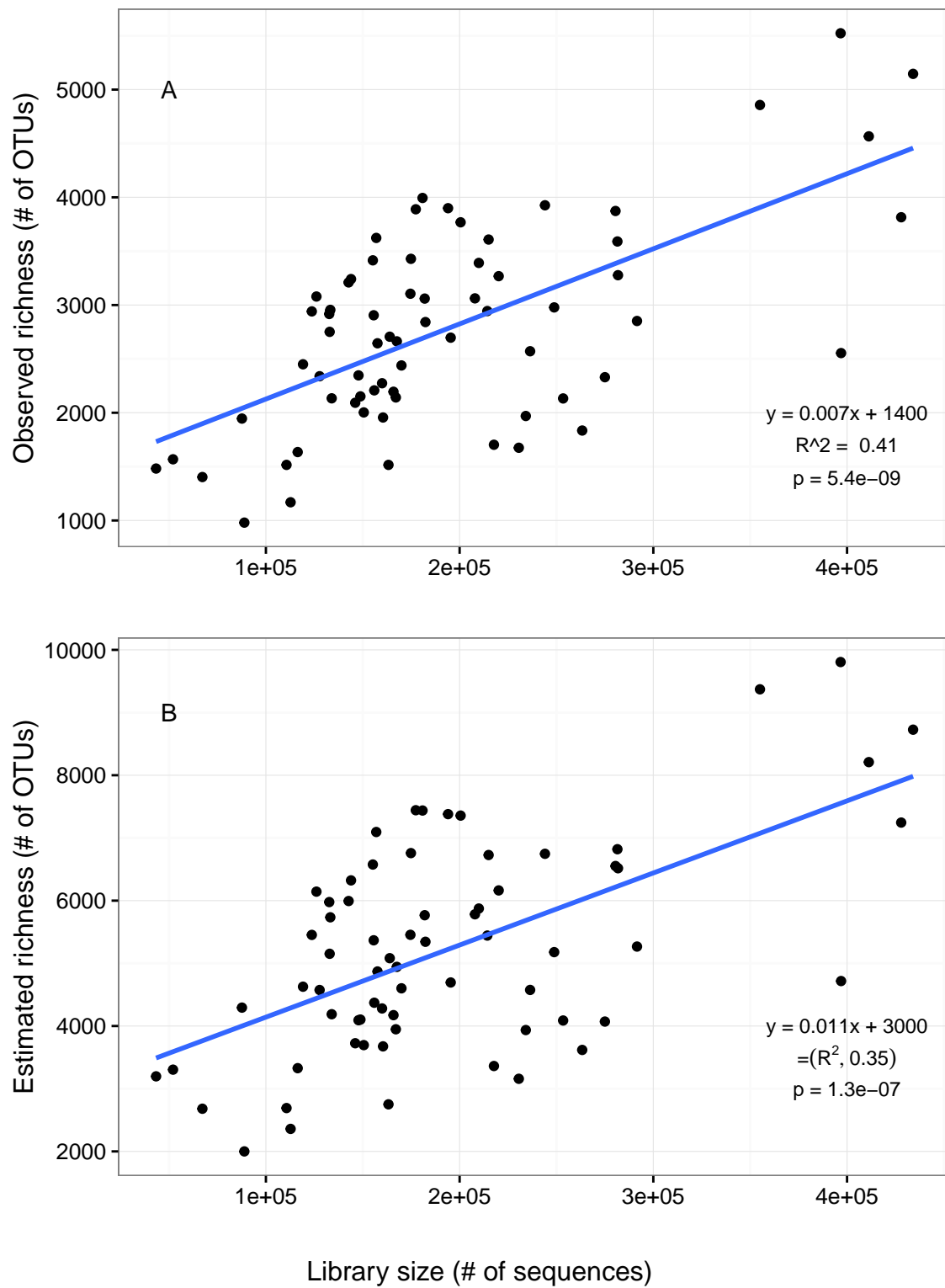


Figure 7: Both observed and estimated richness are significantly correlated to sampling depth.

