# PRACTICAL: Visualization of 16S amplicon data

In this assignment we will be visualizing the results you obtained in the previous QIIME exercise.

### This exercise consists of one part:

1. Visualisation using MicrobiomeAnalyst

---

## 1. Visualisation using MicrobiomeAnalyst

**QIIME2** is under continous development and the visulaization tools aren't always optimal. An alternative to this is to export the results fenerated in **QIIME2** and make plots using other tools such as the **MicrobiomeAnalyst**. MicrobiomeAnalyst can perform various statistical, visual and meta-analysis of microbiome data on different type of input data (amplicon and shotgun), but we will focus on the amplicon part (Marker Data Profiling).

💡 **MicrobiomeAnalyst** can perform comprehensive composition & diversity analysis supporting various methods of data overview, alpha diversity and beta-diversity; comparative analysis supporting multiple differential abundance methods (metagenomeSeq, LEfSe, edgeR, DESeq2, etc.); as well as prediction of metabolic potentials.


In this exercise you will be working with a dataset consiting of 10 samples. In the directory `/home/practical/1/prerun_all/` we have analysed 10 samples from a research project that includes the four samples you have been analysing in the previous exercise. The output artifact from **QIIME2** are named `table_all.qza` , `taxonomy_all.qza` and `metadata_all.tsv`

In order to import **QIIME2** data into **MicrobiomeAnalyst**, you have to manipulate the files first. In this exercise you will choose to import BIOM files, but there are other file input options that we won't cover here.

First you need to create a BIOM table with taxonomy annotations. You should still have the **Conda** environment with QIIME2 activated.

I] Export the OTU table `table_all.qza` and taxonomy table `taxonomy_all.qza` you generated in the previous exercise:

```
qiime tools export --input-path table_all.qza --output-path exported_all
qiime tools export --input-path taxonomy_all.qza --output-path exported_all

Output: exported_all/feature-table.biom and exported_all/taxonomy.tsv
```

II] Modify the exported taxonomy file's header before using it with BIOM software. The first few lines in `exported_all/taxonomy.tsv` should look something like this (the actual data will differ):

```
Feature ID  Taxon   Confidence
0204aa97a655fab087a6f6902be35eb6    k__Bacteria; p__Bacteroidetes; c__Sphingobacter
5cfeeb662c643e512cbfc724c984f53d    k__Bacteria; p__Firmicutes; c__Bacilli; o__Lact
4dc10ad4afd35b9803d853e819d3cec5    k__Bacteria; p__Firmicutes; c__Clostridia; o__C
...
```

III] Change the first line of `taxonomy.tsv` (i.e. the header) to this using tab characters:

```
#OTUID  taxonomy    confidence
```

IV] Finally, add the taxonomy data to your BIOM file:

```
biom add-metadata -i exported_all/feature-table.biom -o table-with-taxonomy.biom --

Output file: table-with-taxonomy.biom
```

V] Make a copy of the `metadata_all.tsv` file and change .tsv to .txt:

