Note : Most updated version of this document is avialable at

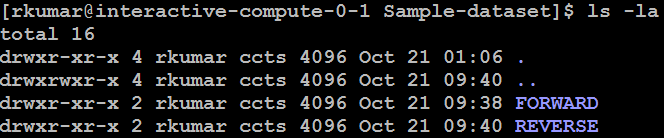
[https://github.com/QWRAP/QWRAP](https://github.com/QWRAP/QWRAP%20)

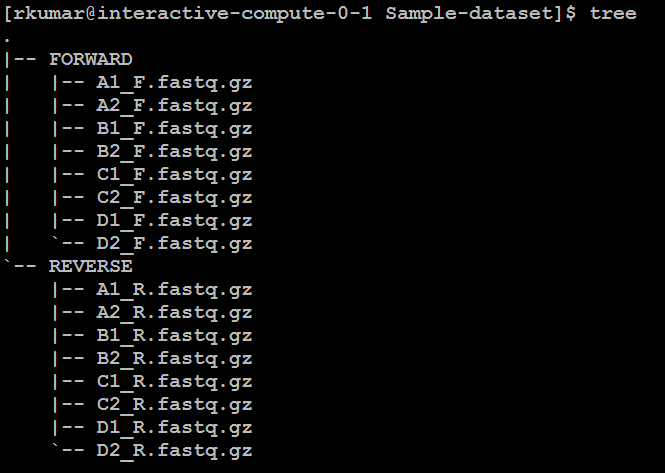
**PAIRED END READS ANALYSIS**

Locate the folders with forward and reverse reads.

Here in this example they are stored in separate folders FORWARD & REVERSE for forward and reverse end reads.

Just check the files.





**Step A: Create directory for analysis**: Create a directory to store analysis results and get (cd) into it.

mkdir ANALYSIS

cd ANALYSIS

**Step B: For Quality check before merging:**  run the program “quality\_check\_before.sh” with location of FWD reads and REV reads.

quality\_check\_before.sh ../FORWARD/ ../REVERSE/

This creates two folders fastqc\_beforef and fastqc\_beforer with FASTQC results for the forward and reverse files and stats.

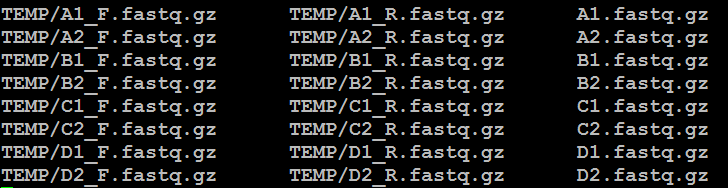
**Step C**: **Merge the forward and reverse files**: Run program: merge\_reads\_F\_R.sh with parameters containing the location of forward and reverse folder. This creates a TEMP folder which has all the reads for the analysis.

merge\_reads\_F\_R.sh ../FORWARD/ ../REVERSE/

**Step D:** **Prepare for merging reads:** The program “prepare\_merge\_fastq.sh” requires the location of TEMP folder (containing both forward and reverse files) as a command line argument. Run the program as

prepare\_merge\_fastq.sh TEMP/

This program creates a mapping file "Paired\_Filelist.txt" with each line containing the location of forward read and reverse reads and the predicted name of merged read.



You can edit the column 3 of the merged file in a text editor if required (especially if you want to rename the files. Make sure that the names in 3rd column should not include underscore (\_).

**Step E: Merge the reads:** This is done using program merge\_fastq.sh. This script uses program USEARCH for merging reads. The program needs 5 parameters which includes the quality control parameters for merging.

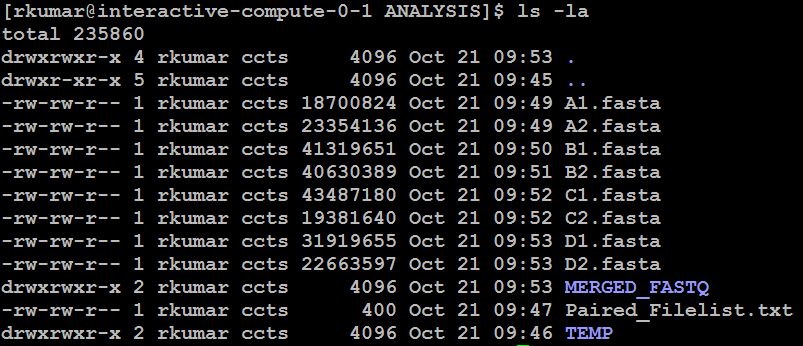
1. Name of mapping file (Paired\_Filelist.txt)
2. Length for trimming forward reads. Provide reads full length if no trimming is required.
3. Length for trimming reverse reads. Provide reads full length if no trimming is required.
4. Max mismatch allowed between forward and reverse reads when aligned.
5. Minimum overlap required between forward and reverse reads when aligned.

