Supplemental Information 1: Data cleaning

Simpson, Bettauer et al.

April 2020

This supplemental methods describes additional analysis of non-microbial taxa in our dataset, and our code to remove those taxa. We start by loading the R data frame `tree` that contains the read counts on the tree of life (from `exp4`).

```
options(warn = -1)
library(xtable); library(ggplot2); library(vcd); library(MASS); library(FNN); library(rlang)
source("~/repo/reefmicrobiome/src/functions.R")

# Load the tree data.frame with Bracken counts etc.
REEF_DIR <- "/home/data/refined/reef/R/"
load( paste0(REEF_DIR, "raw.tree.april.9.RData" ) )  # loads tree data.frame
original <- tree # for safe keeping
date <- "april.13"
```

Examine the four children from the root of the tree, we see that 95.5% (Bellairs, B) and 96.6% (Maycocks, M) of all reads map to 'cellular organisms' which include Archaea, Bacteria and Eukaryota. A small fraction of the reads could not be classified by Kraken/Bracken (0.4% B 0.6% M). Lastly, a small fraction (0.6% B 0.3% M) of reads mapped to plasmids and synehtic seqences, and were therefore removed from further analysis.

```
make_table(1)        # root
```

```
##                     Name Tax. Id. Parent         Rank Local.Freq.Bel
## 1     cellular organisms   131567      1      no rank          0.961
## 2               Viruses    10239      1 superkingdom          0.035
## 3 unclassified sequences    12908      1      no rank          0.004
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.969    -0.009         0.961         0.969    -0.008
## 2          0.025     0.333         0.035         0.025     0.010
## 3          0.006    -0.351         0.004         0.006    -0.002
```

```
make_table(131567)   # cellular organisms
```

```
##        Name Tax. Id. Parent         Rank Local.Freq.Bel Local.Freq.May
## 1  Bacteria        2 131567 superkingdom          0.497          0.608
## 2 Eukaryota     2759 131567 superkingdom          0.480          0.379
## 3   Archaea     2157 131567 superkingdom          0.023          0.014
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1    -0.201         0.477         0.589    -0.112
## 2     0.237         0.461         0.367     0.094
## 3     0.535         0.023         0.013     0.009
```

```
make_table(28384)    # other sequences
```

```
## [1] 0
## [1] Name     Tax. Id. Parent   Rank
## <0 rows> (or 0-length row.names)
```

```
void <- remove_update_tree( 28384 )
```

```
## [1] 0
```

```
## Error in -t2i(tmp$tax_id): invalid argument to unary operator
```

```
#save( tree, file = paste0(REEF_DIR, "tree.other_sequences.april.9.RData" ))
#write.csv( tree, file = paste0(REEF_DIR, "tree.other_sequences.april.9.csv" ))
```

The updated frequencies at the root are now as follows.

```
make_table(1)        # root
```

```
##                    Name Tax. Id. Parent         Rank Local.Freq.Bel
## 1     cellular organisms   131567      1      no rank          0.961
## 2               Viruses    10239      1 superkingdom          0.035
## 3 unclassified sequences    12908      1      no rank          0.004
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.969    -0.009         0.961         0.969    -0.008
## 2          0.025     0.333         0.035         0.025     0.010
## 3          0.006    -0.351         0.004         0.006    -0.002
```

```
make_table(131567)   # cellular organisms
```

```
##        Name Tax. Id. Parent         Rank Local.Freq.Bel Local.Freq.May
## 1  Bacteria        2 131567 superkingdom          0.497          0.608
## 2 Eukaryota     2759 131567 superkingdom          0.480          0.379
## 3   Archaea     2157 131567 superkingdom          0.023          0.014
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1    -0.201         0.477         0.589    -0.112
## 2     0.237         0.461         0.367     0.094
## 3     0.535         0.023         0.013     0.009
```

Note here the global frequency refers to the number of reads mapped to that taxa divided by the total number of reads at that site. The local frequency for a taxa is the number of reads mapped to that taxa divided by the total number of reads mapped to the taxa and all of its siblings in the tree.

