

# Finding the core vs flexible microbiome

Bonnie Hurwitz

## Abstract

In a study comparing obese versus lean twins, Turnbaugh and colleagues found that there were "shared microbial genes among sampled individuals, comprising an extensive, identifiable 'core microbiome' at the gene, rather than at the organismal lineage level".

This means that despite having differences in bacterial composition, the "core" functional repertoire of microbiome remains the same in obese versus lean twins.

In this tutorial, we are going to explore these functional differences in two lean twins (TS1 and TS2) and two obese twins (TS19 and TS20) from this study.

Turnbaugh, Peter J., et al. "A core gut microbiome in obese and lean twins." *nature* 457.7228 (2009): 480-484.

**Citation:** Bonnie Hurwitz Finding the core vs flexible microbiome. **protocols.io**

[dx.doi.org/10.17504/protocols.io.d2y8fv](https://dx.doi.org/10.17504/protocols.io.d2y8fv)

**Published:** 20 Oct 2015

## Guidelines

Read the paper here:

<http://www.nature.com/nature/journal/v457/n7228/abs/nature07540.html>

## Protocol

### Step 1.

Login to iPlant

### Step 2.

Look for the obese lean data files in the community data directory:

/iplant/home/shared/imicrobe/class/obese\_lean

SRR029687 - lean T1

SRR029692 - lean T2

SRR029686 -obese T19

SRR029688 - obese T20

You will use the fasta files in this directory for the next step.

### Step 3.

Click on the Apps button, navigate to experimental->imicrobe->uproc

Click on uproc

Click on "Required Args"

Use the Browse function to navigate to

" /iplant/home/shared/imicrobe/class/obese\_lean/SRR029687.fasta" under "Sequence file"

Run uproc with default parameters on the fasta file

Repeat for all of the other fasta files in the obese lean dataset

#### **Step 4.**

Login to the HPC

#### **NOTES**

**Bonnie Hurwitz** 26 Oct 2015

Some people are having issues with running the sra toolkit on the hpc. You can download these files from iPlant too:

```
% module load irods
```

```
% icd /iplant/home/shared/imicrobe/class
```

```
% iget -r obese_lean
```

#### **Step 5.**

Load irods on the hpc

```
% module load irods
```

If this is the first time you have used irods, you will need to run:

```
% iinit
```

use the configurations:

```
host-> data.iplantcollaborative.org; port -> 1247; zone -> iplant; user-> your iplant user; login -> your iplant login
```

#### **Step 6.**

Copy your files from the uproc runs to iPlant and rename them by the file\_name.uproc

```
% mkdir obese_lean
```

```
% cd obese_lean
```

```
% icd /iplant/home/your_user/analyses/uproc_SRR029687-your-analysis
```

```
% iget uproc-out
```

```
% mv uproc-out SRR029687.uproc
```

Repeat for each fasta file

#### **Step 7.**

Write a program that finds all kegg ids that are "core" to each file (in T1, T2, T19, and T20) and have at least 50 read matches.

The uproc output files have the following:

```
<kegg_id>,<read count>
```

"core" means that the kegg id is present in each of the uproc files

Program name: get\_core.pl

Output: should be a list of kegg ids

#### **Step 8.**

Write a program that finds all kegg ids that are "variable" in the files (in T1, T2, T19, and T20) and have at least 50 read matches.

The uproc output files have the following:

```
<kegg_id>,<read count>
```

"variable" means that the kegg id is present in 1-3 uproc files, but not all 4.

Program name: get\_variable.pl

Output: should be a list of kegg ids

**Step 9.**

Combine the two programs into a single program: `get_core_and_variable.pl`

**Step 10.**

Submit your scripts and results to github under the folder `abe487/obese_lean`

**Step 11.**

Write up a discussion in the forum on what you think "core" and "variable" means based on the article. What are we trying to find out when we do this?