In this examples since we had reads of length 251 and decided not to trim them, we used following parameters.

merge\_fastq.sh Paired\_Filelist.txt 251 251 10 200

This script after merging creates all fasta files in the current directory. It also stores the FASTQ files in a folder MERGED\_FASTQ.

So the current ANALYIS directory will look something like this



**Step F: For Quality check after merging:**  run the program “quality\_check\_after.sh” with location of merged FASTQ reads.

quality\_check\_after.sh MERGED\_FASTQ/

This creates two folders fastqc\_afterf and fastqc\_afterr with FASTQC results and stats.

**Step G: Run the microbiome workflow1**.

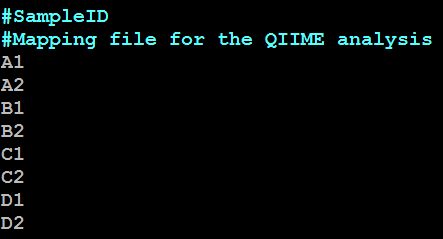
microbiome-workflow1.sh

This script runs inside the ANALYSIS directory and creates many files which are required for QIIME analysis.

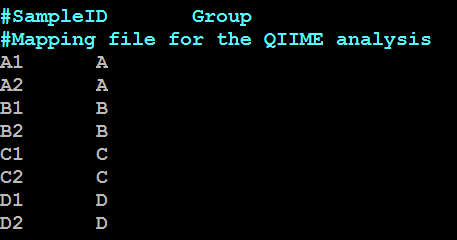
Seqs.fna: Here all fasta files present in a current folder are merged into a single fasta file called seqs.fna. The sample name becomes the header for the reads to distinguish different samples. A snapshot looks as below. Since read headers are not required they are renamed as 1,2… to save space.



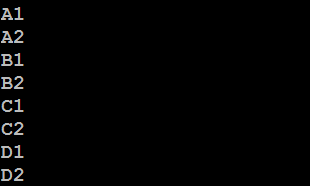
Mapping.txt: The default mapping file has single column containing a header and sample names as shown below



The mapping file can/should be edited to add grouping information about the sample. For example a column can be added as shown below (separated by tab). Similarly more attributed of samples can be added.



Sample\_order.txt: This determines how the samples are ordered in OTU table. This is especially important as the same order is used for taxonomy bas charts. You can change the order of samples here to reorder them.



**Step H: Run the QIIME script:**

The script microbiome-workflow1.sh has created a script “script.sh” which contains a set of QIIME commands to generate taxonomy bar charts. You can change/edit/add the commands if needed and run the script as

sh script.sh

**Step I:** **Run the microbiome workflow2**. This workflow generates script “script\_adv.sh” to do sequence alignment, generate tree, alpha diversity, beta diversity, plots etc. It requires a parameter as even sampling depth. Since all samples have different sampling depth, a minimum sampleing depth should be choosen to normalize all samples to same depth. It the number is 50000, you can run the script as

microbiome-workflow2.sh 5000

This creates a script named “script\_adv.sh”

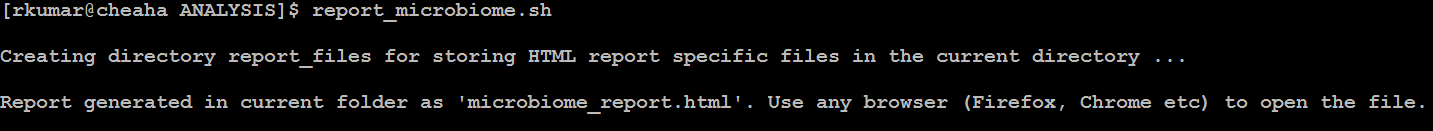
**Step J: Run the QIIME script: You can edit this file as per your requirements and run it as**

sh script\_adv.sh

**Step K: Generate the whole report:** Run the following script to generate the web report.

report\_microbiome.sh

This creates a html report which can be browsed using file microbiome\_report.html



**SINGLE END READS ANALYSIS**

**Step A: Create directory for analysis**: Create a directory to store analysis results and get (cd) into it.

mkdir ANALYSIS

cd ANALYSIS

**Step X: For Quality check before merging:**  run the program “quality\_check\_before.sh” with location of single end reads.

quality\_check\_before.sh ../RAWDATA/

This creates folders fastqc\_beforef with FASTQC results for the forward and reverse files and stats.

**Step Y: Quality filtering** : run the program “quality\_filter\_single.sh” with parameters as described below. This performs trimming of reads if required and two rounds of quality filtering. The parameters of quality filtering can be modified via command line arguments.

**Step Z: For Quality check after merging:**  run the program “quality\_check\_after.sh” with location of filtered FASTQ reads.

quality\_check\_after.sh FILTERED\_FASTQ/

This creates folder fastqc\_afterf with FASTQC results and stats.