We next focus on cleaning our data of obvious non-microbial taxa. Note that 46.1% B and 36.7%$MofallreadsmaptoEukaryota.Withi$

```
make_table(2759, relative_taxa = 2759)  # euk
```

```
##                       Name Tax. Id. Parent    Rank Local.Freq.Bel
## 1              Opisthokonta    33154   2759 no rank          0.587
## 2             Viridiplantae    33090   2759 kingdom          0.369
## 3                       Sar  2698737   2759 no rank          0.027
## 4                      <NA>  2611352   2759 no rank          0.004
## 5                      <NA>  2608109   2759  phylum          0.004
## 6                Rhodophyta     2763   2759  phylum          0.003
## 7                      <NA>   554915   2759 no rank          0.003
## 8             Cryptophyceae     3027   2759   class          0.002
## 9     environmental samples    61964   2759 no rank          0.001
## 10                     <NA>  2611341   2759 no rank          0.001
## 11                     <NA>   554296   2759 no rank          0.000
## 12      Glaucocystophyceae    38254   2759   class          0.000
## 13         Malawimonadidae   136087   2759  family          0.000
## 14                     <NA>  2683617   2759 no rank          0.000
## 15                     <NA>  2608240   2759 no rank          0.000
## 16 unclassified eukaryotes    42452   2759 no rank          0.000
##    Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1           0.571     0.029         0.271         0.209     0.061
## 2           0.386    -0.046         0.170         0.142     0.028
## 3           0.024     0.130         0.012         0.009     0.004
```

```
## 4           0.004    -0.183         0.002         0.002     0.000
## 5           0.006    -0.587         0.002         0.002    -0.001
## 6           0.002     0.449         0.001         0.001     0.001
## 7           0.003     0.003         0.001         0.001     0.000
## 8           0.001     0.322         0.001         0.000     0.000
## 9           0.002    -0.199         0.001         0.001     0.000
## 10          0.000     0.108         0.000         0.000     0.000
## 11          0.001    -0.431         0.000         0.000    -0.000
## 12          0.000     0.148         0.000         0.000     0.000
## 13          0.000    -0.208         0.000         0.000     0.000
## 14          0.000    -2.213         0.000         0.000    -0.000
## 15          0.000    -0.459         0.000         0.000    -0.000
## 16          0.000     0.170         0.000         0.000     0.000
##    Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1         0.587        0.571         0.016        0.000
## 2         0.369        0.386        -0.017        0.000
## 3         0.027        0.024         0.003        0.000
## 4         0.004        0.004         0.000        0.000
## 5         0.004        0.006        -0.002        0.000
## 6         0.003        0.002         0.001        0.000
## 7         0.003        0.003         0.000        0.826
## 8         0.002        0.001         0.001        0.000
## 9         0.001        0.002        -0.001        0.000
## 10        0.001        0.000         0.001        0.001
## 11        0.000        0.001        -0.001        0.000
## 12        0.000        0.000         0.000        0.130
## 13        0.000        0.000         0.000        0.406
## 14        0.000        0.000         0.000        0.000
## 15        0.000        0.000         0.000        0.241
## 16        0.000        0.000         0.000        0.857
```

Metazoa has 23.2% B and 17.7% M of all reads, and therefore represents a significant source of non-microbioal organims. The remaining taxa which includes a well-represented fungal component consist of single cell or basal Eukaryotic organisms and not excluded from the analysis.

```
make_table(33154, relative_taxa = 2759) # opisthokonta
```

```
##              Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1          Metazoa    33208  33154 kingdom          0.859          0.847
## 2            Fungi     4751  33154 kingdom          0.139          0.150
## 3 Choanoflagellata    28009  33154   class          0.001          0.001
## 4    Rotosphaerida  2686024  33154   order          0.000          0.001
## 5    Ichthyosporea   127916  33154   class          0.000          0.000
## 6       Filasterea  2687318  33154   class          0.000          0.001
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.014         0.232         0.177     0.055        0.504
## 2    -0.077         0.038         0.031     0.006        0.082
## 3    -0.441         0.000         0.000    -0.000        0.001
## 4    -0.159         0.000         0.000     0.000        0.000
## 5     0.077         0.000         0.000     0.000        0.000
## 6    -0.384         0.000         0.000    -0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.483         0.021        0.000
## 2        0.086        -0.004        0.000
## 3        0.001         0.000        0.000
## 4        0.000         0.000        0.002
```

```
## 5        0.000         0.000        0.018
## 6        0.000         0.000        0.000
```

```
make_table(33208, relative_taxa = 2759) # metazoa 23.2% B and 17.7% M of all reads
```

```
##        Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Eumetazoa     6072  33208 no rank          0.997          0.999
## 2  Porifera     6040  33208  phylum          0.003          0.001
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.002         0.232         0.177     0.055        0.503
## 2     1.152         0.001         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.483         0.020        0.000
## 2        0.000         0.001        0.000
```

Although we remove Metazoa from further analysis, we comment briefly on differences bewteen the Bellairs and Maycocks sites here. We start with Porifera, the phylum that contains sponges.

