

GraPhlAn visualization of single and multiple samples

Curtis Huttenhower

Abstract

In this protocol we describe some approaches to graphically represent single profiled samples or a merged table of relative abundances.

Citation: Curtis Huttenhower GraPhlAn visualization of single and multiple samples. **protocols.io**

dx.doi.org/10.17504/protocols.io.d9s96d

Published: 14 Jan 2016

Guidelines

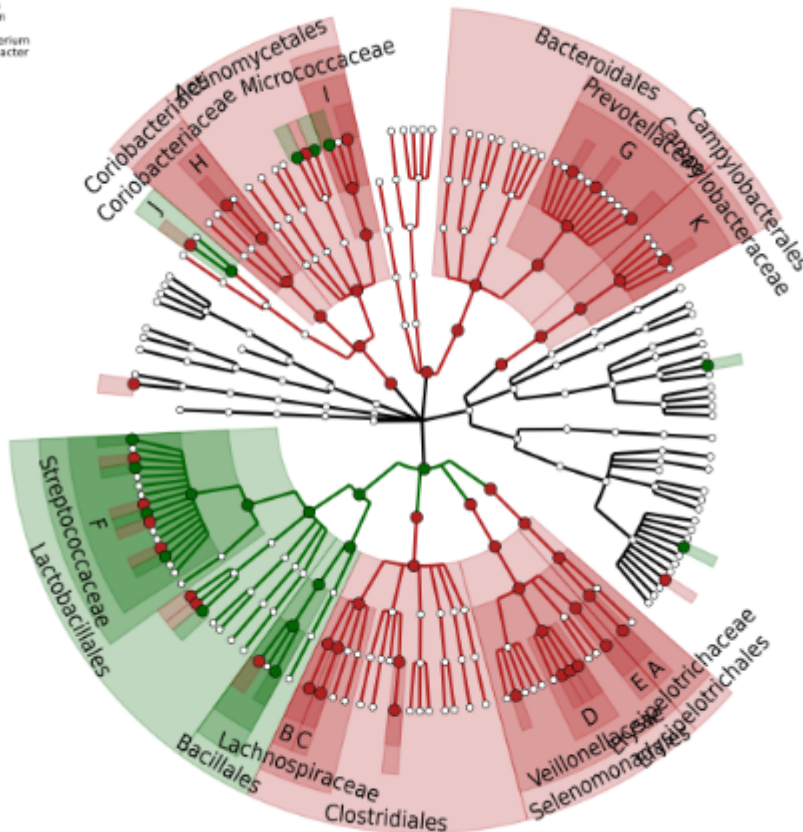
The annotation file can be edited manually and/or with ad-hoc scripts for a higher degree of customization. Please refer to the [GraPhlAn readme file](#) and to these [examples](#) for additional information.

The same script can also be applied to the relative abundance table of 20 merged samples. In this case, the abundances reported refer to the overall average of each clade across all the samples in the table.

```
$ plotting_scripts/metaphlan2graphlan.py output/merged_abundance_table.txt -  
-tree_file tmp/merged.tree.txt --annot_file tmp/merged.annot.txt  
$ graphlan_annotate.py --annot tmp/merged.annot.txt tmp/merged.tree.txt  
tmp/merged.xml  
$ graphlan.py --dpi 200 tmp/merged.xml output_images/merged.png
```

As an example of manually and script-based modification of the "annotation" file provided in [this file](#), we can produce comparisons between classes. Note, this script also includes information about biomarkers (detailed in Step 4: Taxonomic Biomarker discovery with LEfSe). The resulting GraPhlAn image is reported below:

A: Solobacterium
 B: Catonella
 C: Coriobacterium
 D: Veillonella
 E: Megaspheera
 F: Streptococcus
 G: Prevotella
 H: Atopobium
 I: Lactobacillus
 J: Bifidobacterium
 K: Campylobacter



Notice also that MetaPhlAn output can be exported to [Krona](#) (another popular visualization tool) using the script `metaphlan2krona.py` in the `conversion_scripts` folder downloaded as part of the MetaPhlAn package.

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

Before start

REQUIREMENTS: [GraPhlAn](#) installed (and in the system path), and the [matplotlib](#) python library.

