ECOGEO 'Omics Training: 3.2 Amplicon Analysis - QIIME

Version 2

Melissa Dsouza

Abstract

Commands for hands-on component of QIIME, which should be run from a local installation. Details can be found online at http://qiime.org/install/index.html.

The files used in this hands-on component are also available the virtual machine (details in 'Start Instructions'), but you will need to upload them to your Google Drive or Dropbox from within the VM in order to have them available for QIIME, which is installed locally.

Citation: Melissa Dsouza ECOGEO 'Omics Training: 3.2 Amplicon Analysis - QIIME. **protocols.io** dx.doi.org/10.17504/protocols.io.fjbbkin
Published: 18 Aug 2016

Guidelines

1. Filter low sequence count samples from table (minimum sequence count: 7500)

$ filter_samples_from_otu_table.py -i denovo_otu.biom -o core_diversity7500/table_mc7500.biom -n 7500

2. Rarify the OTU table to 7500 sequences/sample
$ single_rarefaction.py -i core_diversity7500/table_mc7500.biom -
o core_diversity7500s/table_even7500.biom -d 7500

3. Beta Diversity (weighted_unifrac)
$ beta_diversity.py -i core_diversity7500s/table_even7500.biom -
o core_diversity7500/bdiv_even7500/ --metrics weighted_unifrac -t ref_set.tre

4. Principal coordinates (weighted_unifrac)
$principal_coordinates.py -i core_diversity7500/bdiv_even7500//weighted_unifrac_dm.txt -
o core_diversity7500/bdiv_even7500//weighted_unifrac_pc.txt

5. Make emperor plots, weighted_unifrac)
$make_emperor.py -i core_diversity7500s/bdiv_even7500//weighted_unifrac_pc.txt -
o core_diversity7500/bdiv_even7500//weighted_unifrac_emperor_pcoa_plot/ -m map_file.txt

6. Alpha rarefaction
$ multiple_rarefactions.py -i core_diversity7500/table_mc7500.biom -m 10 -x 7500 -s 749 -

o core_diversity7500/arare_max7500//rarefaction/

7. Alpha diversity on rarefied OTU tables
$ alpha_diversity.py -i core_diversity7500/arare_max7500//rarefaction/ -
o core_diversity7500_1105noblanks/arare_max7500//alpha_div/ -t ref_set.tre

8. Collate alpha
$ collate_alpha.py -i core_diversity7500/arare_max7500//alpha_div/ -
o core_diversity7500_/arare_max7500//alpha_div_collated/

9. Rarefaction plot: All metrics
$ make_rarefaction_plots.py -i core_diversity7500/arare_max7500//
alpha_div_collated/ -m 150701_CAWSMF_1104.txt -
o core_diversity7500/arare_max7500//alpha_rarefaction_plots/

10. Summarize Taxonomy
$ summarize_taxa.py -i core_diversity7500s/taxa_plots/table_mc7500_sorted.biom -
o core_diversity7500/taxa_plots/

Before start

Before starting, please visit the ECOGEO website for more information on this "Introduction to Environmental 'Omics" training series. The site contains a pre-packaged virtual machine that can be downloaded and used to run all of the protocols in this protocols.io collection. In addition to the VM, the website contains video and presentations from our initial "Intro to Env 'Omics" workshop held at the Univ. of Hawai'i at Manoa on 25-26 Jul 2016.

Please email ‘ecogeo-join@earthcube.org’ to join the ECOGEO listserv for future updates.

Protocol

Step 1.

1. Checking mapping file format

cmd COMMAND
```
$ validate_mapping_file.py -m map_file.txt -o mapping_file_output
```

Step 2.

2.1 Join paired end reads

cmd COMMAND
```
$ join_paired_ends.py -f Undetermined_S0_L001_R1_001.fastq -
r Undetermined_S0_L001_R2_001.fastq -b Undetermined_S0_L001_I1_001.fastq -o reads
```

Step 3.

2.2 Depmultiplex & quality filter

cmd COMMAND

```
$ split_libraries_fastq.py -i fastqjoin.join.fastq -b fastqjoin.join_barcodes.fastq -
o split -m map_file.txt -q 29 --barcode_type 12 --store_demultiplexed_fastq
```

Step 4.

3.1 *De novo* OTU picking

cmd COMMAND

```
$ pick_de_novo_otus.py -i subs_seqs_q.fasta -o uclust_otus/
```

Step 5.

3.2 Closed-reference OTU picking

cmd COMMAND

```
$ pick_closed_reference_otus.py -i subs_seqs_q.fasta -
r $PWD/gg_13_8_otus/rep_set/97_otus.fasta -
t $PWD/gg_13_8_otus/taxonomy/97_otu_taxonomy.txt -o_ref_otus
```

Step 6.

3.3 Open-reference OTU picking

cmd COMMAND

```
$ pick_open_reference_otus.py -i subs_seqs_q.fasta -
r $PWD/gg_13_8_otus/rep_set/97_otus.fasta -o uclust_open_otu
```

Step 7.

BIOM files

cmd COMMAND

```
$ biom summarize-table -i uclust_otus/otu_table_even10.biom
$ biom summarize-table -i uclust_otus/otu_table_even10.biom --qualitative

$ biom add-metadata -i $PWD/otu_table_even10.biom -o denovo_otu.biom --sample-metadata-
fp map_file.txt --observation-metadata-
fp $PWD/uclust_assigned_taxonomy/subs_seqs_q_rep_set_tax_assignments.txt --observation-
header OTUID,taxonomy --sc-separated taxonomy

$ biom convert -i otu_table.biom -o otu_table.txt --table-type 'otutable' --header-
key taxonomy -b
```

Step 8.

More commands

cmd COMMAND

```
$ identify_chimeric_seqs.py -m ChimeraSlayer -i rep_set_aligned.fasta -a gold.fa -
o chimeric_seqs.txt

$ core_diversity_analyses.py -i denovo_otu.biom -o core_diversity -e 10 -m map_file.txt -
t ref_set.tre
```