```
make_table(6040, relative_taxa = 2759) # porifera
```

```
##             Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May
## 1   Demospongiae     6042   6040 class          0.997          0.978
## 2       Calcarea    27929   6040 class          0.003          0.016
## 3 Hexactinellida    60882   6040 class          0.000          0.006
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.019         0.001         0.000     0.000        0.001
## 2    -1.759         0.000         0.000    -0.000        0.000
## 3      -Inf         0.000         0.000    -0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.000         0.001        0.000
## 2        0.000         0.000        0.129
## 3        0.000         0.000        0.013
```

Here we see the tables for corals

```
make_table(6072, relative_taxa = 2759) # eumetazoa  23.2% B and 17.7% M
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1  Bilateria    33213   6072 no rank          0.987          0.992
## 2   Cnidaria     6073   6072  phylum          0.010          0.008
## 3   Placozoa    10226   6072  phylum          0.003          0.000
## 4 Ctenophora    10197   6072  phylum          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.005         0.229         0.176     0.053        0.496
## 2     0.244         0.002         0.001     0.001        0.005
## 3     2.366         0.001         0.000     0.001        0.002
## 4    -0.277         0.000         0.000    -0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.479         0.017        0.000
## 2        0.004         0.001        0.000
## 3        0.000         0.002        0.000
## 4        0.000         0.000        0.564
```

```
make_table(6073, relative_taxa = 2759) # Cnidaria
```

```
##        Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May log(BvsM)
## 1  Anthozoa     6101   6073 class          0.907          0.886     0.023
## 2  Hydrozoa     6074   6073 class          0.090          0.113    -0.226
## 3 Scyphozoa     6142   6073 class          0.003          0.001     1.454
##   Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel Rel.Freq.May
## 1         0.002         0.001     0.001        0.005        0.003
```

```
## 2         0.000         0.000     0.000        0.000        0.000
## 3         0.000         0.000     0.000        0.000        0.000
##   Rel.DeltaFreq Two-Portions
## 1         0.002        0.000
## 2         0.000        0.083
## 3         0.000        0.000
```

```
make_table(6101, relative_taxa = 2759) # anthozoa
```

```
##           Name Tax. Id. Parent     Rank Local.Freq.Bel Local.Freq.May
## 1 Hexacorallia     6102   6101 subclass          0.909          0.890
## 2 Octocorallia     6132   6101 subclass          0.091          0.110
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.021         0.002         0.001     0.001        0.004
## 2    -0.191         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.003         0.001        0.000
## 2        0.000         0.000        0.001
```

```
make_table(6102, relative_taxa = 2759) # hexacorallia
```

```
##           Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May
## 1 Scleractinia     6125   6102 order          0.613          0.691
## 2   Actiniaria     6103   6102 order          0.387          0.309
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.119         0.001         0.001     0.000        0.003
## 2     0.223         0.001         0.000     0.000        0.002
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.002         0.001        0.000
## 2        0.001         0.001        0.000
```

```
make_table(33208, relative_taxa = 2759) # metazoa
```

```
##        Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Eumetazoa     6072  33208 no rank          0.997          0.999
## 2  Porifera     6040  33208  phylum          0.003          0.001
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.002         0.232         0.177     0.055        0.503
## 2     1.152         0.001         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.483         0.020        0.000
## 2        0.000         0.001        0.000
```

```
make_table(6072, relative_taxa = 2759)  # eumetazoa
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1  Bilateria    33213   6072 no rank          0.987          0.992
## 2   Cnidaria     6073   6072  phylum          0.010          0.008
## 3   Placozoa    10226   6072  phylum          0.003          0.000
## 4 Ctenophora    10197   6072  phylum          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.005         0.229         0.176     0.053        0.496
## 2     0.244         0.002         0.001     0.001        0.005
## 3     2.366         0.001         0.000     0.001        0.002
## 4    -0.277         0.000         0.000    -0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.479         0.017        0.000
## 2        0.004         0.001        0.000
## 3        0.000         0.002        0.000
## 4        0.000         0.000        0.564
```

```
make_table(33213, relative_taxa = 2759) # biltaeria       22.9% B and 17.6% M
```

```
##              Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1   Deuterostomia    33511  33213 no rank          0.788          0.783
## 2     Protostomia    33317  33213 no rank          0.212          0.217
## 3 Xenacoelomorpha  1312402  33213  phylum          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.006         0.180         0.138     0.043        0.391
## 2    -0.022         0.049         0.038     0.010        0.105
## 3    -0.115         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.375         0.016        0.000
## 2        0.104         0.001        0.000
## 3        0.000         0.000        0.964
```

