# Diversity in Pmt2 and S1

## Haydeé

We analysed the two metagenomes Pmt2 and S1 using phyloseq. First we process the biom file and subset by Bacteria.

There are 40 unique Phylums and 1705 unique Genus across the two metagenomes.

The most abundant Phylums are in the following table.

```
summary(raw_metagenomes@tax_table@.Data== "")
```

```
##   Kingdom            Phylum            Class            Order
##  Mode :logical    Mode :logical    Mode :logical    Mode :logical
##  FALSE:8192        FALSE:8188        FALSE:7941        FALSE:8126
##                    TRUE :4           TRUE :251         TRUE :66
##     Family            Genus           Species
##  Mode :logical    Mode :logical    Mode :logical
##  FALSE:8003        FALSE:7879        FALSE:7221
##  TRUE :189         TRUE :313         TRUE :971
```

```
raw_metagenomes <- subset_taxa(raw_metagenomes, Genus != "")
relative_metagenomes <- transform_sample_counts(raw_metagenomes, function(x) x*100 / sum(x) )
relative_phylums <- tax_glom(relative_metagenomes, taxrank = 'Phylum')
relative_phylums_df <- psmelt(relative_phylums)
relative_phylums_df[order(-relative_phylums_df$Abundance),][1:11,]
```

```
##         OTU    Sample Abundance          Id  Kingdom          Phylum
## 47      286 Pmt2_S287 60.882418 Pmt2_S287 Bacteria Proteobacteria
## 48      286   S1_S286 48.249401   S1_S286 Bacteria Proteobacteria
## 21     1883   S1_S286 26.444820   S1_S286 Bacteria Actinobacteria
## 22     1883 Pmt2_S287 25.413994 Pmt2_S287 Bacteria Actinobacteria
## 52 2908210   S1_S286 14.642142   S1_S286 Bacteria  Bacteroidetes
## 35 2527989   S1_S286  6.877866   S1_S286 Bacteria Planctomycetes
## 36 2527989 Pmt2_S287  3.599600 Pmt2_S287 Bacteria Planctomycetes
## 18 1855912 Pmt2_S287  2.650876 Pmt2_S287 Bacteria  Acidobacteria
## 51 2908210 Pmt2_S287  2.005273 Pmt2_S287 Bacteria  Bacteroidetes
## 32 2233542 Pmt2_S287  1.667221 Pmt2_S287 Bacteria      Firmicutes
## 31 2233542   S1_S286  1.481893   S1_S286 Bacteria      Firmicutes
```

```
relative_phylums_df2 <- relative_phylums_df[-c(1,4,5)]
rel_phyls <- spread(relative_phylums_df2, Sample, Abundance)
```

```
rel_phyls
```

```
##                             Phylum      Pmt2_S287       S1_S286
## 1                     Acidobacteria 2.650876e+00  3.216666e-01
## 2                    Actinobacteria 2.541399e+01  2.644482e+01
## 3                         Aquificae 8.707546e-03  5.939865e-03
## 4                    Armatimonadetes 3.249159e-02  1.208742e-02
## 5                      Atribacterota 7.463611e-04  3.738377e-04
## 6                      Bacteroidetes 2.005273e+00  1.464214e+01
## 7                       Balneolaeota 5.224528e-03  5.233727e-03
## 8                        Caldiserica 1.094663e-03  3.323002e-04
## 9                     Calditrichaeota 7.811913e-03  3.073776e-03
## 10 Candidatus Absconditabacteria 7.961185e-04  2.907626e-04
## 11    Candidatus Bipolaricaulota 7.712398e-03  2.699939e-03
## 12      Candidatus Cloacimonetes 1.393207e-03  1.744576e-03
## 13       Candidatus Omnitrophica 5.921132e-03  1.079976e-03
## 14   Candidatus Saccharibacteria 1.841024e-03  9.719780e-03
## 15                         Chlamydiae 8.906576e-03  8.224429e-03
## 16                            Chlorobi 6.597832e-02  2.820398e-02
## 17                          Chloroflexi 3.312848e-01  1.197111e-01
## 18                      Chrysiogenetes 5.224528e-03  3.572227e-03
## 19              Coprothermobacterota 1.592237e-03  2.076876e-04
## 20                        Cyanobacteria 6.619726e-01  4.798830e-01
## 21                      Deferribacteres 5.821617e-03  4.319902e-03
## 22                Deinococcus-Thermus 4.753325e-01  3.570150e-01
## 23                          Dictyoglomi 1.741509e-03  2.699939e-03
## 24                       Elusimicrobia 3.383504e-03  2.450714e-03
## 25                        Fibrobacteres 3.134717e-03  1.536888e-03
## 26                            Firmicutes 1.667221e+00  1.481893e+00
## 27                       