Metabolic profiling with HUMAnN

Curtis Huttenhower

Abstract

In this protocol, we perform a metabolic profiling of metagenomic datasets using [HUMAnN](#) applied to the 20 samples taxonomically profiled in the other protocols contained in this collection.

Once the metabolic profile has been generated, the following protocols can be performed:

- MetaPhlAn output merge and visualizations
- GraPhlAn visualization of single and multiple samples
- Taxonomic biomarker discovery with LEfSe

Guidelines

Metabolic profiles of the 20 HMP samples with HUMAnN can be generated with the following steps:

- Perform a translated search (using [blastx](#) or [usearch](#)) against the KEGG DB. Since HUMAnN's development KEGG has become commercial, we are currently developing support for other data sources.
- Place the translated BLAST results using KEGG gene identifiers in the `input` directory (optionally can be gzipped or bzipped, several other formats can be enabled editing the settings in the `SConstruct` file).
- Run the `scons` command, optionally parallelizing multiple analyses using the `-j` flag. Results will be placed in the "output" directory.

We provide the HUMAnN output for users who want to perform the downstream analysis pipeline but avoid the computational intensive steps above. The 20 samples profiled with HUMAnN are available [here](#).

Using HUMAnN output, you can perform the metabolic counterpart of the taxonomic pipeline presented in this tutorial. With a table of metabolomic abundances, the previous protocols can be performed with little to no modification. We report below the first command to obtain the merged table of metabolic abundances.

```
$ mkdir output
$ utils/merge_metaphlan_tables.py humann_profiling/*.txt >
output/merged_humann_abundance_table.txt
```

Before start

REQUIREMENTS: [HUMAnN](#), [scons](#), the KEGG protein DB. HUMAnN can be obtained using [Mercurial](#): `hg clone ssh://hg@bitbucket.org/chuttenh/humann` or using the direct links to the [zip](#), [gz](#), or [bz2](#) archives.

Protocol

Step 1.
Perform a translated search (using [blastx](#) or [usearch](#)) against the KEGG DB.

> ✚ NOTES
> **Bahar Sayoldin** 07 Dec 2015
> Since HUMAnN's development KEGG has become commercial, we are currently developing support for other data sources.

Step 2.
Place the translated BLAST results using KEGG gene identifiers in the `input` directory.

> ✚ NOTES
> **Bahar Sayoldin** 07 Dec 2015
> Optionally can be gzipped or bzipped, several other formats can be enabled editing the settings in the `SConstruct file`.

Step 3.
Run the `scons` command, optionally parallelizing multiple analyses using the `-j` flag. Results will be placed in the "output" directory.

Step 4.
We provide the HUMAnN output for users who want to perform the downstream analysis pipeline but avoid the computational intensive steps above. The 20 samples profiled with HUMAnN are available [here](#).

Step 5.
Using HUMAnN output, you can perform the metabolic counterpart of the taxonomic pipeline presented in this tutorial. We report below the first command to obtain the merged table of metabolic abundances.

> cmd COMMAND
```
$ mkdir output
$ utils/merge_metaphlan_tables.py humann_profiling/*.txt > output/merged_humann_abundance_table.txt
```