```
make_table(33511, relative_taxa = 2759) # deuterostomia    18% B and 13.8% M
```

```
##            Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1      Chordata     7711  33511 phylum          0.987          0.988
## 2 Echinodermata     7586  33511 phylum          0.012          0.011
## 3  Hemichordata    10219  33511 phylum          0.001          0.001
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.001         0.178         0.136     0.042        0.386
## 2     0.042         0.002         0.002     0.001        0.005
## 3     0.044         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.371         0.015        0.000
## 2        0.004         0.001        0.000
## 3        0.000         0.000        0.020
```

```
make_table(7711, relative_taxa = 2759)  # chordata    17.8% 13.6%
```

```
##              Name Tax. Id. Parent      Rank Local.Freq.Bel Local.Freq.May
## 1        Craniata    89593   7711 subphylum          0.997          0.997
## 2 Cephalochordata     7735   7711 subphylum          0.002          0.002
## 3        Tunicata     7712   7711 subphylum          0.001          0.001
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.177         0.135     0.042        0.385
## 2     0.030         0.000         0.000     0.000        0.001
## 3    -0.074         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.369         0.016        0.000
## 2        0.001         0.000        0.006
## 3        0.000         0.000        0.341
```

```
make_table(89593, relative_taxa = 2759) # craniata
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Vertebrata     7742  89593 no rank          1.000          1.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.177         0.135     0.042        0.385
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.369         0.016        0.000
```

```
make_table(7742, relative_taxa = 2759)  # vertebrate
```

```
##            Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Gnathostomata     7776   7742 no rank          1.000          1.000
## 2  Cyclostomata  1476529   7742 no rank          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
```

```
## 1     0.000         0.177         0.135     0.042        0.385
## 2    -0.076         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.369         0.016        0.000
## 2        0.000         0.000        0.660
```

```
make_table(7776, relative_taxa = 2759)  # Gnathostomata
```

```
##             Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1     Teleostomi   117570   7776 no rank          0.998          0.998
## 2 Chondrichthyes     7777   7776   class          0.002          0.002
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.000         0.177         0.135     0.042        0.384
## 2     0.019         0.000         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.368         0.016        0.000
## 2        0.001         0.000        0.015
```

```
make_table(117570, relative_taxa = 2759) # Teleostomi
```

```
##           Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Euteleostomi   117571 117570 no rank          1.000          1.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.177         0.135     0.042        0.384
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.368         0.016        0.000
```

```
make_table(117571, relative_taxa = 2759) # Euteleostomi
```

```
##             Name Tax. Id. Parent       Rank Local.Freq.Bel Local.Freq.May
## 1 Actinopterygii     7898 117571 superclass          0.546          0.533
## 2  Sarcopterygii     8287 117571 superclass          0.454          0.467
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.023         0.097         0.072     0.025        0.210
## 2    -0.027         0.080         0.063     0.017        0.174
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.196         0.014        0.000
## 2        0.172         0.002        0.000
```

```
make_table(7898, relative_taxa = 2759) # Actinopterygii The subtaxa are different types of fish
```

```
##          Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May
## 1 Actinopteri   186623   7898 class          0.994          0.994
## 2   Cladistia  1338366   7898 class          0.006          0.006
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.000         0.096         0.072     0.024        0.208
## 2     0.070         0.001         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.195         0.013        0.000
## 2        0.001         0.000        0.000
```

```
# 9.7% B and 7.2% M of all reads
```

```
make_table(8287, relative_taxa = 2759) # Sarcopterygii ~56%
```

```
##                   Name Tax. Id. Parent    Rank Local.Freq.Bel
## 1 Dipnotetrapodomorpha  1338369   8287 no rank          0.997
## 2    Coelacanthimorpha   118072   8287   class          0.003
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.997    -0.000         0.080         0.063     0.017
## 2          0.003     0.126         0.000         0.000     0.000
```

```
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.174        0.172         0.002        0.000
## 2        0.001        0.000         0.001        0.000
```

```
# 8.0% B and 6.3% M
```

```
make_table(1338369, relative_taxa = 2759) # Dipnotetrapodomorpha  8% B and 6.3% M
```

```
