**Microbiome Processing Pipeline**

**Manual**

**Author:**

Niyati Uppadhyaya

Anirban Bhaduri

Contents

[1. Background 3](#_Toc517428403)

[2. Approach 3](#_Toc517428404)

[3. Requirements and Installation 4](#_Toc517428405)

[4. Content and Processing 4](#_Toc517428406)

[4.1 Sample data and Execution 5](#_Toc517428407)

[5. Reference 5](#_Toc517428408)

# Background

The Microbiome Processing Pipeline aims to process 16S RNA/DNA sequence data from different platforms. It considers raw sequenced data and reports out OTU (Operational Taxonomy Units) tables. The pipeline is customizable and options to use two different processing methods are provided. Further mapping of the input sequence onto various 16S rRNA databases is possible. It should be noted that considering potential liability owing to license requirements, end users are requested to obtain the individual methods and databases.

# Approach

The pipeline processes input files based on the below flow diagram. Based on selection, the user could process an input file using the QIIME 1or the MOTHUR2[[1]](#endnote-1) approach. Choice of the database for assigning taxonomy rests with the user.

**Figure1:** Overview of processing pipeline

**Input Data** (FASTQ files)

**Clustering / OTU Picking** (OTU assignment commands which are clusters of different algorithm are responsible for forming these taxonomic tables)

**Filtering** (In QIIME; Phred score is used and in MOTHUR; FASTQ files are split first and duplicate sequences are removed through unique command)

**Processing** (MOTHUR has 145 commands, many of these commands implement various methods to do the same thing while generating the same output formats and in QIIME, various python functions are used for processing)

# Requirements and Installation

The Microbiome Processing Pipeline is developed in python and is executable across all operating systems having a python interpreter. The pipeline was tested on Intel(R) Core i3 processor supported on Microsoft Windows 10 desktop, 64-bit platform. It is advised that the user considers a configuration equivalent or higher.

The Microbiome Processing Pipeline provides an option of processing the data using 2 different methods, QIIME and MOTHUR. To plug them into the pipeline, user is required to download the methods from the below mentioned cites and comply with the license requirements.

- MOTHUR : <http://www.mothur.org>

- QIIME : <http://qiime.org/install/install.html>

The different methods could be functionalized by referencing the respective installation instructions. Please make sure that the software’s are properly installed and configured in the system.

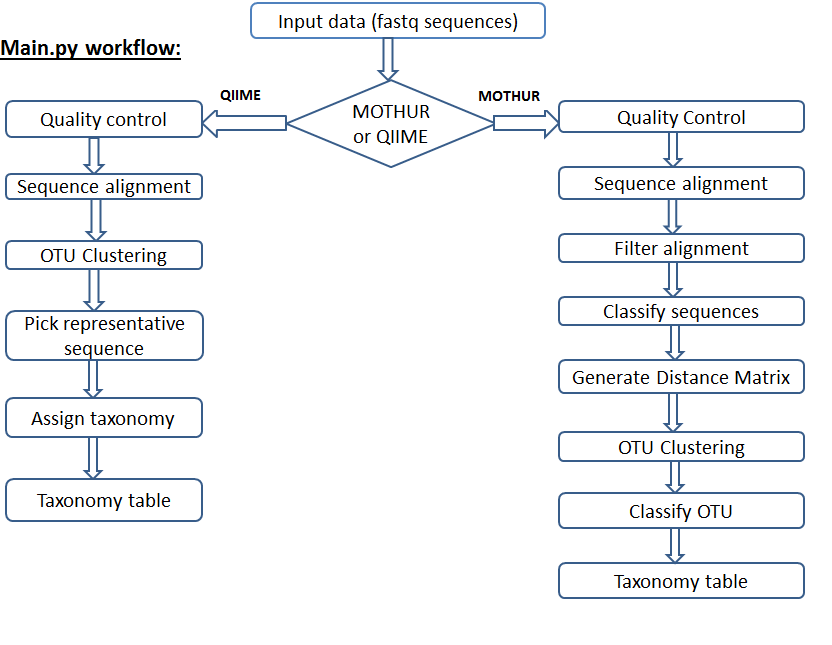
The Microbiome Processing Pipeline is enabled to process multiple reference sequence databases. Standard databases such as Greengenes3, Silva4 and RDP5 database could be plugged in. To plug in users would require the taxonomy information file and FASTA format sequence information.