10.4.2 RLE Bray Curtis

```
##
## > my_bray_curtis <- function(df) {
## +   dis <- sparseBC(as.matrix(df))
## +   colnames(dis) <- row.names(df)
## +   row.names(dis) <- row.names(df)
## +   .... [TRUNCATED]
```

10.4.3 TMM Bray Curtis

```
##
## > my_bray_curtis <- function(df) {
## +   dis <- sparseBC(as.matrix(df))
## +   colnames(dis) <- row.names(df)
## +   row.names(dis) <- row.names(df)
## +   .... [TRUNCATED]
```

10.5 Six panel ordination

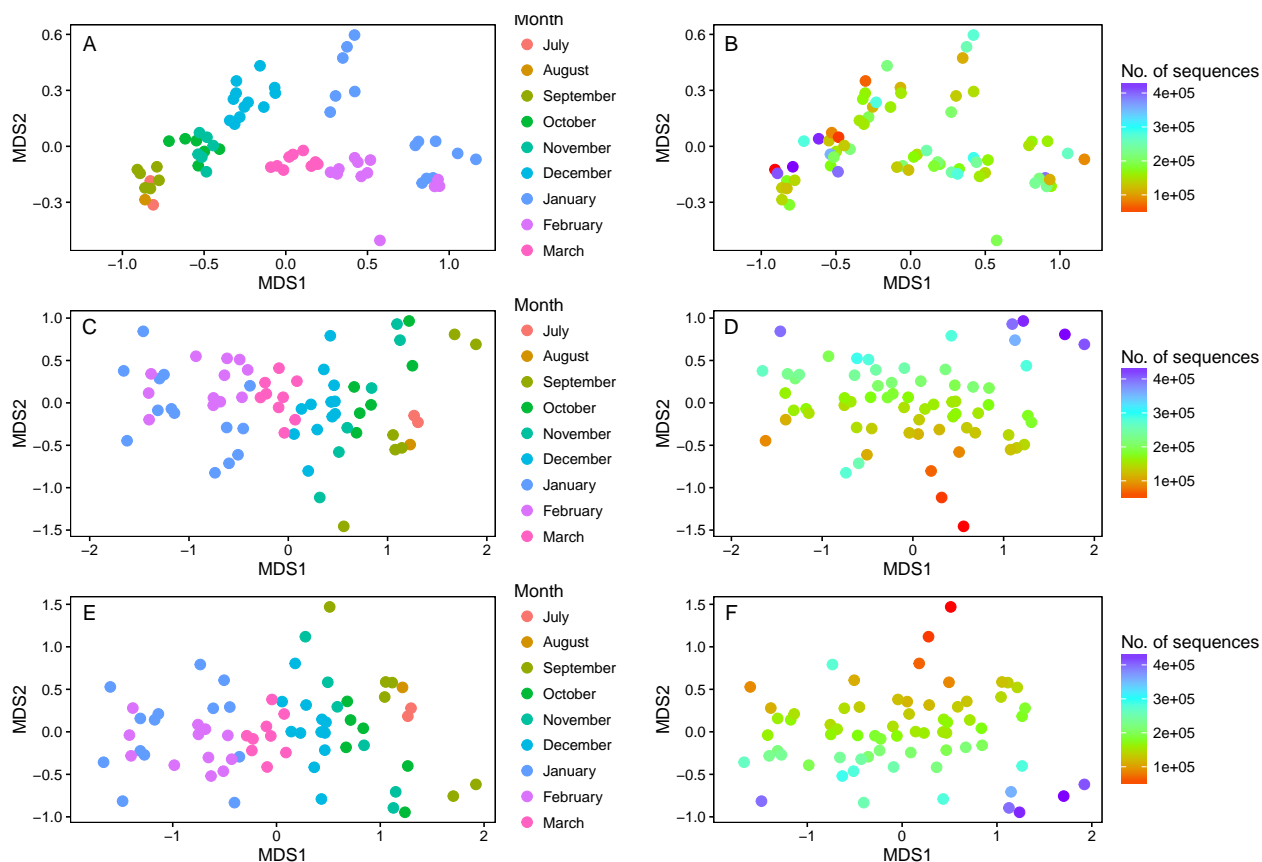


Figure 8: NMDS based on unrarefied libraries, three types of normalization (relative abundance, RLE, and TMM), color-coded according to both month and sampling depth.