##        Name Tax. Id.  Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Tetrapoda    32523 1338369 no rank          1.000          1.000
## 2    Dipnoi     7878 1338369   class          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.000         0.080         0.063     0.017        0.174
## 2     0.205         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.172         0.002        0.000
## 2        0.000         0.000        0.680
```

```
make_table(32523, relative_taxa = 2759) # Tetrapoda
```

```
##       Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May log(BvsM)
## 1  Amniota    32524  32523 no rank          0.944          0.948    -0.004
## 2 Amphibia     8292  32523   class          0.056          0.052     0.064
##   Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel Rel.Freq.May
## 1         0.076         0.060     0.016        0.164        0.163
## 2         0.004         0.003     0.001        0.010        0.009
##   Rel.DeltaFreq Two-Portions
## 1         0.001        0.000
## 2         0.001        0.000
```

```
make_table(32524, relative_taxa = 2759) # Amniota  splits 68%/32% Mammalia and Sarospida
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1   Mammalia    40674  32524   class          0.683          0.684
## 2 Sauropsida     8457  32524 no rank          0.317          0.316
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.002         0.052         0.041     0.011        0.112
## 2     0.003         0.024         0.019     0.005        0.052
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.111         0.001        0.000
## 2        0.051         0.001        0.000
```

```
make_table(8457, relative_taxa = 2759) # Saurospida (reptiles and birds) 2.4% B and 1.9% M of all reads
```

```
##     Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May log(BvsM)
## 1 Sauria    32561   8457 no rank          1.000          1.000     0.000
##   Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel Rel.Freq.May
## 1         0.024         0.019     0.005        0.052        0.051
##   Rel.DeltaFreq Two-Portions
## 1         0.001        0.000
```

```
make_table(40674, relative_taxa = 2759) # mammalia 5.2% B and 4.1% M
```

```
##          Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1      Theria    32525  40674 no rank          0.994          0.994
## 2 Prototheria     9254  40674 no rank          0.006          0.006
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.051         0.041     0.011        0.111
## 2    -0.065         0.000         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
```

```
## 1        0.110         0.001        0.000
## 2        0.001         0.000        0.041
```

```
make_table(32525, relative_taxa = 2759) # Theria
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1   Eutheria     9347  32525 no rank          0.973          0.972
## 2 Metatheria     9263  32525 no rank          0.027          0.028
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.050         0.039     0.011        0.108
## 2    -0.010         0.001         0.001     0.000        0.003
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.107         0.001        0.000
## 2        0.003         0.000        0.941
```

```
make_table(9347, relative_taxa = 2759)  # Eutheria 5% B and 3.9% M (We did detect about 0.1% B and B reads
```

```
##            Name Tax. Id. Parent       Rank Local.Freq.Bel Local.Freq.May
## 1 Boreoeutheria  1437010   9347    no rank          0.980          0.979
## 2    Afrotheria   311790   9347 superorder          0.016          0.017
## 3     Xenarthra     9348   9347 superorder          0.004          0.004
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.001         0.049         0.039     0.010        0.106
## 2    -0.058         0.001         0.001     0.000        0.002
## 3    -0.090         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.105         0.001        0.000
## 2        0.002         0.000        0.004
## 3        0.000         0.000        0.017
```

```
make_table(1437010, relative_taxa = 2759) # Boreoeutheria 4.9% B and 3.9% M
```

```
##               Name Tax. Id.  Parent       Rank Local.Freq.Bel
## 1 Euarchontoglires   314146 1437010 superorder          0.606
## 2   Laurasiatheria   314145 1437010 superorder          0.394
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.599     0.013         0.030         0.023     0.007
## 2          0.401    -0.020         0.019         0.015     0.004
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.064        0.063         0.001        0.000
## 2        0.042        0.042         0.000        0.005
```

```
make_table(314146, relative_taxa = 2759) # Euarchontoglires   3% B and 2.3% M
```

```
##         Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1     Glires   314147 314146 no rank          0.556          0.551
## 2   Primates     9443 314146   order          0.433          0.437
## 3 Scandentia     9392 314146   order          0.007          0.007
## 4 Dermoptera    30656 314146   order          0.004          0.004
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.009         0.017         0.013     0.004        0.036
## 2    -0.010         0.013         0.010     0.003        0.028
## 3    -0.045         0.000         0.000     0.000        0.000
## 4    -0.090         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.035         0.001        0.000
## 2        0.028         0.000        0.002
## 3        0.000         0.000        0.539
## 4        0.000         0.000        0.132
```

```
