H3ABioNet
Pan African Bioinformatics Network for H3Africa

Data analysis of 16S rRNA amplicons

Computational Metagenomics Workshop University of Mauritius

Practical

December 2014

Exercise options

1) We will be going through a 16S pipeline using QIIME and 454 data

- Create a working directory e.g.:

```
$ mkdir -p /home/trainee/practical/day4/
$ cd  /home/trainee/practical/day4/
```

- Download the dataset and uncompress

```
$ wget http://172.22.64.219:8080/day4/dataset1.tgz
$ tar -xzvf dataset1.tgz
```

- For each command you can get a detailed list of settings by running

```
$ split_libraries.py -h
```

or

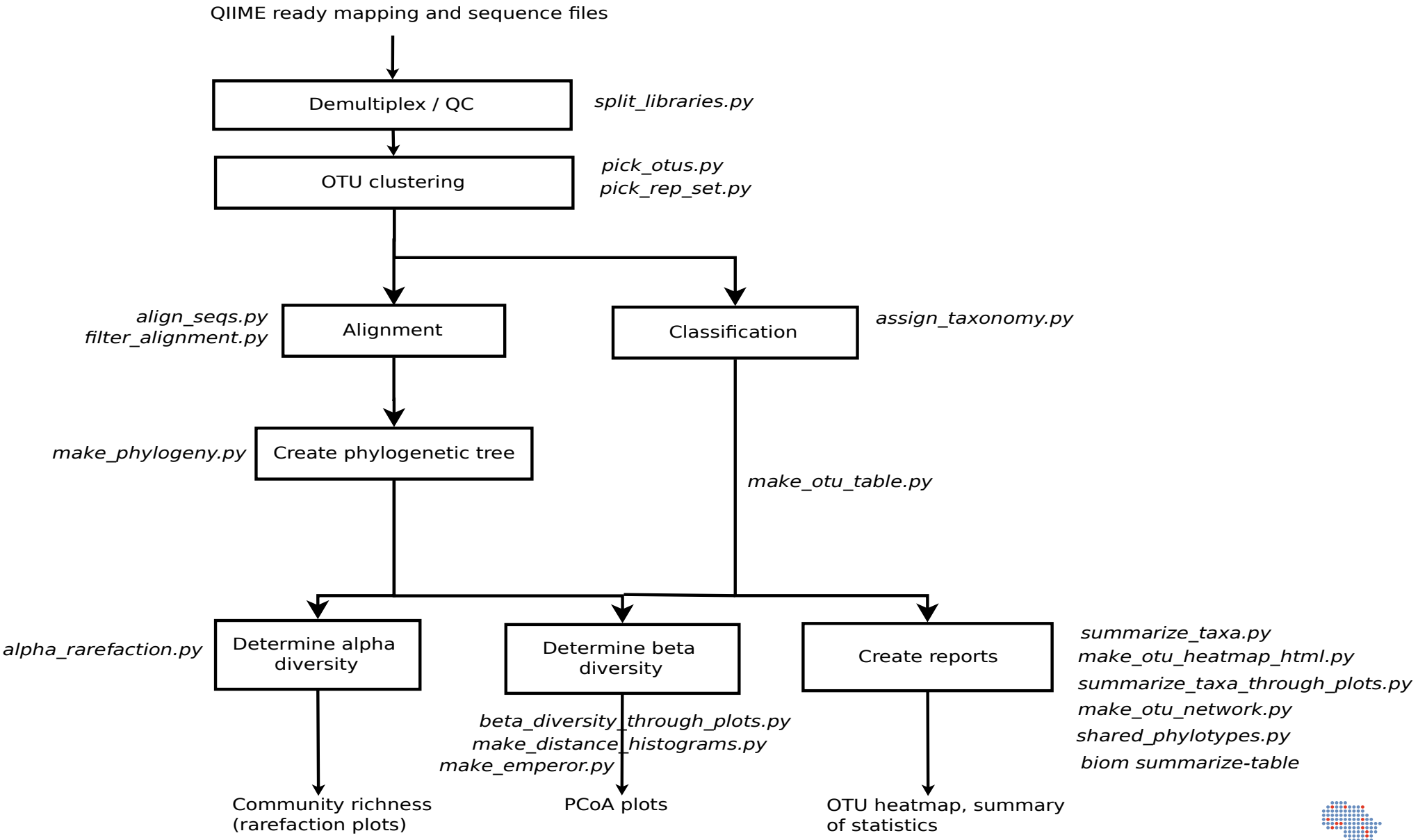
go to the documentation on your machine:

`~/16SrRNADiversityAnalysis/docs/QIIME/index.html`

2) If you are familiar with the analysis please go ahead and tackle the practice dataset on the H3ABioNet 16S SOP page (You do not need to use the QIIME workflow for the analysis). A copy of the dataset are on the local network and instructions are downloadable here:

- http://172.22.64.219:8080/day4/dataset2.tgz
- http://172.22.64.219:8080/day4/16SrRNAdiversityanalysis-Questions.pdf

16S analysis in QIIME (1)



16S analysis in QIIME (2)

- Check if everything is configured

```
$ print_qiime_config.py
```

- *Look at the mapping file (Fasting_Map.txt)*

- Convert sff to fasta and qual

```
$ process_sff.py -i input/Fasting_Example.sff -o output
```