Fusobacteria 6.667493e-03  2.085183e-02
## 28                   Gemmatimonadetes 6.533148e-01  4.586988e-01
## 29                     Ignavibacteriae 6.169919e-03  6.272165e-03
## 30                 Kiritimatiellaeota 2.209229e-02  1.316739e-02
## 31                          Nitrospirae 5.496701e-01  3.053008e-02
## 32                      Planctomycetes 3.599600e+00  6.877866e+00
## 33                       Proteobacteria 6.088242e+01  4.824940e+01
## 34                     Rhodothermaeota 5.876350e-02  6.351087e-02
## 35                        Spirochaetes 8.767255e-02  5.021886e-02
## 36                        Synergistetes 3.552679e-02  1.831805e-02
## 37                          Tenericutes 1.184226e-02  2.446560e-02
## 38             Thermodesulfobacteria 9.155363e-03  3.489152e-03
## 39                         Thermotogae 1.706679e-02  1.113206e-02
## 40                     Verrucomicrobia 6.798852e-01  2.301179e-01
```

```r
rel_phyls[order(-rel_phyls$Pmt2_S287), ]
```

```
##                             Phylum      Pmt2_S287       S1_S286
## 33                   Proteobacteria 6.088242e+01  4.824940e+01
## 2                    Actinobacteria 2.541399e+01  2.644482e+01
## 32                   Planctomycetes 3.599600e+00  6.877866e+00
## 1                     Acidobacteria 2.650876e+00  3.216666e-01
## 6                      Bacteroidetes 2.005273e+00  1.464214e+01
## 26                         Firmicutes 1.667221e+00  1.481893e+00
## 40                   Verrucomicrobia 6.798852e-01  2.301179e-01
## 20                     Cyanobacteria 6.619726e-01  4.798830e-01
## 28                   Gemmatimonadetes 6.533148e-01  4.586988e-01
```

```
## 31                    Nitrospirae 5.496701e-01 3.053008e-02
## 22              Deinococcus-Thermus 4.753325e-01 3.570150e-01
## 17                     Chloroflexi 3.312848e-01 1.197111e-01
## 35                     Spirochaetes 8.767255e-02 5.021886e-02
## 16                         Chlorobi 6.597832e-02 2.820398e-02
## 34                   Rhodothermaeota 5.876350e-02 6.351087e-02
## 36                    Synergistetes 3.552679e-02 1.831805e-02
## 4                   Armatimonadetes 3.249159e-02 1.208742e-02
## 30                 Kiritimatiellaeota 2.209229e-02 1.316739e-02
## 39                       Thermotogae 1.706679e-02 1.113206e-02
## 37                       Tenericutes 1.184226e-02 2.446560e-02
## 38              Thermodesulfobacteria 9.155363e-03 3.489152e-03
## 15                         Chlamydiae 8.906576e-03 8.224429e-03
## 3                          Aquificae 8.707546e-03 5.939865e-03
## 9                    Calditrichaeota 7.811913e-03 3.073776e-03
## 11          Candidatus Bipolaricaulota 7.712398e-03 2.699939e-03
## 27                      Fusobacteria 6.667493e-03 2.085183e-02
## 29                    Ignavibacteriae 6.169919e-03 6.272165e-03
## 13            Candidatus Omnitrophica 5.921132e-03 1.079976e-03
## 21                    Deferribacteres 5.821617e-03 4.319902e-03
## 7                       Balneolaeota 5.224528e-03 5.233727e-03
## 18                      Chrysiogenetes 5.224528e-03 3.572227e-03
## 24                      Elusimicrobia 3.383504e-03 2.450714e-03
## 25                       Fibrobacteres 3.134717e-03 1.536888e-03
## 14          Candidatus Saccharibacteria 1.841024e-03 9.719780e-03
## 23                        Dictyoglomi 1.741509e-03 2.699939e-03
## 19              Coprothermobacterota 1.592237e-03 2.076876e-04
## 12            Candidatus Cloacimonetes 1.393207e-03 1.744576e-03
## 8                         Caldiserica 1.094663e-03 3.323002e-04
## 10 Candidatus Absconditabacteria 7.961185e-04 2.907626e-04
## 5                       Atribacterota 7.463611e-04 3.738377e-04
```

```
rel_phyls[order(-rel_phyls$S1_S286), ]
```

```