make_table(314145, relative_taxa = 2759) # Laurasiatheria 1.9% B and 1.5% M (bats pangolin whale dolphin e
```

```
##             Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May
## 1   Artiodactyla    91561 314145 order          0.557          0.549
## 2      Carnivora    33554 314145 order          0.297          0.299
## 3     Chiroptera     9397 314145 order          0.081          0.085
## 4 Perissodactyla     9787 314145 order          0.036          0.037
## 5   Eulipotyphla     9362 314145 order          0.019          0.021
## 6      Pholidota     9971 314145 order          0.009          0.009
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.015         0.011         0.009     0.002        0.023
## 2    -0.007         0.006         0.005     0.001        0.012
## 3    -0.045         0.002         0.001     0.000        0.003
## 4    -0.014         0.001         0.001     0.000        0.002
## 5    -0.092         0.000         0.000     0.000        0.001
## 6    -0.012         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.023         0.000        0.248
## 2        0.013        -0.001        0.011
## 3        0.004        -0.001        0.000
## 4        0.002         0.000        0.208
## 5        0.001         0.000        0.000
## 6        0.000         0.000        0.578
```

```
make_table(314147, relative_taxa = 2759) # Glires (from Euarchontoglires,  (rodents, hamster etc,) 1.7% an
```

```
##         Name Tax. Id. Parent  Rank Local.Freq.Bel Local.Freq.May log(BvsM)
## 1   Rodentia     9989 314147 order          0.981          0.979     0.002
## 2 Lagomorpha     9975 314147 order          0.019          0.021    -0.100
##   Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel Rel.Freq.May
## 1         0.016         0.012     0.004        0.035        0.034
## 2         0.000         0.000     0.000        0.001        0.001
##   Rel.DeltaFreq Two-Portions
## 1         0.001        0.000
## 2         0.000        0.012
```

```
make_table(9443, relative_taxa = 2759)   # Primates 1.3% and 1.0%
```

```
##            Name Tax. Id. Parent     Rank Local.Freq.Bel Local.Freq.May
## 1   Haplorrhini   376913   9443 suborder          0.962          0.959
## 2 Strepsirrhini   376911   9443 suborder          0.038          0.041
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.003         0.012         0.010     0.003        0.027
## 2    -0.077         0.000         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.026         0.001        0.000
## 2        0.001         0.000        0.003
```

```
make_table(376913, relative_taxa = 2759) # Haplorrhini
```

```
##           Name Tax. Id. Parent       Rank Local.Freq.Bel Local.Freq.May
## 1  Simiiformes   314293 376913 infraorder          0.991          0.990
## 2 Tarsiiformes   376912 376913 infraorder          0.009          0.010
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.001         0.012         0.010     0.003        0.027
## 2    -0.080         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.026         0.001        0.000
## 2        0.000         0.000        0.165
```

```
make_table(314293, relative_taxa = 2759) # Simiiformes
```

```
##          Name Tax. Id. Parent      Rank Local.Freq.Bel Local.Freq.May
## 1  Catarrhini     9526 314293 parvorder          0.939          0.936
## 2 Platyrrhini     9479 314293 parvorder          0.061          0.064
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.004         0.012         0.009     0.003        0.025
## 2    -0.053         0.001         0.001     0.000        0.002
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.024         0.001        0.000
## 2        0.002         0.000        0.040
```

```
make_table(9526, relative_taxa = 2759) # Catarrhini
```

```
##               Name Tax. Id. Parent        Rank Local.Freq.Bel
## 1       Hominoidea   314295   9526 superfamily          0.834
## 2 Cercopithecoidea   314294   9526 superfamily          0.166
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.831     0.004         0.010         0.007     0.002
## 2          0.169    -0.018         0.002         0.002     0.000
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.021        0.020         0.001        0.000
## 2        0.004        0.004         0.000        0.803
```

```
make_table(314294, relative_taxa = 2759) # Old world monkeys .2% B and M of all reads
```

```
##              Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1 Cercopithecidae     9527 314294 family          1.000          1.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.002         0.002     0.000        0.004
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.004         0.000        0.803
```

```
make_table(314295, relative_taxa = 2759) # Hominoidea   1% B 0.7% M of all reads
```

```
##          Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1   Hominidae     9604 314295 family          0.971          0.963
## 2 Hylobatidae     9577 314295 family          0.029          0.037
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.008         0.009         0.007     0.002        0.020
## 2    -0.231         0.000         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.020         0.000        0.000
## 2        0.001         0.000        0.000
```

All there is a statistically signficant difference in the number of reads between Bellairs and Maycocks for all of these taxa, the differences are generally at most $1-2\%$. We remove the subtree reooted at Metazoa from further analysis.