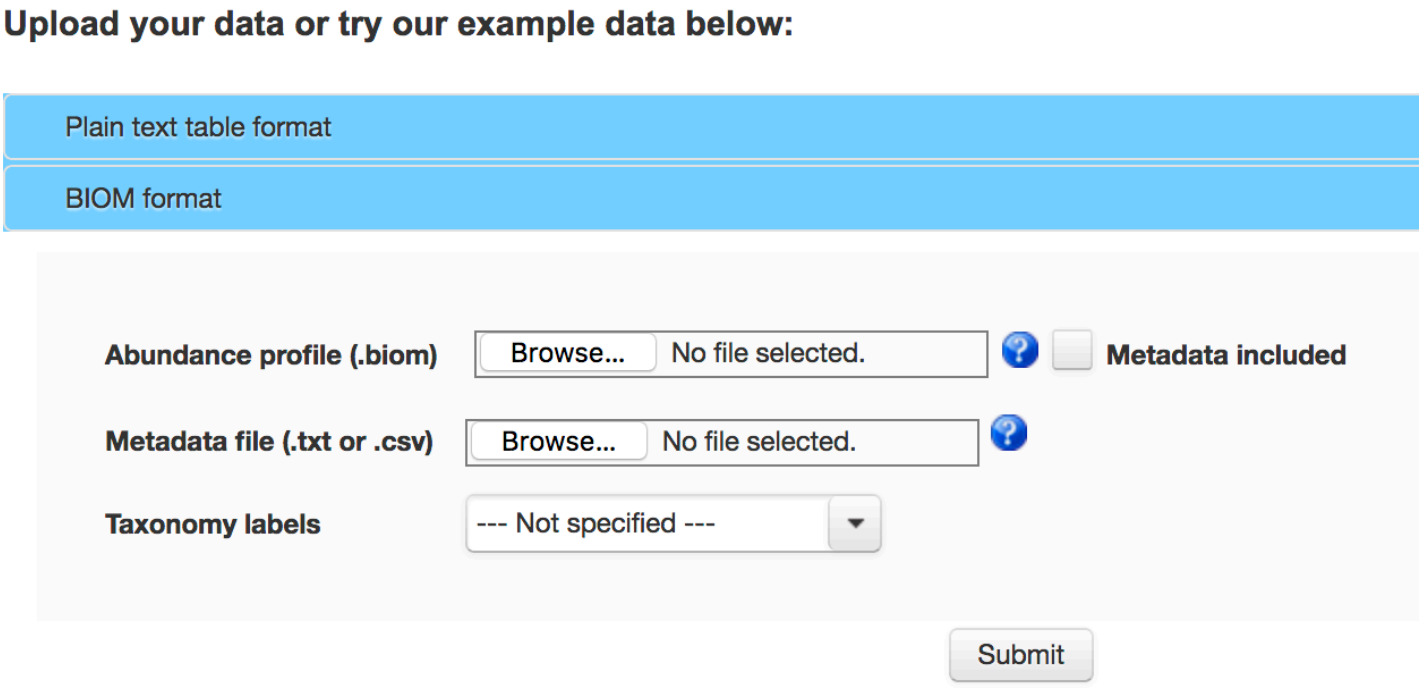
```
Output file: metadata_all.txt
```

VI] Change header in `metadata_all.txt` file:

```
From #SampleID to #NAME
```

VII] Open a web browser and go to https://www.microbiomeanalyst.ca/faces/home.xhtml. Click on "**Marker Data Profiling (MDP)**"

VIII] Select "**BIOM format**", upload the two files you just made ( `table-with-taxonomy.biom` and `metadata_all.txt` ), select "**Taxonomic labels**" "**SILVA taxonomy**" and press "**Submit**"




❓ What is the number of samples included in your data, and how many reads does the smallest data set contain?

▶ **Solution** - Click to expand

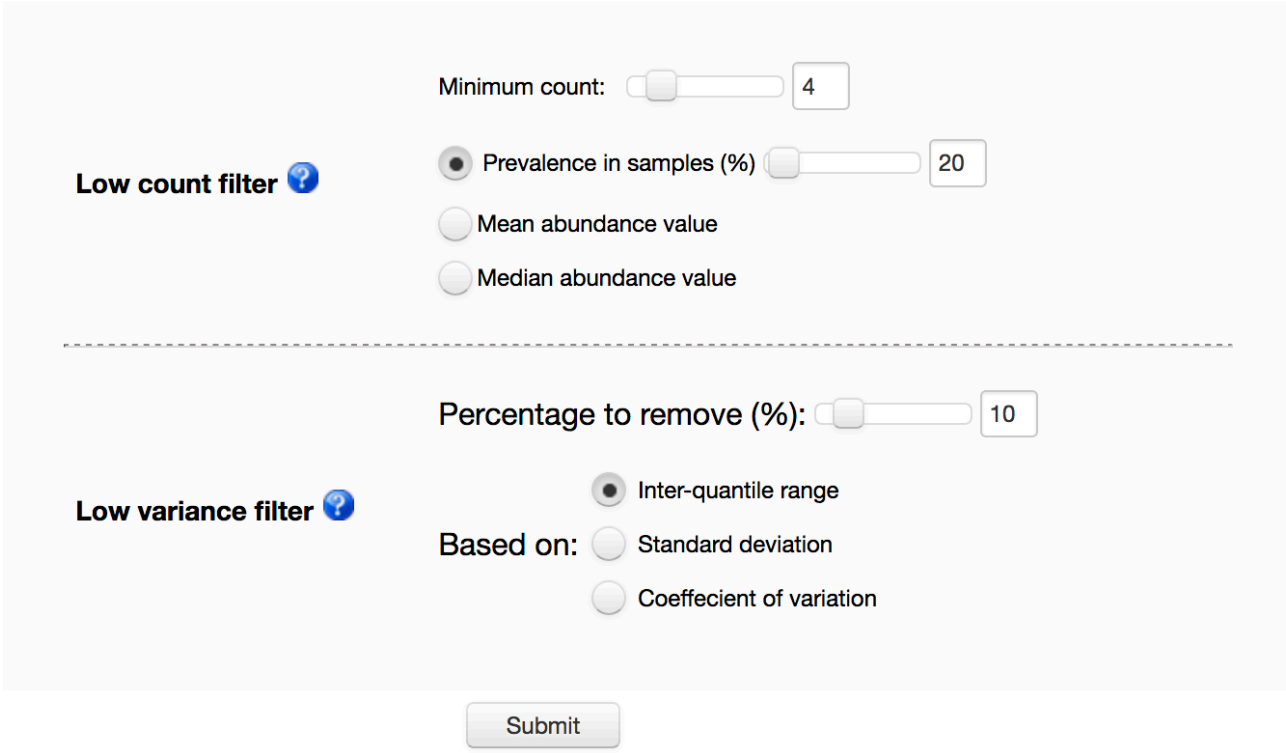
❓ How many OTUs are there in total, and how many OTUs with more than 2 reads are there?

