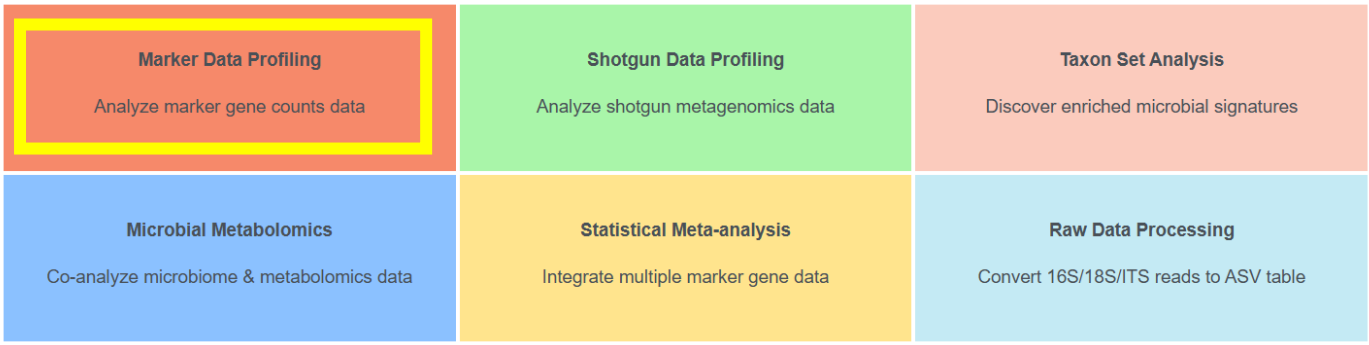
Pipeline – Phase 1: Selection of Microorganisms of Interest (Part 2 – MicrobiomeAnalyst & Selection of microorganisms)

Preparation of files

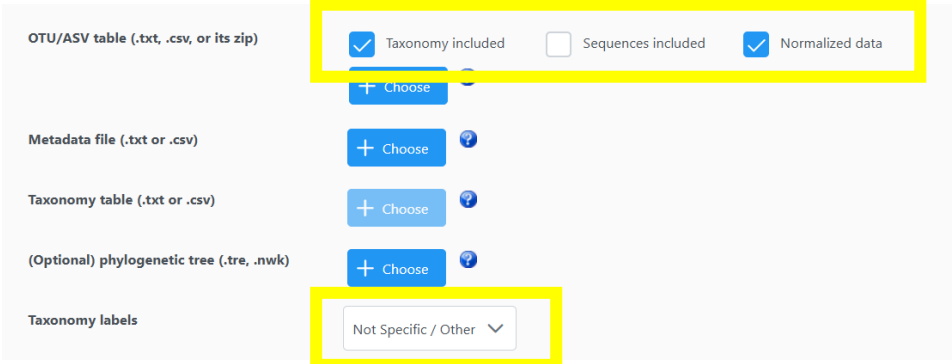
1. Metadata – to create the **metadata.csv** file, use the following template and fill it in: https://github.com/VaneBR/Project/blob/main/Files/metada.csv or https://drive.google.com/drive/u/0/folders/18yojrGbjKc_-GAFdOLlppHP2UWXFu3f1.
2. OUT/ASV table – this file is the output file, **matrix_normalized_microbiomeanalyst.csv**, obtained from running the **normalize_16S.py** script.

Using MicrobiomeAnalyst

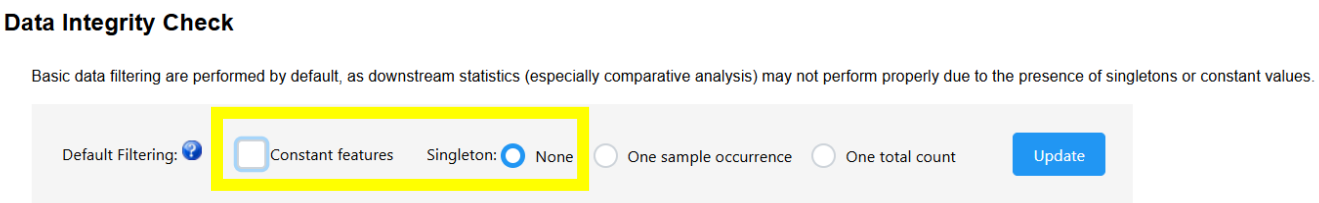
1. Open MicrobiomeAnalyst: https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/ModuleView.xhtml and select