##                           Phylum    Pmt2_S287       S1_S286
## 33              Proteobacteria 6.088242e+01 4.824940e+01
## 2                Actinobacteria 2.541399e+01 2.644482e+01
## 6                 Bacteroidetes 2.005273e+00 1.464214e+01
## 32               Planctomycetes 3.599600e+00 6.877866e+00
## 26                    Firmicutes 1.667221e+00 1.481893e+00
## 20                 Cyanobacteria 6.619726e-01 4.798830e-01
## 28              Gemmatimonadetes 6.533148e-01 4.586988e-01
## 22            Deinococcus-Thermus 4.753325e-01 3.570150e-01
## 1                  Acidobacteria 2.650876e+00 3.216666e-01
## 40               Verrucomicrobia 6.798852e-01 2.301179e-01
## 17                    Chloroflexi 3.312848e-01 1.197111e-01
## 34                Rhodothermaeota 5.876350e-02 6.351087e-02
## 35                   Spirochaetes 8.767255e-02 5.021886e-02
## 31                    Nitrospirae 5.496701e-01 3.053008e-02
## 16                        Chlorobi 6.597832e-02 2.820398e-02
## 37                      Tenericutes 1.184226e-02 2.446560e-02
## 27                      Fusobacteria 6.667493e-03 2.085183e-02
## 36                    Synergistetes 3.552679e-02 1.831805e-02
## 30                 Kiritimatiellaeota 2.209229e-02 1.316739e-02
```

```
## 4              Armatimonadetes 3.249159e-02 1.208742e-02
## 39                 Thermotogae 1.706679e-02 1.113206e-02
## 14  Candidatus Saccharibacteria 1.841024e-03 9.719780e-03
## 15                   Chlamydiae 8.906576e-03 8.224429e-03
## 29               Ignavibacteriae 6.169919e-03 6.272165e-03
## 3                     Aquificae 8.707546e-03 5.939865e-03
## 7                   Balneolaeota 5.224528e-03 5.233727e-03
## 21                Deferribacteres 5.821617e-03 4.319902e-03
## 18                Chrysiogenetes 5.224528e-03 3.572227e-03
## 38           Thermodesulfobacteria 9.155363e-03 3.489152e-03
## 9                 Calditrichaeota 7.811913e-03 3.073776e-03
## 11  Candidatus Bipolaricaulota 7.712398e-03 2.699939e-03
## 23                   Dictyoglomi 1.741509e-03 2.699939e-03
## 24                 Elusimicrobia 3.383504e-03 2.450714e-03
## 12      Candidatus Cloacimonetes 1.393207e-03 1.744576e-03
## 25                 Fibrobacteres 3.134717e-03 1.536888e-03
## 13     Candidatus Omnitrophica 5.921132e-03 1.079976e-03
## 5                   Atribacterota 7.463611e-04 3.738377e-04
## 8                    Caldiserica 1.094663e-03 3.323002e-04
## 10 Candidatus Absconditabacteria 7.961185e-04 2.907626e-04
## 19            Coprothermobacterota 1.592237e-03 2.076876e-04
```

```
rel_phyls %>% filter(rel_phyls$Pmt2_S287 > 2.2*rel_phyls$S1_S286)
```

```
##                             Phylum    Pmt2_S287        S1_S286
## 1                    Acidobacteria 2.6508756557 0.3216665517
## 2                  Armatimonadetes 0.0324915873 0.0120874182
## 3                      Caldiserica 0.0010946630 0.0003323002
## 4                  Calditrichaeota 0.0078119130 0.0030737765
## 5  Candidatus Absconditabacteria 0.0007961185 0.0002907626
## 6     Candidatus Bipolaricaulota 0.0077123982 0.0026999388
## 7        Candidatus Omnitrophica 0.0059211315 0.0010799755
## 8                          Chlorobi 0.0659783227 0.0282039758
## 9                      Chloroflexi 0.3312848209 0.1197111315
## 10           Coprothermobacterota 0.0015922370 0.0002076876
## 11                     Nitrospirae 0.5496700835 0.0305300769
## 12          Thermodesulfobacteria 0.0091553630 0.0034891516
## 13               Verrucomicrobia 0.6798852196 0.2301178586
```

```
rel_phyls %>% filter(1.5*rel_phyls$Pmt2_S287 < rel_phyls$S1_S286)
```

```
##                            Phylum   Pmt2_S287      S1_S286
## 1                    Bacteroidetes 2.005273290 14.642141807
## 2 Candidatus Saccharibacteria 0.001841024  0.009719780
## 3                     Dictyoglomi 0.001741509  0.002699939
## 4                    Fusobacteria 0.006667493  0.020851835
## 5                   Planctomycetes 3.599600149  6.877865570
## 6                      Tenericutes 0.011842263  0.024465599
```

```
rel_phyls %>% filter(abs(rel_phyls$Pmt2_S287 - rel_phyls$S1_S286) < 0.001)
```

```
##                             Phylum    Pmt2_S287       S1_S286
```