```
make_table(33090, relative_taxa = 2759) # viridiplantae
```

```
##           Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1 Streptophyta    35493  33090 phylum          0.908          0.843
## 2  Chlorophyta     3041  33090 phylum          0.092          0.157
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.074         0.154         0.119     0.035        0.335
## 2    -0.535         0.016         0.022    -0.007        0.034
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.326         0.009        0.000
## 2        0.060        -0.026        0.000
```

```
make_table(35493, relative_taxa = 2759) # Streptophyta
```

```
##                 Name Tax. Id. Parent      Rank Local.Freq.Bel
## 1      Streptophytina   131221  35493 subphylum          1.000
## 2 Klebsormidiophyceae   131220  35493     class          0.000
## 3   Chlorokybophyceae   131213  35493     class          0.000
## 4 Mesostigmatophyceae    96475  35493     class          0.000
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          1.000     0.000         0.154         0.119     0.035
## 2          0.000    -0.126         0.000         0.000     0.000
## 3          0.000    -0.008         0.000         0.000     0.000
## 4          0.000    -0.102         0.000         0.000     0.000
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.335        0.326         0.009        0.000
## 2        0.000        0.000         0.000        0.560
## 3        0.000        0.000         0.000        1.000
## 4        0.000        0.000         0.000        0.831
```

```
# 131221, 3193 (Embryophyta)
# 58023, 78536, 58024, 3398, 1437183, 71240, 91827,
make_table(91827, relative_taxa = 2759)
```

```
##           Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1 Pentapetalae  1437201  91827 no rank          1.000          1.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.124         0.095     0.029        0.269
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.259         0.010        0.000
```

```
make_table( 71275 , relative_taxa = 2759) # rosids
```

```
##                   Name Tax. Id. Parent    Rank Local.Freq.Bel
## 1               fabids    91835  71275 no rank          0.639
## 2              malvids    91836  71275 no rank          0.335
## 3 rosids incertae sedis    91834  71275 no rank          0.026
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.639    -0.001         0.048         0.037     0.011
## 2          0.334     0.002         0.025         0.019     0.006
## 3          0.026    -0.007         0.002         0.002     0.000
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.105        0.101         0.004        0.000
## 2        0.055        0.053         0.002        0.000
## 3        0.004        0.004         0.000        0.006
```

```
# fabids 91835 # 72025 # 3803 # 3814
# malvids 91836

make_table( 71274 , relative_taxa = 2759) # astrids
```

```
##          Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1     lamiids    91888  71274 no rank          0.801          0.803
## 2 campanulids    91882  71274 no rank          0.191          0.189
## 3    Ericales    41945  71274   order          0.008          0.008
## 4    Cornales    41934  71274   order          0.000          0.000
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.002         0.037         0.028     0.009        0.080
## 2     0.010         0.009         0.007     0.002        0.019
## 3    -0.026         0.000         0.000     0.000        0.001
## 4     0.480         0.000         0.000     0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.077         0.003        0.000
```

```
## 2        0.018         0.001        0.000
## 3        0.001         0.000        0.607
## 4        0.000         0.000        0.000
```

```
# lamiids
```

We remove the subtrees …. I have to revisit this to find where multicellularity begins and cut those brances.

We now examine fungi, comment on mulicellular fungi, and remove these branches from the tree of life for further anlsysi.

```
make_table(4751) # fungi  # 3-3.5% are uncertain. Let's ignore.
