

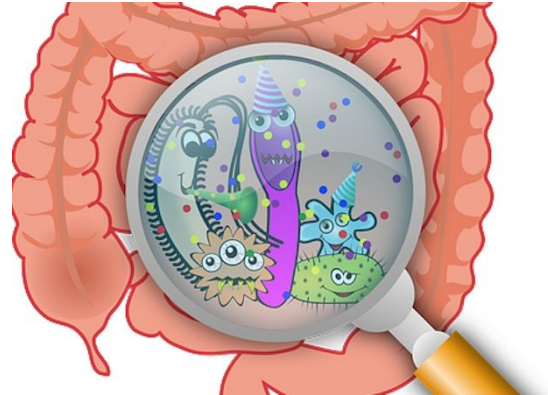
Tutorial #9: Tax4fun



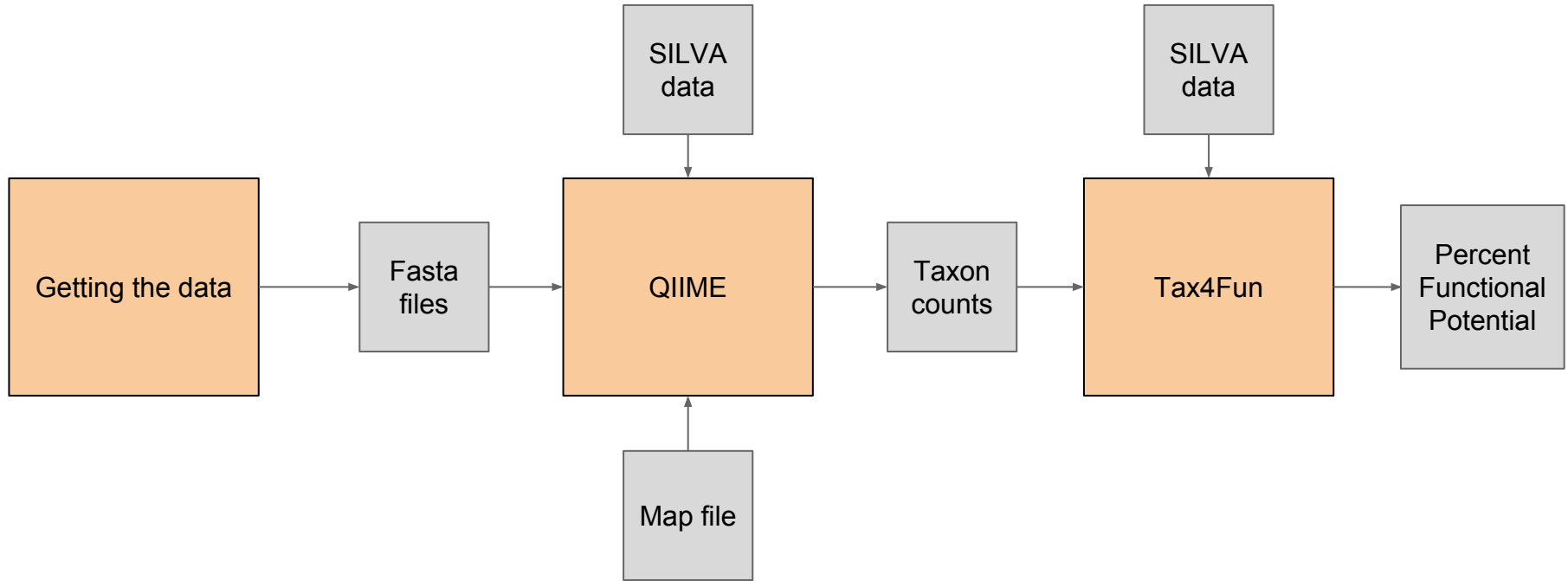
Sean Stevens & Nicholas Falkowski

Background - Gut microbiome

- Contains thousands of different micro organisms which each perform some function
- Could be bacteria, archaea, viruses, protists, or other forms of micro-organisms
- By classifying these micro-organisms we can attempt to determine their function in the gut using tax4fun



Overview



Prerequisites

1. Install Qiime

- Proteus -- module add
- Simple install <http://qiime.org/install/install.html>
- Full install <https://github.com/qiime/qiime-deploy>

2. Download Tax4Fun

- http://tax4fun.gobics.de/Tax4Fun/Tax4Fun_0.3.1.tar.gz

3. Download SILVA Reference data and database

- <http://tax4fun.gobics.de/Tax4Fun/ReferenceData/SILVA123.zip>
- http://tax4fun.gobics.de/SilvaSSURef_123_NR.zip