4

```
## 1                Atribacterota 0.0007463611 0.0003738377
## 2                Balneolaeota 0.0052245278 0.0052337275
## 3                  Caldiserica 0.0010946630 0.0003323002
## 4 Candidatus Absconditabacteria 0.0007961185 0.0002907626
## 5      Candidatus Cloacimonetes 0.0013932074 0.0017445758
## 6                  Chlamydiae 0.0089065760 0.0082244289
## 7                  Dictyoglomi 0.0017415093 0.0026999388
## 8                Elusimicrobia 0.0033835037 0.0024507137
## 9              Ignavibacteriae 0.0061699186 0.0062721655
```

The deepth of the samples are the following:

```
sample_sums(raw_metagenomes)
```

```
## Pmt2_S287    S1_S286
##    2009751    2407462
```

```
relative_genus <- tax_glom(relative_metagenomes, taxrank = 'Genus')
relative_genus_df <- psmelt(relative_genus)
relative_genus_df$Genus[relative_genus_df$Abundance < 1] <- "Genus < 1% abundance"
```

Genus with abundance > 1% in Pmt2.

```
relative_genus_df %>% filter(Sample=="Pmt2_S287" & Abundance >= 1)%>% count(Genus)
```

```
##                 Genus n
## 1       Bradyrhizobium 1
## 2         Burkholderia 1
## 3          Cupriavidus 1
## 4           Luteitalea 1
## 5           Lysobacter 1
## 6        Mesorhizobium 1
## 7     Mycolicibacterium 1
## 8         Nocardioides 1
## 9     Paraburkholderia 1
## 10         Pseudomonas 1
## 11        Sphingomonas 1
## 12        Streptomyces 1
## 13          Variovorax 1
```

Genus with abundance > 1% in Topaz.

```
relative_genus_df %>% filter(Sample=="S1_S286" & Abundance >= 1)%>% count(Genus)
```

```
##                    Genus n
## 1           Alienimonas 1
## 2     Antarcticibacterium 1
## 3        Bradyrhizobium 1
## 4          Conexibacter 1
## 5               Devosia 1
## 6              Halomonas 1
```

```
## 7         Hymenobacter 1
## 8         Nocardioides 1
## 9          Paracoccus 1
## 10        Pontibacter 1
## 11        Pseudomonas 1
## 12        Sphingomonas 1
## 13        Streptomyces 1
```

```
relative_genus_df %>% filter(Sample=="Pmt2_S287" & Abundance < 1)%>% count(Genus)
```

```
##                     Genus    n
## 1 Genus < 1% abundance 1692
```

```
relative_genus_df %>% filter(Sample=="S1_S286" & Abundance < 1)%>% count(Genus)
```

```
##                     Genus    n
## 1 Genus < 1% abundance 1692
```

If we subset just by Actinobacteria, we obtain the following.

```
relative_actinos_df2 <- psmelt(relative_actinos_genus)
length(sort(unique(relative_actinos_df2[,"Genus"])))
```

```
## [1] 222
```

There are 222 unique genus across Actinobacteria.