▶ **Solution** - Click to expand

IX] Press "**Proceed**"

It is reccomended to filter low quality and/or uninformative features to improve downstream statisitcal analysis

X] In the new window appering, mouse over the question mark in the "**Feature filtering**" tab and read the filtering options. Continue with the default setting and press "**Submit**".




A box in the upper corner of the web browser will tell you how many features that were removed.

❓ What is the difference between an OTU and a feature?

▶ **Solution** - Click to expand

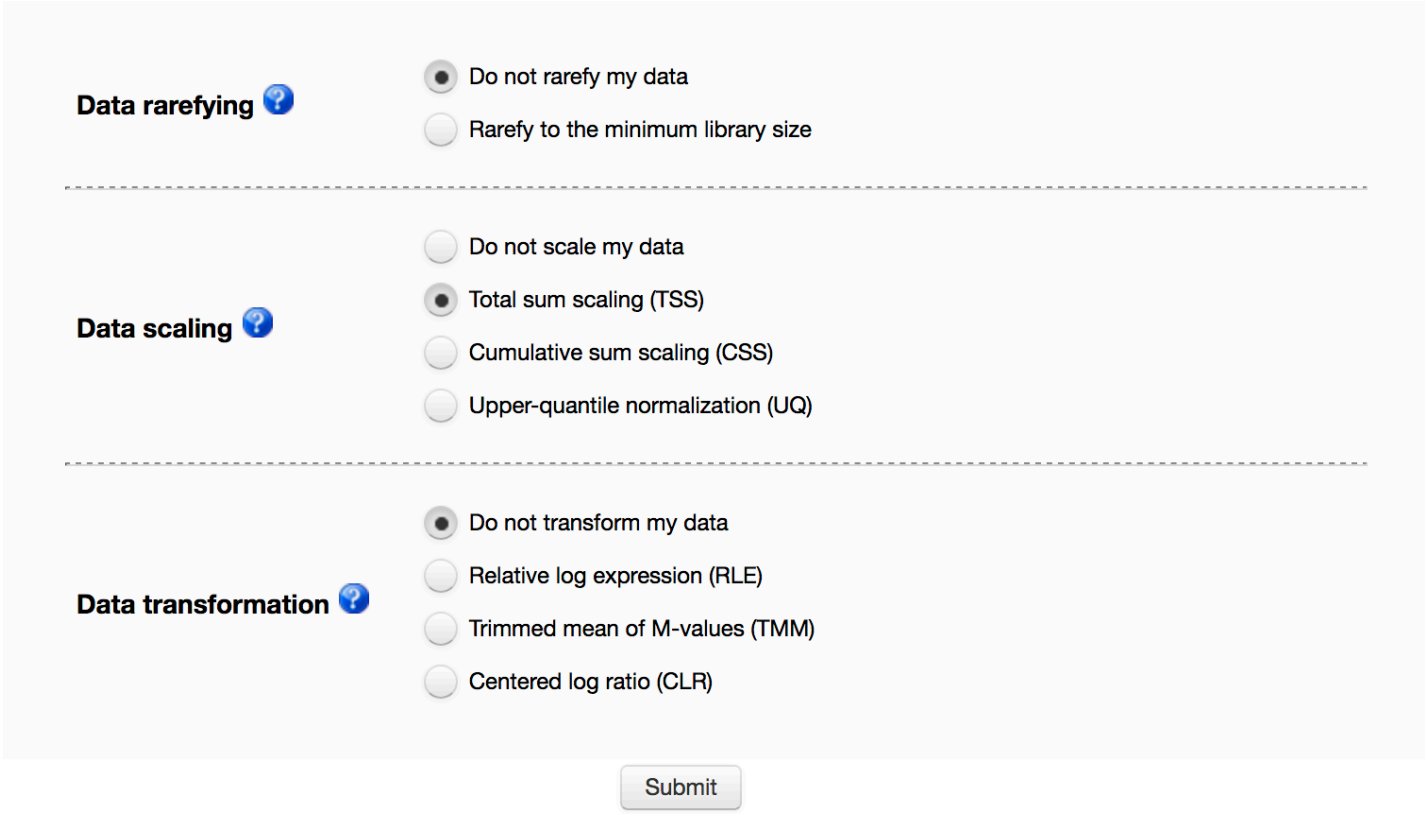
❓ How many features remains after filtering?

