README ITS - Hispanic Cervical Samples

UPR

86 Samples Overlapping 16S with ITS

Data analysis: Daniela Vargas Robles

PI: Filipa Godoy

Process performed using DADA2 in QIIME2:

II. Download the UNITE Database & Fit a Classifier

Download the UNITE Taxonomy Database

Download the UNITE taxonomy database from UNITE Repository.

Note: Reason for Using Dynamic Sets

Because DADA2, our chosen denoising method, identifies features at a 99% similarity threshold, using dynamic sets allows it to assign taxonomy at varying thresholds (down to 97%). This is beneficial for less well-developed databases like those for fungi.

Process in QIIME2

```
conda activate qiime2-amplicon-2024.5
```

Import Reference Sequences into QIIME2

```
qiime tools import \
  --type 'FeatureData[Sequence]' \
  --input-path sh_qiime_release_all_25.07.2023/developer/sh_refs_qiime_ver9_dynamic_all_25.07.2023_dev.fasta \
  --output-path unite-sh_refs_qiime_ver9_dynamic_all_25.07.2023_dev.qza
```

Import Reference Taxonomy into QIIME2

```
qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-path sh_qiime_release_all_25.07.2023/developer/sh_taxonomy_qiime_ver9_dynamic_all_25.07.2023_dev.txt \
  --output-path unite-sh_taxonomy_qiime_ver9_dynamic_all_25.07.2023_dev.qza
```

III. Train the Classifier in QIIME2

Fit the Naive Bayes Classifier Note: This step is computationally intensive and may take several hours (approximately 4 hours with 64 GB RAM). If your computer lacks sufficient resources, consider using a more powerful machine.

```
bash qiime feature-classifier fit-classifier-naive-bayes \ --i-reference-reads unite-sh_refs_qiime_ver9_dynamic_all_25.07.2023_dev.qza
```

Alternative Solution

```
I (Daniela Vargas) received the pre-trained classifier unite-ver9-dynamic-all-classifier-25.07.2023.qza from Elif via WeTransfer, beca
```

IV. Importing Sequence Data into QIIME2 (FASTQ Files)

```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path /Users/danielavargasrobles/Library/CloudStorage/GoogleDrive-danielavargasrobles@gmail.com/My\ Drive/Filipa_Daniela_2021
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path demux-paired-end.qza
```

Summarize Demultiplexed Data

```
qiime demux summarize \
  --i-data demux-paired-end.qza \
  --o-visualization demux-paired-end.qzv
```

Vizualize

```
qiime tools view demux-paired-end.qzv
```

Quality vizualition

```
fastqc *.fastq.gz
```

Verifying primers in sequences in the terminal

It seems that no illumina adapters are present based on fastqc FWD:GAACGCAGCRAAIIGYGA; REV:TCCTCCGCTTATTGATATGC

Run in the terminal:

```
COMPLETE_SEQ="TCCTCCGCTTATTGATATGC"
COUNT=$(zgrep -c "$COMPLETE_SEQ" *_001.fastq.gz)
echo "$COUNT"
```

```
COMPLETE_SEQ="GAACGCAGCAAAAAGTGA"
COUNT=$(zgrep -c "$COMPLETE_SEQ" *_001.fastq.gz)
echo "$COUNT"
```

Conclusion: no primers found

VI. Denoising the sequences using DADA2 in QIIME2

Since I downloaded the trainer, I have to install the qiime version that were used to generate the classifier:

Install version 2023.7

```
conda deactivate
```

```
wget https://data.qiime2.org/distro/core/qiime2-2023.7-py38-osx-conda.yml
```

```
conda env create -n qiime2-amplicon-2023.7 --file qiime2-2023.7-py38-osx-conda.yml
```

Activate old qiime2 version

```
conda activate qiime2-amplicon-2023.7
```

Note: --p-trim-left-r 90 \ # esto corta la secuencia reverse (segun la imagen de calidad) 90 bases de derecha a izq

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs demux-paired-end.qza \
  --p-trim-left-f 10 \
  --p-trim-left-r 90 \
  --p-trunc-len-f 0 \
  --p-trunc-len-r 0 \
  --o-representative-sequences dada2_paired_end_rep_seqs.qza \
  --o-table dada2_paired_end_table.qza \
  --o-denoising-stats dada2_paired_end_stats.qza
```

“--p-trunc-len 0” indicates that you don’t cut anything – remember, these ITS sequences have different sizes so you can’t set one specific size or you lose information.

Vizualize

```
qiime metadata tabulate \
  --m-input-file dada2_paired_end_stats.qza \
  --o-visualization dada2_paired_end_stats.qzv
```

```
qiime tools view dada2_paired_end_stats.qzv
```

VII. Taxonomic classification of denoised sequences with UNITE classifier

Using the classifier `unite-ver9-dynamic-all-classifier-25.07.2023.qza` that Elif provided (I did not have to run it! too heavy).

```
time qiime feature-classifier classify-sklearn \
--i-reads dada2_paired_end_rep_seqs.qza \
--i-classifier  QIIME2/unite-ver9-dynamic-all-classifier-25.07.2023.qza \
--o-classification ITS-UNITE9dyna-dada2-taxonomy
--p-n-jobs 1
```

export ASV table

```
qiime tools export --input-path dada2_paired_end_table.qza --output-path exported-asv-table
```

export rep seqs table

```
qiime tools export --input-path dada2_paired_end_rep_seqs.qza --output-path exported-rep_seqs
```

Export files into R

...