Getting the data

Download amplicon sequence data from MG-RAST

- Using the website
 - Search for project
 - Filter by sequence type “amplicon”
 - Use download page to get file from the dereplication step
- Using MG-RAST-Tools (<https://github.com/MG-RAST/MG-RAST-Tools>)
 - Get the mgm ids for your samples (i.e. mgm4539180.3, mgm4539181.3)
 - Use mg-download.py with file parameter “150.1” to get the de-replicated pipeline file

```
# download mgm
while read p; do
    mg-download.py --metagenome $p --file 150.1 --dir download
done <mgm
```

Qiime Outline

1. Create map text file for downloaded samples
2. Add qiime labels to downloaded fasta files
 - `add_qiime_labels.py`
3. Pick OTUs using closed reference
 - `pick_otus.py`
4. Obtain reference sequences for each OTU
 - `pick_rep_set.py`
5. Retrieve taxonomic information for each OTU
 - `parallel_assign_taxonomy_blast.py`
6. Generate OTU table
 - `make_otu_table.py`

Qiime Map

- Tab delimited file
- Mandatory headers
 - SampleID
 - BarcodeSequence
 - LinkerPrimerSequence
 - InputFileName
 - Description
- Only need SampleID and InputFileName for the next step

#SampleID	BarcodeSequence	LinkerPrimerSequence	InputFileName	Description
mgm4539180.3			mgm4539180.3.150.dereplication.passed.fna	
mgm4539181.3			mgm4539181.3.150.dereplication.passed.fna	

Qiime Pipeline - Bash script

```
add_qiime_labels.py -m map -i download -c InputFileName -o working
```

```
cd working
```

```
pick_otus.py -i seqs.fna -o picked_otus -r /home/nyid/tutorial9/SilvaSSURef_123_NR/  
SILVA_123_SSURef_Nr99_tax_silva.fasta -m uclust_ref --suppress_new_clusters --enable_rev_strand_match
```

```
pick_rep_set.py -i picked_otus/seqs_otus.txt -f seqs.fna -o rep_set.fna -m most_abundant
```

```
parallel_assign_taxonomy_blast.py -i rep_set.fna -b /home/nyid/tutorial9/SilvaSSURef_123_NR/  
SILVA_123_SSURef_Nr99_tax_silva.fasta -t /home/nyid/tutorial9/SilvaSSURef_123_NR/SILVA_123_SSURef_Nr99_tax_silva.tax  
-o rep_set_taxonomy -O 6 -U start_parallel_jobs.py
```

```
make_otu_table.py -i picked_otus/seqs_otus.txt -t rep_set_taxonomy/rep_set_tax_assignments.txt -o otu_table.biom
```


Tax4fun

- R package
- Can use on Proteus or local computer

Usage

1. Import Qiime “biom” or text file
2. Run Tax4Fun
3. View or export Tax4Fun profile

Tax4Fun - R script

```
library("Tax4Fun")
help(Tax4Fun)

# location of reference data downloaded from Tax4Fun website
refdata="F:/tutorial9/SILVA123"

# import the qiime biom file
b=importQIIMEBiomData("F:/tutorial9/working/otu_table.biom")

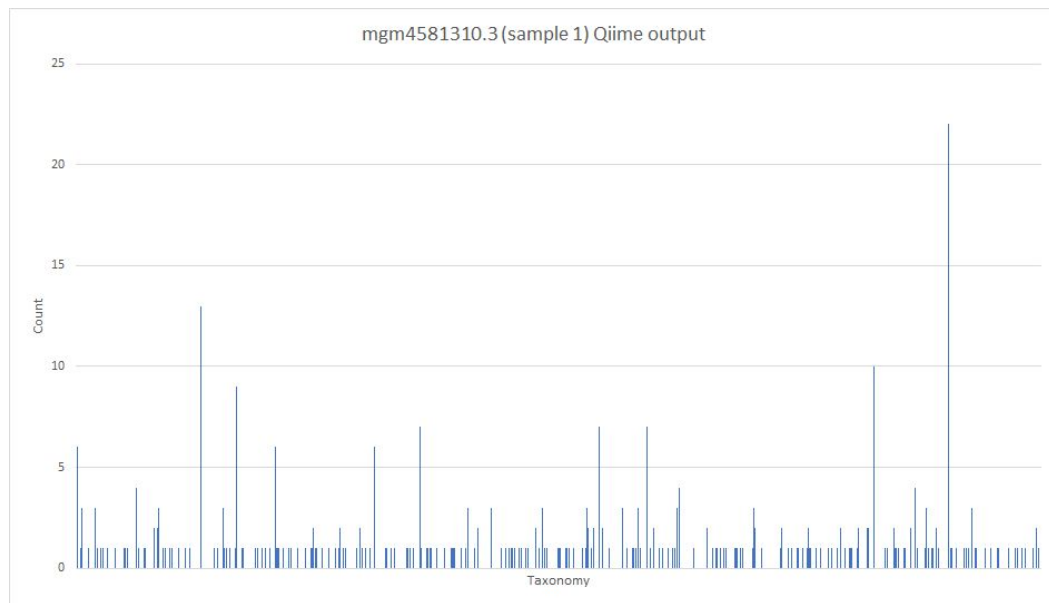
# run Tax4Fun
Tax4FunOutput=Tax4Fun(b, refdata)

# get output to a readable state
Tax4FunProfile <- Tax4FunOutput$Tax4FunProfile
Tax4FunProfile <- data.frame(t(Tax4FunOutput$Tax4FunProfile))
View(Tax4FunProfile)

# save output to disk
write.table(Tax4FunProfile, "F:/tutorial9/working/Tax4FunProfile.txt", sep="\t")
write.table(Tax4FunProfile, "F:/tutorial9/working/Tax4FunProfile.csv", sep=",")
```