There are 17 unique Genus in sample Pmt2 with abundance >= 1.

```
relative_actinos_df %>% filter(Sample=="Pmt2_S287" & Abundance >= 1)%>% count(Genus)
```

```
##                  Genus n
## 1        Actinomadura 1
## 2        Actinoplanes 1
## 3       Amycolatopsis 1
## 4        Arthrobacter 1
## 5        Cellulomonas 1
## 6     Corynebacterium 1
## 7            Gordonia 1
## 8      Microbacterium 1
## 9      Micromonospora 1
## 10      Mycobacterium 1
## 11 Mycolicibacterium 1
## 12           Nocardia 1
## 13       Nocardioides 1
## 14         Nonomuraea 1
## 15     Pseudonocardia 1
## 16       Rhodococcus 1
## 17       Streptomyces 1
```

There are 19 unique Genus in sample S1 with abundance >= 1.

```
relative_actinos_df %>% filter(Sample=="S1_S286" & Abundance >= 1)%>% count(Genus)
```

```
##                  Genus n
## 1         Actinoplanes 1
## 2         Amycolatopsis 1
## 3          Arthrobacter 1
## 4              Baekduia 1
## 5           Cellulomonas 1
## 6           Conexibacter 1
## 7        Corynebacterium 1
## 8               Frankia 1
## 9         Microbacterium 1
## 10        Micromonospora 1
## 11         Modestobacter 1
## 12          Mycobacterium 1
## 13 Mycolicibacterium 1
## 14              Nocardia 1
## 15          Nocardioides 1
## 16         Pseudonocardia 1
## 17           Rhodococcus 1
## 18           Rubrobacter 1
## 19           Streptomyces 1
```

And in total, there are 22 Genus with abundance over than 1%.

```
relative_actinos_df %>% filter(Abundance >= 1)%>% count(Genus)
```

```
##                  Genus n
## 1         Actinomadura 1
## 2         Actinoplanes 2
## 3         Amycolatopsis 2
## 4          Arthrobacter 2
## 5              Baekduia 1
## 6           Cellulomonas 2
## 7           Conexibacter 1
## 8        Corynebacterium 2
## 9               Frankia 1
## 10              Gordonia 1
## 11        Microbacterium 2
## 12        Micromonospora 2
## 13         Modestobacter 1
## 14          Mycobacterium 2
## 15 Mycolicibacterium 2
## 16              Nocardia 2
## 17          Nocardioides 2
## 18            Nonomuraea 1
## 19         Pseudonocardia 2
## 20           Rhodococcus 2
## 21           Rubrobacter 1
## 22           Streptomyces 2
```

There are 205 Genus with abundance less than 1% in sample Pmt2 and 203 in Sample S1.

```r
relative_actinos_df %>% group_by(Sample) %>%filter(Abundance < 1)%>% count(Genus)
```

```
## # A tibble: 2 x 3
## # Groups:    Sample [2]
##   Sample    Genus                    n
##   <chr>     <chr>                <int>
## 1 Pmt2_S287 Genus < 1% abundance   205
## 2 S1_S286   Genus < 1% abundance   203
```

```r
relative_genus_df %>% filter(Sample=="Pmt2_S287" & Phylum=="Actinobacteria" & Abundance >= 1)%>% count(
```

```
##                Genus n
## 1 Mycolicibacterium 1
## 2        Nocardioides 1
## 3        Streptomyces 1
```

```r
relative_genus_df %>% filter(Sample=="S1_S286" & Phylum=="Actinobacteria" & Abundance >= 1)%>% count(Ge
```

```
##            Genus n
## 1 Conexibacter 1
## 2 Nocardioides 1
## 3 Streptomyces 1
```

```r
relative_genus_df %>% filter(Sample=="Pmt2_S287" & Phylum=="Actinobacteria" & Abundance < 1)%>% count(G
```

```
##                  Genus   n
## 1 Genus < 1% abundance 219
```

```r
relative_genus_df %>% filter(Sample=="S1_S286" & Phylum=="Actinobacteria" & Abundance < 1)%>% count(Genu
```

```
##                  Genus   n
## 1 Genus < 1% abundance 219
```