# Content and Processing

The Microbiome Processing Pipeline comprises of 3 key python script. They are ‘Main.py’, ‘Qiime.py’ and ‘Mothur.py’. Main.py file is the user interfacing script and based in the inputs provided by the users forwards the information to either Qiime.py or Mothur.py. Based on the user inputs Qiime.py script performs analysis using QIIME method. Mothur.py analyzes data using MOTHUR approach for the assessment.

Input to the pipeline is a FASTQ file. To customize the processing, advanced users may set values to specific parameters of the pipeline. The current pipeline is customized to process single end files.

A high-level flow diagram of the pipeline in provided below.



## Sample data and Execution

A sample input data embedded below.

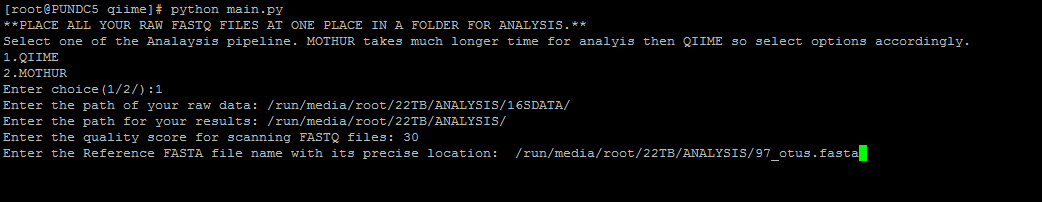
01\_20.fastq

Place this file in folder in named 16SDATA in a particular location.

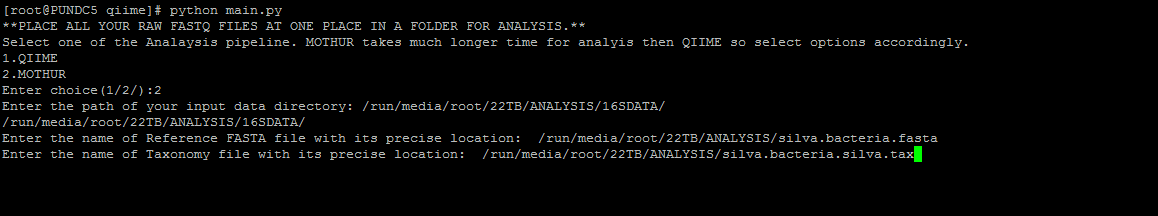
To execute the pipeline please follow the below mentioned steps:

Execute main script as: python main.py

The script will run and execute as follows.



The above figure represents the steps and the information provided by user when QIIME is considered for analysis purpose.



The figure depicts the steps to be followed when MOTHUR is considered for analysis.

# Reference

1. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–6 (2010).

2. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–41 (2009).

3. Al-Hebshi, N. N., Nasher, A. T., Idris, A. M. & Chen, T. Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. *J. Oral Microbiol.* **7,** 28934 (2015).

4. Pfeiffer, S. *et al.* Improved group-specific primers based on the full SILVA 16S rRNA gene reference database. *Environ. Microbiol.* **16,** 2389–407 (2014).

5. Bacci, G. *et al.* Evaluation of the Performances of Ribosomal Database Project (RDP) Classifier for Taxonomic Assignment of 16S rRNA Metabarcoding Sequences Generated from Illumina-Solexa NGS. *J. genomics* **3,** 36–9 (2015).

1. [↑](#endnote-ref-1)