

picrust

November 16, 2017

0.1 Picrust: metagenomic inference from 16S data

0.1.1 if you have 16S sequencing data, you can 'infer' whole genomes

0.1.2 by using picrust

Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Langille, M. G.I.; Zaneveld, J.; Caporaso, J. G.; McDonald, D.; Knights, D.; Reyes, J.; Clemente, J. C.; Burkepile, D. E.; Vega Thurber, R. L.; Knight, R.; Beiko, R. G.; and Huttenhower, C. **Nature Biotechnology**, 1-10. 8 2013.

```
In [3]: # make sure you are in 'metag' folder
        pwd
```

```
/Users/husenzhang/Documents/GitHub/FAES_metagenomics
```

```
In [7]: # change '.' to '_' in seqs.fa
        # necessary for pick_closed_reference_otus script below
        sed 's/\./_/' pre-computed-results/seqs.fa \
            > pre-computed-results/seqs_picrust.fa
```

```
In [12]: # check if "." -> "_" ? yes success
         head -1 pre-computed-results/seqs.fa
         head -1 pre-computed-results/seqs_picrust.fa
```

```
>S1.2
```

```
>S1_2
```

0.1.3 type: qiime

```
In [9]: # Picrust only works with closed OTU pick
        # only OTUs in greengene databases are retained
        # novel OTUs discarded. Necessary for genome imputation
        # Closed OTU pick explanation:
        # http://qiime.org/scripts/pick_closed_reference_otus.html
        pick_closed_reference_otus.py \
            -i pre-computed-results/seqs_picrust.fa \
            -o picrust_biom -f
```

```
In [13]: # check the new otu_table
         biom summarize-table -i picrust_biom/otu_table.biom
```

```
Num samples: 6
Num observations: 372
Total count: 11,414
Table density (fraction of non-zero values): 0.395
```

```
Counts/sample summary:
Min: 1,784.000
Max: 2,022.000
Median: 1,915.000
Mean: 1,902.333
Std. dev.: 78.191
Sample Metadata Categories: None provided
Observation Metadata Categories: taxonomy
```

```
Counts/sample detail:
```

```
S1: 1,784.000
S3: 1,828.000
S4: 1,904.000
S6: 1,926.000
S2: 1,950.000
S5: 2,022.000
```

Num observations: 372 – even bigger than 76 we saw on Tuesday? Don’t worry - 76 OTUs are picked by usearch, which tends to have ‘tighter’ clusters. let’s just move on...

0.1.4 important: exit qiime

0.1.5 picrust step1: normalize genome copies

```
In [ ]: # E. coli genome has 6 copies of 16S
         # Thermotoga maritima has only 1.
         normalize_by_copy_number.py \
         -i picrust_biom/otu_table.biom \
         -o picrust_biom/norm.biom
```

0.1.6 Step 2: Predict Functions For Metagenome

```
In [ ]: # creates the final metagenome functional predictions.
         # It multiplies each normalized OTU abundance
         # by each predicted functional trait abundance
         # to produce a table of functions (rows) by samples (columns).

         # Input is the normalized OTU table created by
         # normalize_by_copy_number.py.
         # Output is in biom format by default:
```

```

predict_metagenomes.py \
    -i picrust_biom/norm.biom \
    -o picrust_biom/predictions.biom

```

0.1.7 step 3: collapse genes into KEGG pathways

```

In [ ]: categorize_by_function.py \
    -i picrust_biom/predictions.biom \
    -c KEGG_Pathways \
    -l 3 \
    -o picrust_biom/L3.biom

```

convert L3.biom to L3.txt - easier to read

```

In [ ]: biom convert \
    --to-tsv \
    --header-key KEGG_Pathways \
    -i picrust_biom/L3.biom \
    -o picrust_biom/L3.txt

```

```

In [ ]: # view the metagenomic pathways
sed 1d picrust_biom/L3.txt | head -4 | cut -f2- | less -S

```

```

In [ ]: S1      S2      S3      S4      S5      S6      KEGG_Pathways
0.0      0.0      4.0      6.0      0.0      0.0      Metabolism; Xenobiotics Bio
29360.0  33367.0  46563.0  35558.0  41137.0  34115.0  Environmental Information B
0.0      0.0      0.0      0.0      0.0      0.0      Cellular Processes; Cell Co
887.0    1013.0    944.0    1115.0    899.0    895.0    Organismal Systems; Endocri

```

very carefully type the following command

```

In [ ]: sed -i 's/KEGG_Pathways/taxonomy/' picrust_biom/L3.txt

```

0.1.8 getting the pathway table ready for plotting

```

In [ ]: biom convert --to-hdf5 \
    --process-obs-metadata taxonomy \
    -o ~/Desktop/L3.hdf5 \
    -i picrust_biom/L3.txt \
    --table-type "Pathway table"

```

0.1.9 important: type: qiime

now we plot the pathways using the same metadata 'map.txt'

we pass on '-nonphylogenetic_diversity' because we do not

have a tree file for pathways - we only have a tree for 16S.

```
In [ ]: # plotting the pathways
#
core_diversity_analyses.py -i ~/Desktop/L3.hdf5 \
-o picrust_core_diversity \
-m data/map.txt \
-e 1000 \
--nonphylogenetic_diversity
```

warning like these are fine /usr/lib/pymodules/python2.7/matplotlib/collections.py:548:
FutureWarning: elementwise comparison failed; returning scalar instead, but in the future will
perform elementwise comparison if self._edgecolors == 'face':
cd picrust_core_diversity ls -lh

0.1.10 double click 'index.html'

here you will find various results

take some time (15 min) to explore these results

```
In [ ]: hzhang@bl8vbox[picrust_core_diversity] ll

drwxrwxr-x 4 hzhang hzhang  4096 Nov 16 18:17 arare_max1000
drwxrwxr-x 3 hzhang hzhang  4096 Nov 16 18:16 bdiv_even1000
-rw-rw-r-- 1 hzhang hzhang   427 Nov 16 18:16 biom_table_summary.txt
-rw-rw-r-- 1 hzhang hzhang  2209 Nov 16 18:17 index.html
-rw-rw-r-- 1 hzhang hzhang  5275 Nov 16 18:17 log_20171116181639.txt
-rw-rw-r-- 1 hzhang hzhang 16019 Nov 16 18:16 table_even1000.biom.gz
-rw-rw-r-- 1 hzhang hzhang 29103 Nov 16 18:16 table_mc1000.biom.gz
drwxrwxr-x 3 hzhang hzhang  4096 Nov 16 18:17 taxa_plots
```

```
In [ ]:
```