Illumina (post-MR DNA) processing pipeline : mothur

Kylie Langlois[1]

[1]State University of New York at Stony Brook

dx.doi.org/10.17504/protocols.io.hnvb5e6

Collier Lab

Kylie Langlois

ABSTRACT

This pipeline is for datasets generated from Illumina (2x300) sequencing by MR DNA. MR DNA runs sequences through a proprietary pipeline for quality control of sequences and OTU clustering. This pipeline begins with the OTU abundnace table and utilizes the representative sequence set generated by the MR DNA pipeline. **Through trial and error, I have discovered the easiest and best practice for my lab is to use a combination of QIIME (full installation) and mothur. This protocol is the mothur portion.** This protocol does not include data visualization.

As of 4/5/17, the following versions were used: QIIME v.1.9.1-amd64.vdi and mothur v.1.39.5

PROTOCOL STATUS

Working

Working for mothur

Before you start this protocol

1 **Begin with Illumina (post-MR DNA) processing pipeline : QIIME** steps 1-7 to rarefy the OTU abundance table. Complete step 12 to convert the rarefied OTU abundance table to a text file.

Install the executable version of mothur in Windows

2 "Downloading an executable version of mothur is the easiest and fastest way to get started." -mothur download page

The best way to use mothur is in Windows or Mac where you can download an executable version of mothur that runs through a terminal window. All scripts are written in C++, a wrapped language, so you can execute a command within the mothur terminal just like you would execute a QIIME command in the QIIME terminal window. Here is the latest release of mothur on github. Follow instructions of the installation page. If you download the executable version of mothur, it will simply show up as an icon that opens to a terminal window when double-clicked. That is mothur! It is simpler than it appears.

Assign taxonomy

3 To compare the taxonomy assigned by the sequencing facility (MR DNA in our case) through a proprietary pipeline and curated database, you can use the classify.seqs command in mothur.

COMMAND

```
classify.seqs(fasta=CCWT_16S_Jan_Feb_Apr_rep_set.fa, template=gg_13_8_99.fasta, taxonomy=gg_13_8_99.tax)
```

fasta = is your unaligned representative sequences (1 OTU = 1 sequence)
template = can be either Greengenes or SILVA database template, unaligned
taxonomy = needs to be from the same database as the template file