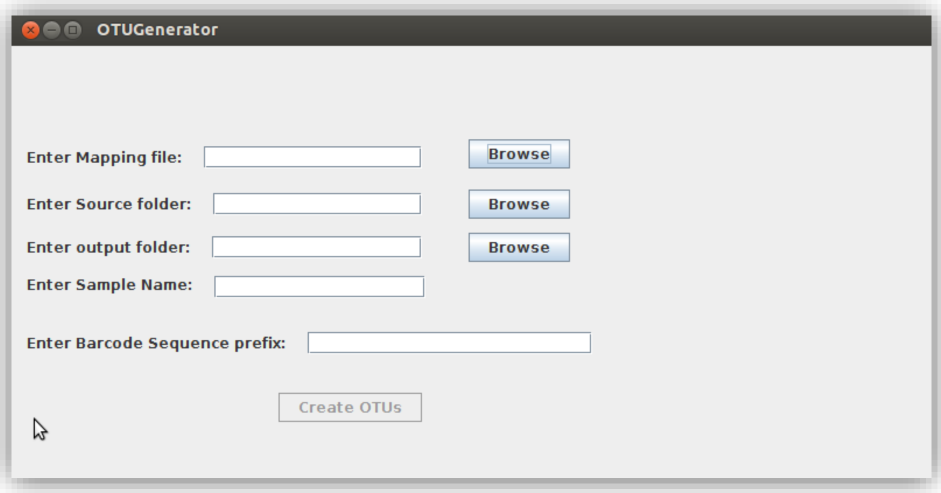
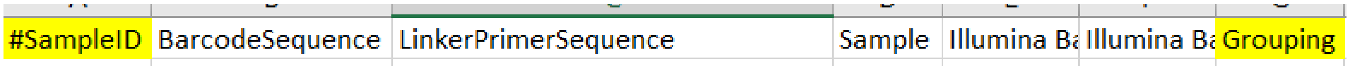
# User Manual – OTUGenerator Software



This utility software facilitates joining of paired-end reads, creation of the seq.fna file and generation of OTU tables using open, closed and De Novo picking.

Step 1:

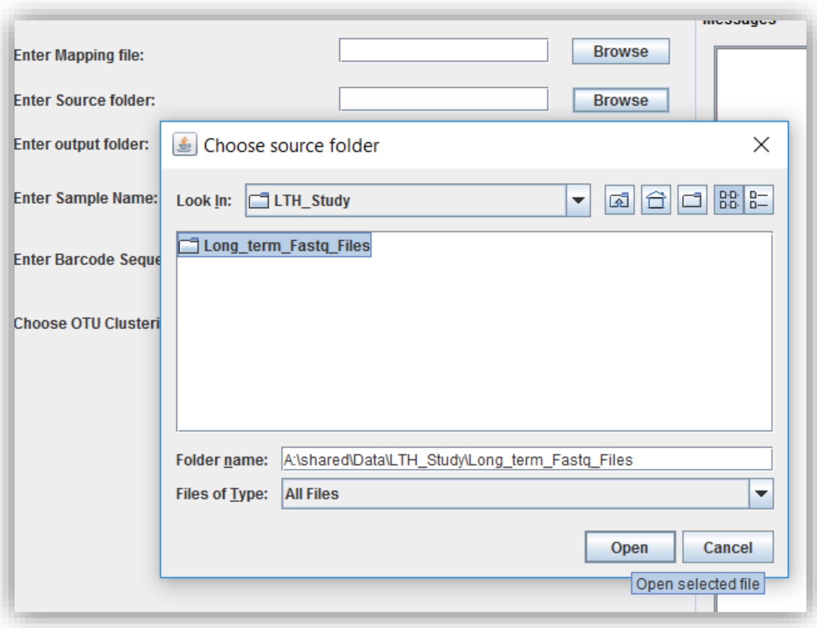
Locate the mapping file using the Browse button. The mapping file MUST contain a column named ‘SampleID’ and a column named ‘Grouping’ as shown below.



| | | | | | | |
|---|---|---|---|---|---|---|
| #SampleID | BarcodeSequence | LinkerPrimerSequence | Sample | Illumina Ba | Illumina Ba | Grouping |

Step 2:

Locate the source folder which contains all the FastQ files and provide its path using the Browse button.



Look In: LTH_Study

Long_term_Fastq_Files

Folder name: A:\shared\Data\LTH_Study\Long_term_Fastq_Files

Files of Type: All Files

Step 3: Locate the output folder you would like the software to generate the OTU Tables and intermediate files in.

Step 4: Type in the sample name which is the common prefix of all fastq files (Illumina naming convention). Example shown below in red.

Step 5: Type in the barcode sequence prefix common to all fastq files. (Shown below in green)

Sample fastq file name: 1-VM1-338rcbc1_S1_L001_R1_001.fastq

Step 6:

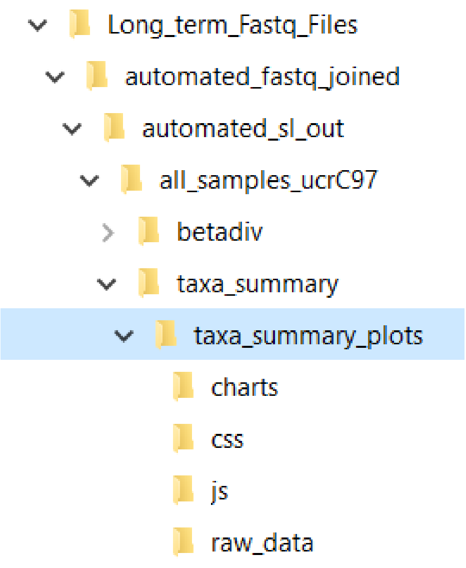
- **Choose OTU clustering method** allows you to choose a clustering method to pick OTUs. For open-reference, closed-reference or Denovo, make the appropriate selection from dropdown and skip to step 7.
- If you select **UPARSE**, you will require an additional package which can be downloaded from here - http://www.drive5.com/usearch/download.html (LINUX)

- You will receive an e-mail with the link to download usearch.
  Rename the downloaded file as "usearch" and put it in the same directory as the java application and python scripts(folder **OTUGenerator**).
- Download some other required scripts here:
  **http://drive5.com/python/**
  Paste this in the folder **OTUGenerator.** Right click on the file and click ‘Extract here’.
- You will also require RDP classifier as the software uses it to assign taxonomy for UPARSE generated cluster. You will find RDP classifier here - https://sourceforge.net/projects/rdp-classifier/files/rdp-classifier/rdp_classifier_2.2.zip/download
- Set path for RDP classifier by executing these commands:
  **RDP_JAR_PATH="/*path to RDP Classifier*/rdp_classifier_2.2/rdp_classifier-2.2.jar"**
  e.g. /home/atharva/Data/rdp_classifier_2.2/rdp_classifier-2.2.jar

Step 7: Click the button Create OTUs. All the outputs would be created in the output directory you mentioned in Step 3.

In the output directory, look for a folder named: “automated_fastq_joined”

This folder contains the following



- Long_term_Fastq_Files
  - automated_fastq_joined
    - automated_sl_out
      - all_samples_ucrC97
        - betadiv
        - taxa_summary
          - taxa_summary_plots
            - charts
            - css
            - js
            - raw_data

The folder taxa_summary_plots will be generated (except for UPARSE) and contains plots representing changes in abundancy of OTUs over samples.

*Note: Diversity plots will not be generated for the sample data provided on the website. The number of samples and the variation in the samples are far below the threshold required to get a diversity plot.*