Results - Qiime output

#OTU ID	mgm4581310.3	mgm4539187.3	mgm4581339.3	mgm4539184.3	mgm4539180.3	mgm4539186.3	mgm4539182.3	mgm4539181.3	mgm4539183.3	mgm4539185.3	taxonomy
EU775506.1.1389	6	1	1	0	0	0	0	0	0	0	0 Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Anaerostipes; uncultured bacterium
EU560798.1.1500	0	0	0	1	0	0	0	0	0	0	0 Bacteria; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Lactococcus; uncultured bacterium
HQ716186.1.1384	1	0	0	0	0	0	0	0	0	0	0 Bacteria; Tenericutes; Mollicutes; Mollicutes RF9; uncultured bacterium
AB506356.1.1529	3	0	0	0	0	1	1	1	6	0	0 Bacteria; Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminococcaceae NK4A214 group; uncultured bacterium
KF590558.1.1432	0	0	0	0	0	1	0	0	0	0	0 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; bacterium 28W313
DQ325948.1.1351	0	0	0	0	0	0	0	0	1	0	0 Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; [Eubacterium] hallii group; uncultured bacterium
AF371630.1.1443	0	0	1	0	0	0	0	0	0	0	0 Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Lachnospiraceae UCG-005; uncultured bacterium
X68176.1.1513	0	0	0	0	0	0	1	0	0	0	0 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae 1; Clostridium sensu stricto 1; uncultured bacterium



Results - Tax4Fun output

Functional profile	mgm4581310.3	mgm4539187.3	mgm4581339.3	mgm4539184.3	mgm4539180.3	mgm4539186.3	mgm4539182.3	mgm4539181.3	mgm4539183.3	mgm4539185.3
K00001; alcohol dehydrogenase [EC:1.1.1.1]	0.00049348	0.000471269	0.000440149	0.000520761	0.000568664	0.000555686	0.000567681	0.000557346	0.00049059	0.000401698
K00002; alcohol dehydrogenase (NADP+) [EC:1.1.1.2]	1.09E-05	1.37E-05	9.59E-06	1.37E-05	1.36E-05	1.40E-05	1.53E-05	1.33E-05	9.92E-06	1.11E-05
K00003; homoserine dehydrogenase [EC:1.1.1.3]	0.000855445	0.000904215	0.000912132	0.00083887	0.000787225	0.000862994	0.000832134	0.000853293	0.000872523	0.000904748
K00004; (R,R)-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4 1.1.1.303]	9.22E-06	3.38E-05	5.77E-06	3.43E-05	4.49E-05	3.20E-05	4.63E-05	3.88E-05	2.22E-05	1.78E-05
K00005; glycerol dehydrogenase [EC:1.1.1.6]	0.000177755	0.000189969	0.000181795	0.000262034	0.000201232	0.000298173	0.000203809	0.000237543	0.000327346	0.000207471
K00007; D-arabinitol 4-dehydrogenase [EC:1.1.1.11]	1.96E-06	1.70E-06	5.85E-07	2.07E-06	1.18E-05	4.84E-06	3.85E-06	2.43E-06	8.87E-07	1.17E-06
K00008; L-iditol 2-dehydrogenase [EC:1.1.1.14]	0.000146196	0.000244723	0.000143256	0.000260155	0.00027731	0.000265711	0.000308191	0.000268543	0.000192961	0.000203481
K00009; mannitol-1-phosphate 5-dehydrogenase [EC:1.1.1.17]	6.39E-05	0.000153866	4.65E-05	0.000239458	0.000235492	0.000221051	0.00023523	0.000206181	0.000129765	0.000151104