▶ **Solution** - Click to expand

XI] Press "**Proceed**"

💡 **Note:** Normalization of the data can be useful to make biologically meaningful comparisons between samples with very diiferent sequencing depths (although rarefication is debated).

XII] In the new window appering, mouse over the question mark in the "**Data rarefying**" tab and read the rarefy options. Continue with the default setting and press "**Submit**"

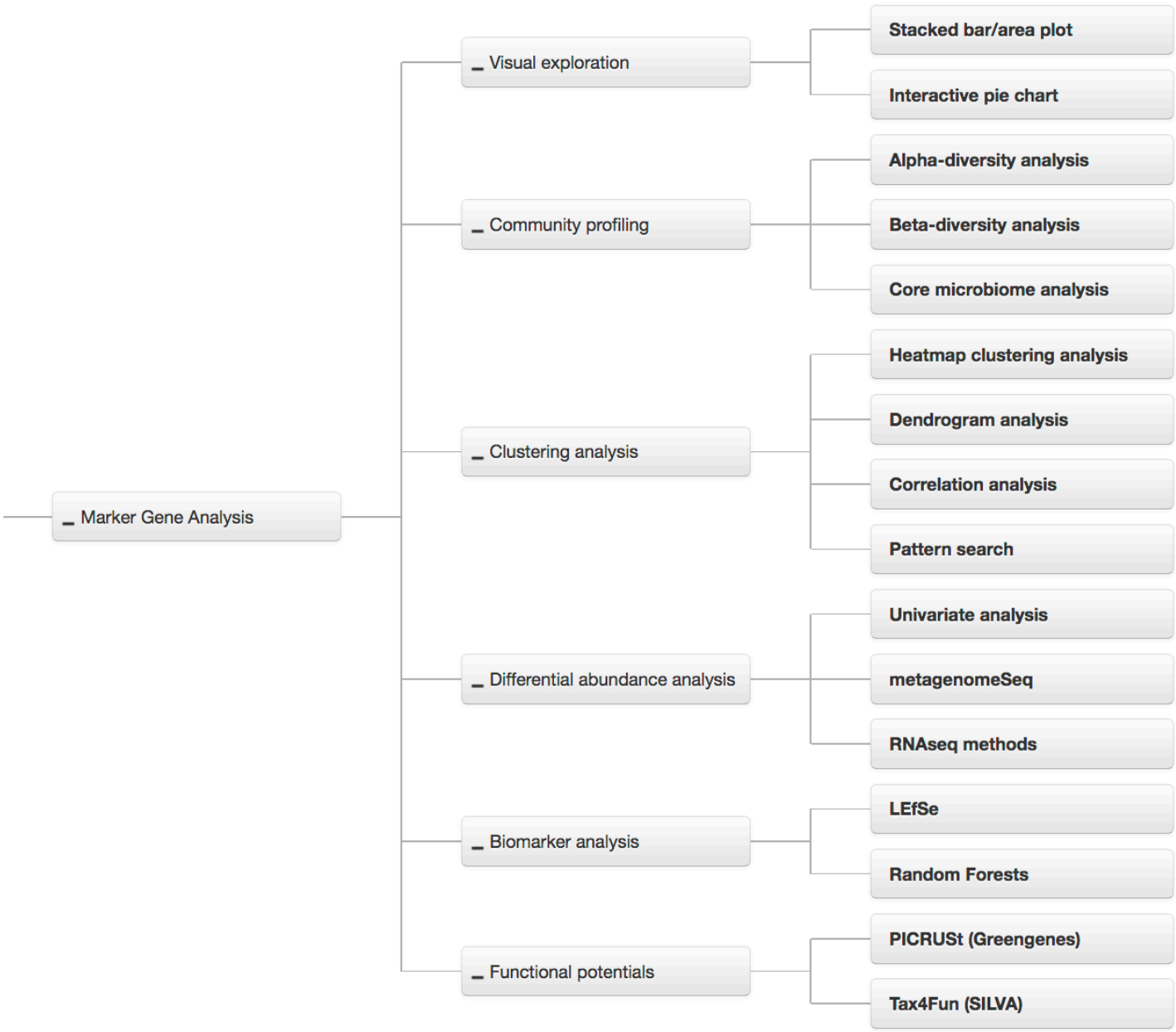



A box in the upper corner of the web browser will tell you that no rarefication was performed on

❓ What rarefication do?

▶ **Solution** - Click to expand

XIII] Press "**Proceed**", and genererate the analysis overview.




XIV] Explore the diffent analysis possibilities and answer the following:

💡 **Note**: The legend for some of the plots may be difficult to read. It is possible to download either the image of the plot or the actual table with counts/relative abundances, etc. for many of the analysis.

❓ What is the most abundant Famlily in sample 37.S37?

▶ **Solution** - Click to expand

❓ What is the most abundant Genus in sample 42.S42?

▶ **Hint** - Click to expand

❓ The metadata "DepthGr" separates the samples into four different sampling depths (). Are there any bacterial families that are relatively more abundant in samples taken from below 400 meters compared to the other samples?

▶ **Hint** - Click to expand

❓ Analysing the diversity of the samples, does the sampling depth seem to have an effect on the alpha diversity?

▶ **Solution** - Click to expand

💡 A alpha diversity analysis will tell you species richness (number of taxa and the distribution of these) within a sample from one environment.

XV] Perform alpha diversity analysis and try changing the metadata (experimental factors) and other parameters.

❓ Are there any diversity measures or statistical methods that shows a significant difference on the samples grouped by sampling depth (depthGr)?

▶ **Solution** - Click to expand

💡 A beta diversity analysis can tell you how different is the microbial composition in one environment compared to another.

XVI] Perform beta diversity analysis and try changing the metadata (experimental factors) and other parameters.

❓ Are there any experimental factors that seem to have an impact on the clustering?

▶ **Solution** - Click to expand

💡 **LEfSe** (Linear discriminant analysis Effect Size) determines the features (organisms, clades, operational taxonomic units, genes, or functions) most likely to explain differences between classes by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance.

**LEfSe** finds taxa which describe a particular group the most, so if you see any taxa colored according to a variable (e.g like sampling depth), you can interpret this as that taxa being significantly increased in abundance compared to the other group. You can read more about **LEfSe** here

XVII] Genereate a **LEfSe** plot.

❓ Does these results correlate with what you found using other plots?

▶ **Solution** - Click to expand

💡 **Note**: For functional analysis with PICRUSt: You need to use closed-reference OTU picking protocol in **QIIME2** and search sequences against the Greengenes reference OTUs (18May2013 version).

We have not provided you with this data for this exercise, therefore you can not do the functional

extrapolation of your data.

XVIII] Before you continue with the next exercise, deactivate the **Conda** environment in the terminal:

```
conda deactivate
```

**Progress tracker**

Complete

# That was the end of the this practical - Good job 👍