Argonne National Laboratories Marker Gene Sequencing Analysis Workflow

1/1/2015-12/31/2015
16S rRNA Sequence Data Processing Protocol

Sequences (Read 1, Read 2 and Index) files were joined to increase specificity during processing using Qiime 1.8. After joining reads, samples were demultiplexed to associate sequences with each sample. These demultiplexed sequences were clustered into representative OTUs (operational taxonomic units) based on sequence similarity (set at 97%) using uclust_ref to bin sequences. To identify sequences, clusters were passed through Greengenes database. Sequences that did not hit towards this database at 97% sequence similarity were discarded. With closed reference otu generation serves as a filter for chimeric sequences, there was no need to check for chimeras using ChimeraSlayer. Since sequences were not aligned and filtered, phylogenetic trees were used from the Greengenes database.

#Neon Workflow - Closed Referenced 16S OTU Picking - 150211#

module load qiime/1.8

#Unzip Reads
gunzip -c rawseq/16S/Undetermined_S0_L001_I1_001.fastq.gz > rawseq/16S/barcodes.fastq &&
gunzip -c rawseq/16S/Undetermined_S0_L001_R1_001.fastq.gz > rawseq/16S/read1.fastq &&
gunzip -c rawseq/16S/Undetermined_S0_L001_R2_001.fastq.gz > rawseq/16S/read2.fastq

#Join Read 1 and 2
chmod a+x scripts/*

```
scripts/fastq-join rawseq/16S/read1.fastq rawseq/16S/read2.fastq -o
rawseq/16S/01_joined/out.%.fastq >
rawseq/16S/01_joined/out.stats.txt

#Join Barcodes to Paired Reads
scripts/fastq-barcode.pl rawseq/16S/barcodes.fastq
rawseq/16S/01_joined/out.join.fastq >
rawseq/16S/01_joined/out.barcodes.fastq

#Demultiplex Reads (16S joined, ITS single read)
split_libraries_fastq.py -i rawseq/16S/01_joined/out.join.fastq -b
rawseq/16S/01_joined/out.barcodes.fastq -m
mapping/neon_16s_mapping_file.txt -o demultiplex --
barcode_type=12 --max_barcode_errors=0

#Zip 16S Joined Sequences
tar -zcvf 16S_fastq.tar.gz rawseq/16S/01_joined/

#Split FNA file by sample
split_fasta_on_sample_ids.py -i seqs.fna -o 16S_sample_fasta

#Closed Reference OTU Picking in Parallel
pick_closed_reference_otus.py -i demultiplex_16S/seqs.fna -r
/gg_13_8_otus/rep_set/97_otus.fasta -o picked_closed_97 -p
ucrss_params2.txt -a -O 4

#Add Taxonomy to OTU table using Greengenes 13.8
biom add-metadata -i picked_closed_97/16S/otu_table.biom --
observation-metadata-fp
/gg_13_8_otus/taxonomy/97_otu_taxonomy.txt -o
picked_closed_97/16S/otu_table_w_tax.biom --sc-separated
taxonomy --observation-header OTUID,taxonomy
```

ITS Sequence Data Processing Protocol

Sequences from Read 1 and Index files were demultiplexed to associate sequences with each sample using Qime 1.8. Reads were not joined as ITS markers are variable in length and cannot be joined with a set confidence. These demultiplexed sequences were clustered into representative OTUs (operational taxonomic units) based on sequence similarity (set at 97%) using uclust_ref to bin sequences. To identify sequences, clusters were passed through UNITE database. Sequences that did not hit towards this database at 97% sequence similarity were discarded. With closed reference otu generation serves as a filter for chimeric sequences, there was no need to check for chimeras using ChimeraSlayer. There is not a clear consensus on phylogenetic relationship as ITS markers are too variable to build a consensus. Downstream analyses should be generated using non-phylogenetic relationships i.e. Bray-Curtis, etc.

#Neon Workflow - Closed Referenced ITS OTU Picking - 150211#

module load qiime/1.8

#Unzip Reads
gunzip -c rawseq/ITS/Undetermined_S0_L001_I1_001.fastq.gz > rawseq/16S/barcodes.fastq &&
gunzip -c rawseq/ITS/Undetermined_S0_L001_R1_001.fastq.gz > rawseq/16S/read1.fastq &&
gunzip -c rawseq/ITS/Undetermined_S0_L001_R2_001.fastq.gz > rawseq/16S/read2.fastq

#Demultiplex Reads
split_libraries_fastq.py -i rawseq/ITS/Undetermined_S0_L001_R1_001.fastq.gz -b rawseq/ITS/Undetermined_S0_L001_I1_001.fastq.gz -m mapping/neon_its_mapping_file.txt -o demultiplex_ITS --barcode_type=12 --max_barcode_errors=0 --rev_comp_mapping_barcodes

#Split FNA file by sample
split_fasta_on_sample_ids.py -i seqs.fna -o ITS_sample_fasta

#Closed Reference OTU Picking in Parallel
pick_closed_reference_otus.py -i demultiplex_ITS/seqs.fna -r /project/gilbertjack/databases/its_12_11_otus/rep_set/97_otus.fasta -o picked_closed_its_97 -p ucrss_params_its.txt -a -O 4

#Add Taxonomy to OTU table using UNITE 12.11
biom add-metadata -i picked_closed_97/ITS/otu_table.biom --observation-/its_12_11_otus/taxonomy/97_otu_taxonomy.txt -o picked_closed_97/ITS/otu_table_w_tax.biom --sc-separated taxonomy --observation-header OTUID,taxonomy