GraPhlAn can be downloaded using [Mercurial](#):

`hg clone ssh://hg@bitbucket.org/nsegata/graphlan.`

Protocol

Step 1.

The "`metaphlan2graphlan.py`" script in the `plotting_scripts` folder can generate the two required input files for GraPhlAn which are (i) a tree structure to represent and (ii) graphical annotation options for the tree.

`cmd` **COMMAND**

```
$ mkdir tmp
$ plotting_scripts/metaphlan2graphlan.py profiled_samples/BM_SRS013506.txt --
tree_file tmp/BM_SRS013506.tree.txt --annot_file tmp/BM_SRS013506.annot.txt
```

Step 2.

With these two generated files we can now run GraPhlAn.

cmd **COMMAND**

```
$ graphlan_annotate.py --
annot tmp/BM_SRS013506.annot.txt tmp/BM_SRS013506.tree.txt tmp/BM_SRS013506.xml
$ graphlan.py --dpi 200 tmp/BM_SRS013506.xml output_images/BM_SRS013506.png
```

📌 **NOTES**

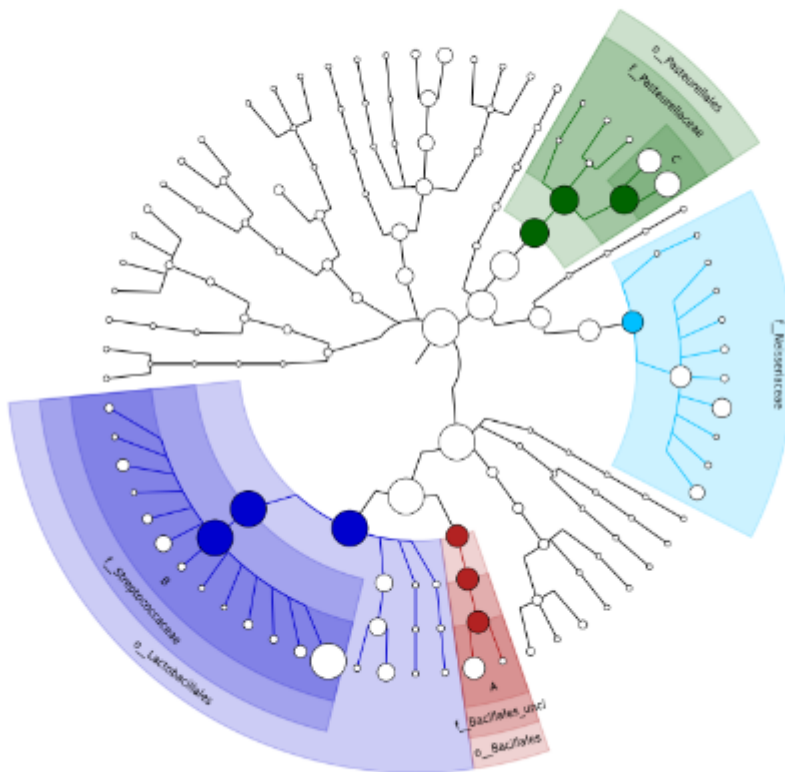
Bahar Sayoldin 09 Dec 2015

Please refer to [GraPhlAn project page](#) for detailed information and additional customization options.

Step 3.

Here is the resulting image:

Ag_Gemella
Bg_Streptococcus
Cg_Haemophilus



Step 4.

Images for all samples can be created with a script similar to this example:

cmd **COMMAND**

```
$ mkdir -p tmp
$ for file in profiled_samples/*
$ do
$   filename=`basename ${file}`
$   samplename=${filename%\. *}
$
```

```
$ plotting_scripts/metaphlan2graphlan.py ${file} --
tree_file tmp/${samplename}.tree.txt --annot_file tmp/${samplename}.annot.txt
$ graphlan_annotate.py --
annot tmp/${samplename}.annot.txt tmp/${samplename}.tree.txt tmp/${samplename}.xml
$ graphlan.py --dpi 200 tmp/${samplename}.xml output_images/${samplename}.png
$ done
```

📌 NOTES

Bahar Sayoldin 02 Dec 2015

There are additional options for customization that can be used to modify the output circular tree. Specifically, several options can be set in the `metaphlan2graphlan.py` script to control the number of annotated clades shown (`--max_annot_clades`, default 10), to set the starting and ending annotated taxonomic levels (`--min_annot_lev` and `--max_annot_lev`), and many more options.