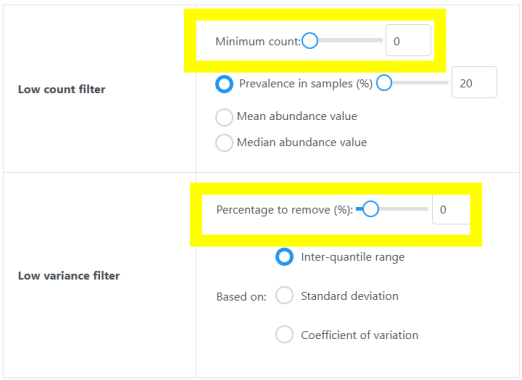
2. Select the following options and upload both files:



3. Select and deselect the following options:

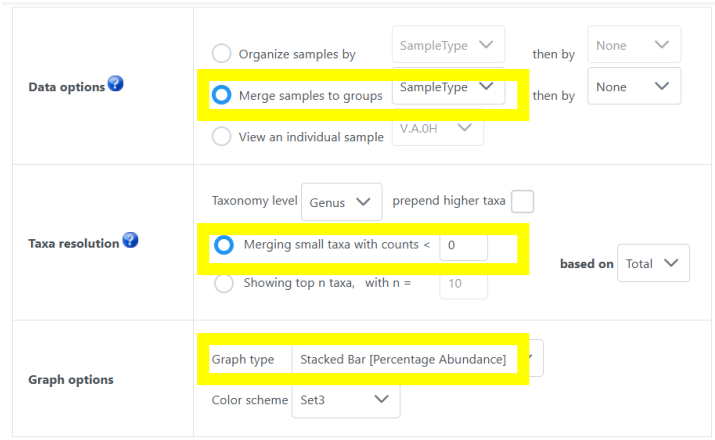


4. Put the “Minimum count” and “Percentage to remove (%)” to 0:



5. In the “Data Normalization” click on submit and then proceed.
6. Now you can visualize and make various statistical assays. For the purpose of this script, we are going to need 4 files from this analysis:
    a. The first one would be “Stacked bar/area plot”

Change the following parameters, hit submit and then download the abundance table:

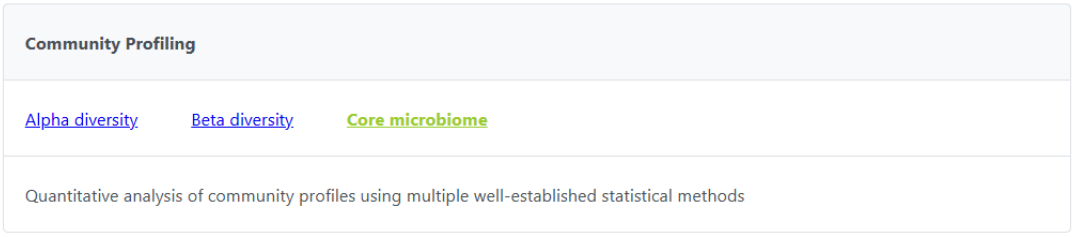


Open the file and substitute “.” with “,” and then create a “Mean Abundance (%)” column in the last column, where you calculate the mean between the different samples, such as:

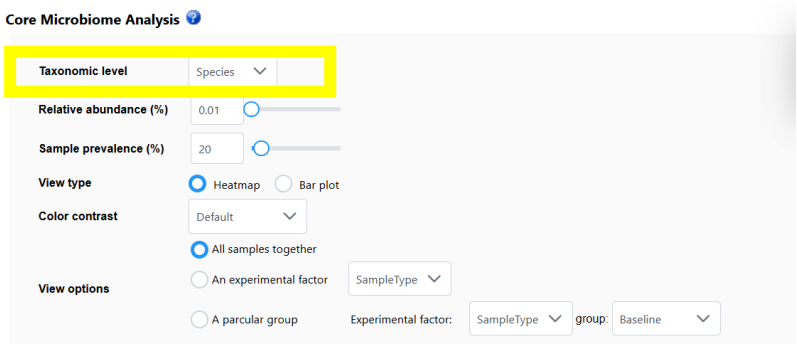
	A	B	C	D	E
1		Baseline	Cellulose	Negative_control	Mean Abundance (%)
2	Acetanaerobacterium	3,33E-05	1,67E-05	0	=MÉDIA(B2:D2)

Finally, substitute “,” with “.” and save the file.