```

```
##                    Name Tax. Id. Parent       Rank Local.Freq.Bel
## 1               Dikarya   451864   4751 subkingdom          0.961
## 2  Fungi incertae sedis   112252   4751    no rank          0.035
## 3 environmental samples    57731   4751    no rank          0.004
## 4    unclassified Fungi    89443   4751    no rank          0.000
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.967    -0.006         0.036         0.030     0.006
## 2          0.030     0.164         0.001         0.001     0.000
## 3          0.003     0.110         0.000         0.000     0.000
## 4          0.000     0.252         0.000         0.000     0.000
```

```
make_table(451864, relative_taxa = 4751) # Dikarya  breask into 80% ascomycota and 20% basidiomycota
```

```
##            Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1    Ascomycota     4890 451864 phylum          0.795          0.778
## 2 Basidiomycota     5204 451864 phylum          0.205          0.222
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.022         0.029         0.024     0.005        0.764
## 2    -0.080         0.007         0.007     0.001        0.197
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.752         0.012        0.000
## 2        0.215        -0.018        0.000
```

```
make_table(4890, relative_taxa = 4751) # ascomycota
```

```
##                    Name Tax. Id. Parent      Rank Local.Freq.Bel
## 1        saccharomyceta   716545   4890   no rank          0.989
## 2      Taphrinomycotina   451866   4890 subphylum          0.011
## 3 environmental samples   136265   4890   no rank          0.000
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.989    -0.001         0.028         0.023     0.005
## 2          0.011     0.045         0.000         0.000     0.000
## 3          0.000     0.179         0.000         0.000     0.000
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.755        0.744         0.011        0.000
## 2        0.008        0.008         0.000        0.026
## 3        0.000        0.000         0.000        0.459
```

```
make_table(5204, relative_taxa = 4751) # basidiomycota
```

```
##                    Name Tax. Id. Parent      Rank Local.Freq.Bel
## 1       Agaricomycotina     5302   5204 subphylum          0.597
## 2    Ustilaginomycotina   452284   5204 subphylum          0.288
## 3      Pucciniomycotina    29000   5204 subphylum          0.100
## 4      Wallemiomycotina  2204096   5204 subphylum          0.015
## 5 environmental samples   136247   5204   no rank          0.001
##   Local.Freq.May log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq
## 1          0.554     0.074         0.004         0.004     0.001
```

```
## 2          0.319    -0.104         0.002         0.002    -0.000
## 3          0.115    -0.135         0.001         0.001    -0.000
## 4          0.011     0.246         0.000         0.000     0.000
## 5          0.001     0.427         0.000         0.000     0.000
##   Rel.Freq.Bel Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.118        0.119        -0.001        0.062
## 2        0.057        0.069        -0.012        0.000
## 3        0.020        0.025        -0.005        0.000
## 4        0.003        0.002         0.001        0.001
## 5        0.000        0.000         0.000        0.105
```

```
make_table(112252, relative_taxa = 4751)  # other incertae sedis
```

```
##                 Name Tax. Id. Parent   Rank Local.Freq.Bel Local.Freq.May
## 1       Mucoromycota  1913637 112252 phylum          0.788          0.771
## 2    Chytridiomycota     4761 112252 phylum          0.133          0.129
## 3      Microsporidia     6029 112252 phylum          0.079          0.096
## 4      Zoopagomycota  1913638 112252 phylum          0.000          0.001
## 5 Blastocladiomycota   451459 112252 phylum          0.000          0.003
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.022         0.001         0.001     0.000        0.028
## 2     0.027         0.000         0.000     0.000        0.005
## 3    -0.187         0.000         0.000     0.000        0.003
## 4      -Inf         0.000         0.000    -0.000        0.000
## 5      -Inf         0.000         0.000    -0.000        0.000
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.023         0.005        0.000
## 2        0.004         0.001        0.000
## 3        0.003         0.000        0.639
## 4        0.000         0.000        0.005
## 5        0.000         0.000        0.000
```

```
make_table(716545, relative_taxa = 4751) # saccharomyceta
```

```
##               Name Tax. Id. Parent      Rank Local.Freq.Bel Local.Freq.May
## 1   Pezizomycotina   147538 716545 subphylum          0.790          0.807
## 2 Saccharomycotina   147537 716545 subphylum          0.210          0.193
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1    -0.020         0.023         0.019     0.004        0.597
## 2     0.081         0.006         0.005     0.001        0.158
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.600        -0.003        0.010
## 2        0.144         0.014        0.000
```

```
make_table( 147538, relative_taxa = 4751 ) # Pezizomycotina
```

```
##             Name Tax. Id. Parent    Rank Local.Freq.Bel Local.Freq.May
## 1   leotiomyceta   716546 147538 no rank          0.996          0.995
## 2 Orbiliomycetes   189478 147538   class          0.003          0.003
## 3  Pezizomycetes   147549 147538   class          0.002          0.002
##   log(BvsM) Glob.Freq.Bel Glob.Freq.May DeltaFreq Rel.Freq.Bel
## 1     0.000         0.022         0.019     0.004        0.594
## 2    -0.052         0.000         0.000     0.000        0.002
## 3    -0.003         0.000         0.000     0.000        0.001
##   Rel.Freq.May Rel.DeltaFreq Two-Portions
## 1        0.597        -0.003        0.013
## 2        0.002         0.000        0.348
## 3        0.001         0.000        0.955
```