Step 5.

To view additional options for customization input the following command:

```
cmd COMMAND
$ plotting_scripts/metaphlan2graphlan.py -h
Returns additional customization options
```

Step 6.

The annotation file can be edited manually and/or with ad-hoc scripts for a higher degree of customization. Please refer to the [GraPhlAn readme file](#) and to these [examples](#) for additional information.

Step 7.

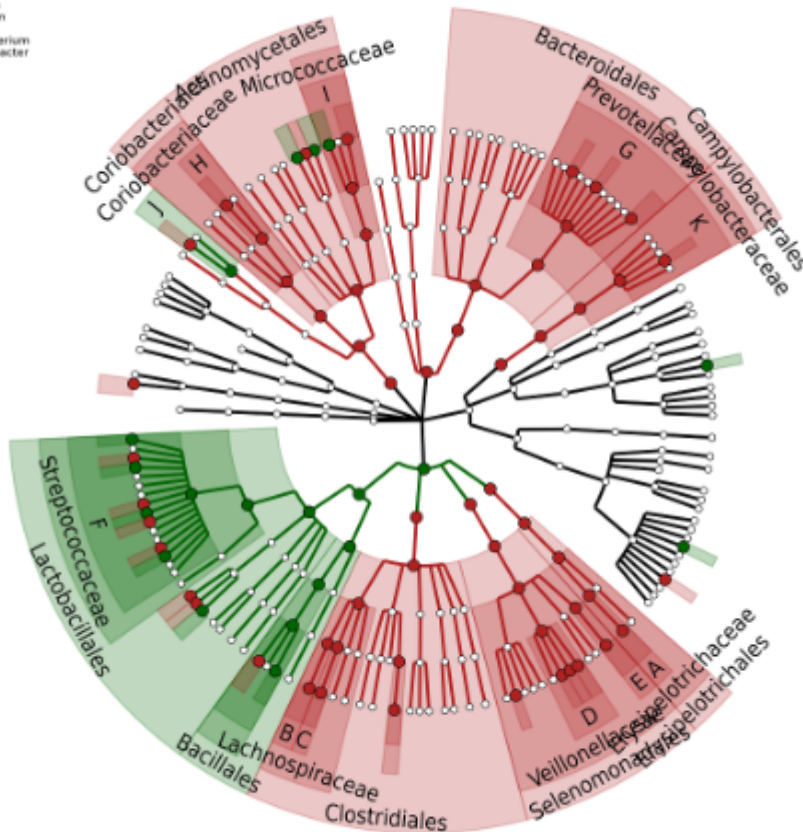
The same script can also be applied to the relative abundance table of 20 merged samples. In this case, the abundances reported refer to the overall average of each clade across all the samples in the table.

```
cmd COMMAND
$ plotting_scripts/metaphlan2graphlan.py output/merged_abundance_table.txt --
tree_file tmp/merged.tree.txt --annot_file tmp/merged.annot.txt
$ graphlan_annotate.py --annot tmp/merged.annot.txt tmp/merged.tree.txt tmp/merged.xml
$ graphlan.py --dpi 200 tmp/merged.xml output_images/merged.png
```

Step 8.

As an example of manually and script-based modification of the "annotation" file provided in [this file](#), we can produce comparisons between classes. The resulting GraPhlAn image is shown below:

A: Solobacterium
B: Catonella
C: Oribacterium
D: Veillonella
E: Megamonas
F: Streptococcus
G: Prevotella
H: Atopobium
I: Butyrivibrio
J: Bifidobacterium
K: Campylobacter



Step 9.

Notice also that MetaPhlAn output can be exported to [Krona](#) (another popular visualization tool) using the script "metaphlan2krona.py" in the "conversion_scripts" folder downloaded as part of the MetaPhlAn package.