- Run fastqc to check quality but we need to be convert fastq first

```
$ convert_fastaqual_fastq.py -c fastaqual_to_fastq -f output/Fasting_Example.fna -q output/Fasting_Example.qual -o output
$ mkdir output/raw.fastqc
$ fastqc output/Fasting_Example.fastq -o output/raw.fastqc
```

- *Look at the fastqc output*

- Demultiplex reads and do internal QIIME quality control. (*Look at the split_library_log.txt*)

```
$ split_libraries.py -m metadata/Fasting_Map.txt -f output/Fasting_Example.fna -q output/Fasting_Example.qual -o output/split_libraries
```

16S analysis in QIIME (3)

- Pick OTUs (using uclust)

```
$ pick_otus.py -i output/split_libraries/seqs.fna -m uclust -o output/otus -s 0.97 -l
```

- Pick representative sequence from otus (picking method=most_abundant)

```
$ pick_rep_set.py -m most_abundant -f output/split_libraries/seqs.fna -i
output/otus/seqs_otus.txt -o output/otus/seqs.otus.mostabundant.repset.fna
```

- Assign taxonomy (using uclust)

```
$ assign_taxonomy.py -m uclust -r /home/qiime/qiime_software/gg_otus-13_8-
release/rep_set/97_otus.fasta -t /home/qiime/qiime_software/gg_otus-13_8-
release/taxonomy/97_otu_taxonomy.txt -i output/otus/seqs.otus.mostabundant.repset.fna
-o output/otus/seqs.otus.mostabundant.repset.uclust
```

- Make otu table

```
$ make_otu_table.py -i output/otus/seqs_otus.txt -t
output/otus/seqs.otus.mostabundant.repset.uclust/seqs.otus.mostabundant.repset_tax_assi
gnments.txt -o output/otus/otu_table.biom
```

- *Look at the OTU output generated in otus/ (representative sequence, OTU table and BIOM table)*

16S analysis in QIIME (4)

- Align with PyNAST

```
$ align_seqs.py -m pynast -i output/otus/seqs.otus.mostabundant.repset.fna -o
output/otus/pynast_aligned_seqs
```

- Filter alignments

```
$ filter_alignment.py -i
output/otus/pynast_aligned_seqs/seqs.otus.mostabundant.repset_aligned.fasta -o
output/otus/pynast_aligned_filtered_seqs
```

- Make phylogenetic tree (using fasttree)

```
$ make_phylogeny.py -t fasttree -i
output/otus/pynast_aligned_filtered_seqs/seqs.otus.mostabundant.repset_aligned_pfiltered.fa
sta -o output/otus/rep_set.tre
```

- *Look at the alignment and tree generation output generated in otus/*

- Get summaries from OTU table *(Look at the output)*

```
$ biom summarize-table -i output/otus/otu_table.biom -o output/otus/otu_table_summary.txt
```

```
$ biom summarize-table --qualitative -i output/otus/otu_table.biom -o
output/otus/otu_table_summary.qualitative.txt
```

16S analysis in QIIME (5)

- Create OTU heatmap *(Look at the output)*

```
$ rm -rf output/otus/heatmap; make_otu_heatmap_html.py -i
output/otus/otu_table.biom -o output/otus/heatmap
```

- Create taxonomy summary *(Look at the output, area and bar charts)*

```
$ rm -rf output/otus/taxa_summary; summarize_taxa_through_plots.py -m
metadata/Fasting_Map.txt -i output/otus/otu_table.biom -o
output/otus/taxa_summary
```

- Create rarefaction plots *(Look at the output)*

```
$ mkdir output/alpha_div; echo "alpha_diversity:metrics
berger_parker_d,brillouin_d,chao1,chao1_confidence,dominance,doubles,enspie,
equitability,esty_ci,fisher_alpha,gini_index,goods_coverage,heip_e,kempton_t
aylor_q,margalef,mcintosh_d,mcintosh_e,menhinick,michaelis_menten_fit,observ
ed_species,osd,simpson_reciprocal,robbins,shannon,simpson,simpson_e,singles,
strong,PD_whole_tree" > output/alpha_div/alpha_params.txt
```

```
$ alpha_rarefaction.py -m metadata/Fasting_Map.txt -i
output/otus/otu_table.biom -o output/alpha_div/rarefaction -p
output/alpha_div/alpha_params.txt -t output/otus/rep_set.tre
```

16S analysis in QIIME (6)

- Calculate beta diversities and make plots *(Look at the output)*

```
$ beta_diversity_through_plots.py -m metadata/Fasting_Map.txt -i
output/otus/otu_table.biom -o output/beta_div/ -t output/otus/rep_set.tre -e 146
```

- Calculate jackknifed beta diversities *(Look at the output)*

```
$ jackknifed_beta_diversity.py -m metadata/Fasting_Map.txt -i
output/otus/otu_table.biom -t output/otus/rep_set.tre -o output/jackknifed -e
110
```

- Make bootstrap tree *(Look at the output)*

```
$ make_bootstrapped_tree.py -m
output/jackknifed/unweighted_unifrac/upgma_cmp/master_tree.tre -s
output/jackknifed/unweighted_unifrac/upgma_cmp/jackknife_support.txt -o
output/jackknifed/unweighted_unifrac/upgma_cmp/jackknife_named_nodes.pdf
```

- Make bi-plots *(Look at the output)*

```
$ make_emperor.py -m metadata/Fasting_Map.txt -i
output/beta_div/unweighted_unifrac_pc.txt -t
output/otus/taxa_summary/otu_table_L3.txt --n_taxa_to_keep 5 -o output/3d_biplot
```