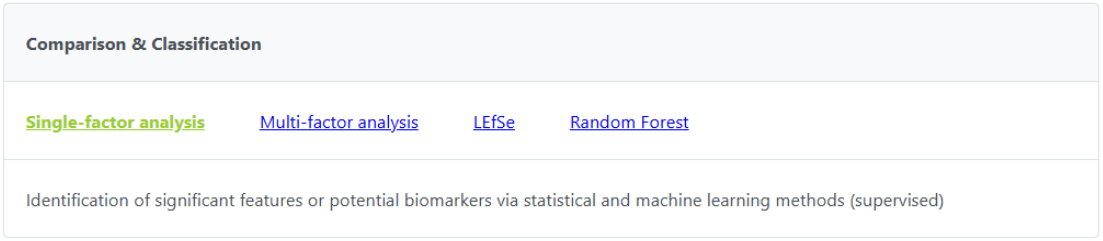
b. In “Core Microbiome” click on the following option:



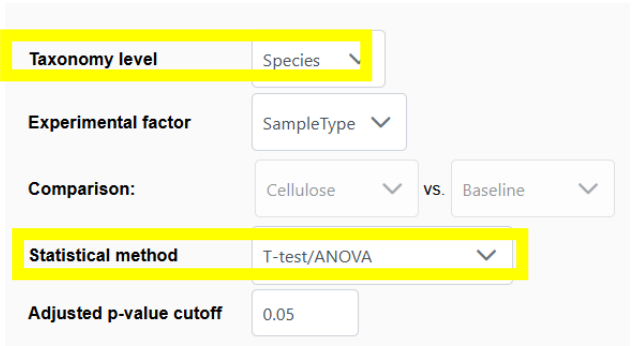
Then change the “Taxonomic level” to “Species”, hit submit and download the result table:



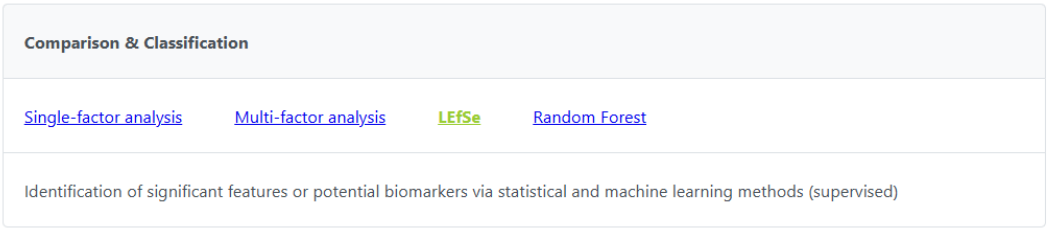
c. For “Single-factor analysis” click the following option:



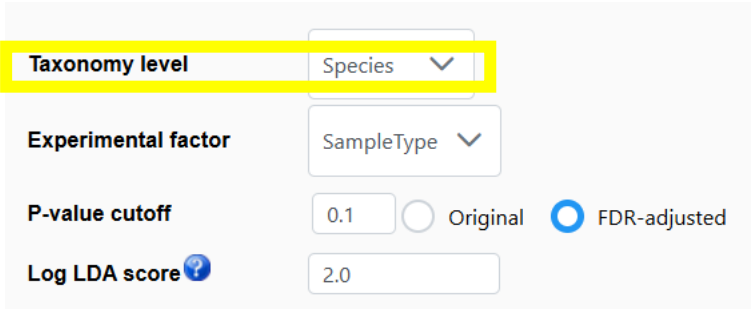
Change the “Taxonomy level” to “Species” and the “Statistical method” to “T-test/ANOVA”, hit submit and download the analysis results:



d. For “LefSe” click the following option:



Change the “Taxonomy level” to “Species”, hit submit and download the analysis results



Steps to use the script:

The script described below is available in the following link: https://colab.research.google.com/drive/1Wk8Fq5X1huzqWaI-HL_7IOpjKeROAXSg or in github: https://github.com/VaneBR/Project/blob/main/Scripts/selection_bacterial_taxa.py.

1. Upload the files: **taxa_abund.csv**, **univar_test_output.csv**, **lefse_de_output.csv** and **core_microbiome.csv** in GoogleColab. Notice that you might need to change the thresholds values since this depends on your data:

```
# Define thresholds
lda_cutoff = 1
pval_cutoff = 0.05
abundance_cutoff = 0.001
core_cutoff = 0.1
```

2. Run the script and obtain two output files: **selected_taxa.csv** and **excluded_taxa.csv**
3. Evaluate your data according to your needs.