MetaPhlAn output merge and visualization

Curtis Huttenhower

Abstract

In this protocol, we will merge the 20 profiled metagenomes in a table of relative abundances and visualize the table with a heatmap.

Guidelines

The commands reported in this protocol can be retrieved as a [bash script](#)

Before start

REQUIREMENTS: the [matplotlib](#) python library installed.

Protocol

Step 1.
Merging the profiled metagenomes is a simple operation that can be performed with a script downloaded with MetaPhlAn and located in the "utils" folder:

⌘ COMMAND
```
$ mkdir output
$ utils/merge_metaphlan_tables.py profiled_samples/*.txt > output/merged_abundance_table.txt
```
The resulting table contains relative abundances with microbial clades as rows and samples as columns.

Step 2.
The "metaphlan_hclust_heatmap.py" script in the MetaPhlAn "plotting_scripts" folder can now be used to perform hierarchical clustering of both samples and clades to generate the heatmap:
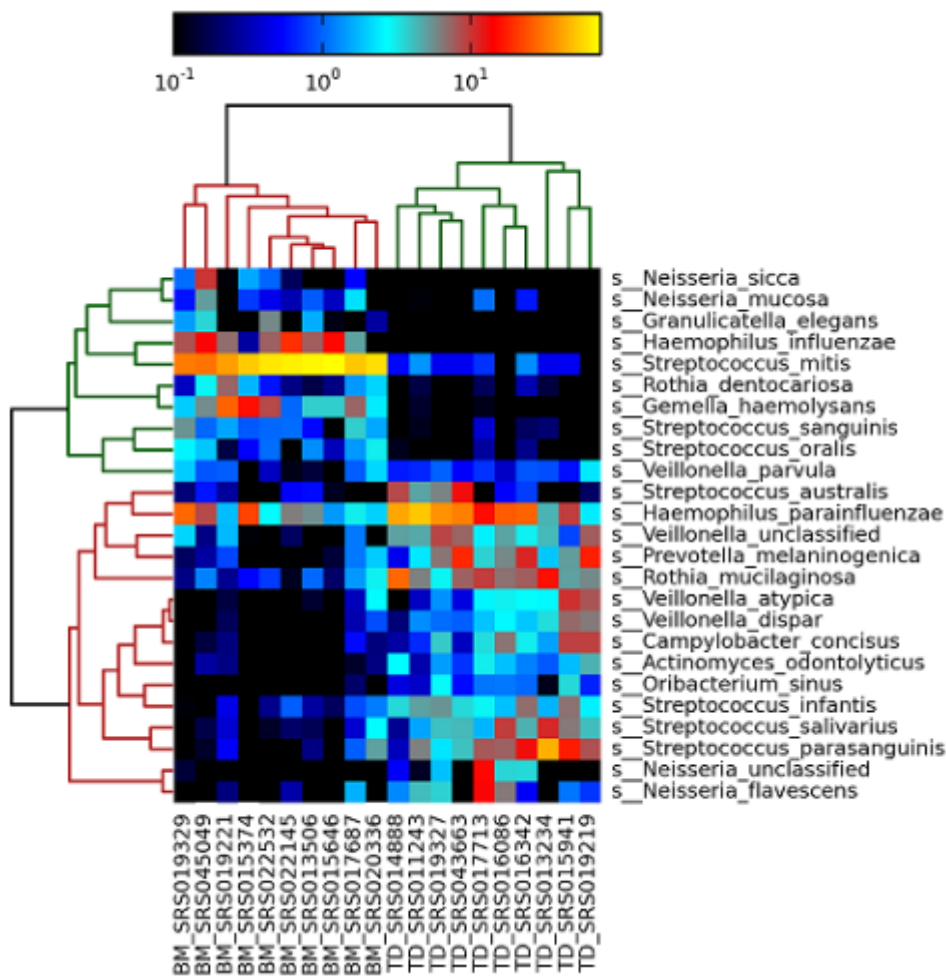
⌘ COMMAND
```
$ mkdir output_images
$ plotting_scripts/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.1 -s log --in output/merged_abundance_table.txt --out output_images/abundance_heatmap.png
```
Specifically, we are displaying the abundances for species only (default --tax_lev s), in logarithmic scale (-s log), reporting only the 25 most abundant clades (--top 25) according to the 90th percentile of the abundances in each clade (default --perc 90) with custom color map (-c bbcry). In this example, the clustering is performed with "average" linkage (default -m average), using "Bray-Curtis" distance for clades (default -d braycurtis) and "correlation" for samples (default -f correlation).

Step 3.

The resulting heatmap is shown here:

Step 4.

For additional customization, enter the following command to list all available options in "metaphlan_hclust_heatmap.py" for the heatmap and/or the clustering.

cmd COMMAND

```
plotting_scripts/metaphlan_hclust_heatmap.py -h
```

Provides a full list of available options for heatmap and/or clustering.