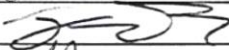



	Title:	Document Number:	Revision:
	COI Metagenomics Qiime2 Analysis	SOP.02.053	A
Document Approval			
Title	Name	Signature	Date
Author	Jin Li		7/19/23
Laboratory Manager	Linda Lam		07/19/23
Laboratory Director	Jin Li, PhD		7/19/23
Quality Assurance	Vattanak Ry		7/19/23

1. Objective

1.1.To define the procedure for analysis COI metagenomics using Qiime2

2. Scope

2.1.This procedure applies to lab personnel who will perform COI metagenomics analysis.

3. Responsibility

3.1.Personnel are responsible for adhering to the details of this SOP and to safety standards when performing this procedure.

4. Definitions

4.1.This protocol was developed to use Qiime2 and COI database to analyze COI metagenomics.

5. Procedure

5.1.Equipment and Materials

5.1.1. Equipment

Computer/server with Qiime2

5.2.Procedure:

5.2.1. Building a COI database from BOLD references


5.2.1.1. Following the tutorial “Building a COI database from BOLD reference” in Qiime2 forum (<https://forum.qiime2.org/t/building-a-coi-database-from-bold-references/16129>)

5.2.1.2. Download derep COI sequence artifact: `bold_derep1_seqs.qza`

5.2.1.3. Download derep COI taxonomy artifact: `bold_derep1_taxa.qza`

5.2.2. Generate taxon chart using Qiime

Using the script below for bioinformatic analysis of BE and F230 libraries.

	Title:	Document Number:	Revision:
	COI Metagenomics Qiime2 Analysis	SOP.02.053	A

```
#import fastq to qiime
```

```
#use bash shell. Qiime V4
```

```
#chsh -s /bin/bash
```

```
#import combined BE run fastq files
```

```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path BE-COMBINED \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path Be-fullrun.qza
```

```
# trim sequences remove primers.
```

```
# This step attempts to find and remove all primer sequences from the ends of each read. This will also remove any Illumina adapter sequence that come before the primer sequence, as per the cutadapt program standard practice. NOTE: the error-rate refers to proportion of error in the primer sequence that is removed. There is no quality filtering at this stage. See https://docs.qiime2.org/2023.5/plugins/available/cutadapt/trim-paired/
```

```
qiime cutadapt trim-paired \
  --i-demultiplexed-sequences Be-fullrun.qza \
  --p-cores 46 \
  --p-front-f CCNGAYATRGCNTTYCCNCG \
  --p-front-r GTRATNGCNCCNGCNARNAC \
  --p-error-rate 0.4 \
  --p-minimum-length 200 \
  --p-discard-untrimmed \
  --p-overlap 10 \
  --o-trimmed-sequences Be-fullruna-trimmed.qza
```

```
#DADA2 DENOISE
```

```
# This step evaluates the error across all sequences and uses that to determine what is ‘real’ variation, and find the likely representative sequences (also called ‘Amplicon Sequence Variants’ (ASVs) and ‘Operational Taxonomic Units’ (OTUs)). Filtering of poor quality sequences is performed at this step, and is based on the error profile of the program’s algorithm. See https://docs.qiime2.org/2023.5/plugins/available/dada2/denoise-paired/
```

```
qiime dada2 denoise-paired \
  --p-n-threads 46 \
  --i-demultiplexed-seqs Be-fullruna-trimmed.qza \
  --p-trim-left-f 0 \
  --p-trim-left-r 0 \
```

```
--p-trunc-len-f 200 \  
--p-trunc-len-r 180 \  
--p-chimera-method consensus \  
--o-representative-sequences trimmed-Be-fullruna-rep-seqs-dada2.qza \  
--o-table trimmed-Be-fullruna-table-dada2.qza \  
--o-denoising-stats trimmed-Be-fullruna-stats-dada2.qza
```

#taxonomy

This step assigns the taxonomy to each representative sequence by aligning each of the representatives to a curated reference sequence database (see links under section 5.2.1 above). The algorithm used to assign taxonomy in the below command is the BLAST program (Basic Local Alignment Search Tool), which finds the best match across any short stretch of sequence. The search usually returns multiple hits, from which a consensus taxonomy is assigned. See <https://docs.qiime2.org/2023.5/plugins/available/feature-classifier/classify-consensus-blast/>

```
qiime feature-classifier classify-consensus-blast \  
--i-query trimmed-Be-fullruna-rep-seqs-dada2.qza \  
--i-reference-taxonomy bold_database/bold_taxa-derep2.qza \  
--i-reference-reads bold_database/bold_seqs-derep2.qza \  
--o-classification bold_Be-fullruna-trimmed-rep-seqs-dada2-taxa-derep2.qza \  
--p-perc-identity 0.97 \  
--p-maxaccepts 10
```

#filter taxonomy

This step will filter the input frequency table to remove sequences occur only in one sample, i.e. remove all singleton sequences

```
qiime feature-table filter-features \  
--i-table trimmed-Be-fullruna-table-dada2.qza \  
--p-min-samples 2 \  
--o-filtered-table FilteredSusan-Be-fullruna-trimmed-table-dada2-derep2.qza
```

#export representative sequences (ASV) with a feature ID to a tsv file

```
qiime tools export \  
--input-path trimmed-Be-fullruna-rep-seqs-dada2.qza \  
--output-path exports
```

rename the generic exported req sequences tsv file based on qiime pipeline

```
mv exports/dna-sequences.fasta exports/trimmed-Be-fullruna-rep-seqs-dada2.fasta
```

export the taxonomy table and assign each feature ID to a taxonomy

```
qiime tools export \  
  --input-path bold_Be-fullruna-trimmed-rep-seqs-dada2-taxa-derep2.qza \  
  --output-path exports
```

```
# and rename taxonomy table based on qiime pipeline
```

```
mv exports/taxonomy.tsv exports/bold_Be-fullruna-trimmed-rep-seqs-dada2-taxa-derep2.tsv
```

```
# export the frequency table in two steps. Frequency table contains feature ID and the frequency  
of the feature per sample
```

```
## first export to export the frequency table to a biom file
```

```
qiime tools export \  
  --input-path trimmed-Be-fullruna-table-dada2.qza \  
  --output-path exports
```

```
# then use the biom command to convert to a tab-delimited table (biom comes packaged with  
qiime)
```

```
biom convert -i exports/feature-table.biom -o exports/FilteredSusan-Be-fullruna-trimmed-table-  
dada2-derep2.tsv --to-tsv
```