How to prepare for the Sheba/TAU microbiome analysis workshop (5.2019)

In order to participate in the hands-on analysis "sandbox" in the workshop, please follow these steps:
1. Bring your laptop (It can be mac/linux or windows)
2. Install qiime2 **LATEST VERSION** (**2019.1**):
   - For windows, use the qiime2 virtualbox image. Follow instructions **here**
   - For mac/linux, follow the qiime2 conda installation instructions **here**

If you would like to analyze your own dataset, you will need to prepare it BEFORE the workshop using these steps:

For Calour/dbBact
(Performing these steps will enable focusing the workshop time on using the new tools rather than spending time on this computationally heavy steps).

1. denoise your fastq reads file using your preferred method (DADA2/Deblur):

1a. Import the reads files. Exact commands depend on the data you start with (see the qiime2 importing data tutorial).

1b. Demultiplex the data:
```
qiime demux emp-single --i-seqs YOUR_DEMUX_FILE.qza --m-barcodes-file YOUR_MAP_FILE.tsv --m-barcodes-column BarcodeSequence --o-per-sample-sequences demux.qza
```
(for more details, see the qiime2 moving pictures tutorial)

1c. Denoise the reads using either DADA2 or Deblur:
IMPORTANT: we want to have exact sequences and not hashes as the bacterial IDS, so must use the `--p-no-hashed-feature-ids` flag.
Dada2 (assuming single end and not paired end sequences):
```
qiime dada2 denoise-single --i-demultiplexed-seqs demux.qza --p-trim-left 0 --p-trunc-len 150 --o-representative-sequences rep-seqs-dada2.qza --o-table table-dada2.qza --o-denoising-stats stats-dada2.qza --p-no-hashed-feature-ids
```

Debur: use these 2 commands
```
qiime quality-filter q-score --i-demux demux.qza --o-filtered-sequences demux-filtered.qza --o-filter-stats demux-filter-stats.qza
```

```
qiime deblur denoise-16S --i-demultiplexed-seqs demux-filtered.qza --p-trim-length 150 --o-representative-sequences rep-seqs-deblur.qza --o-table table-deblur.qza --p-sample-stats --o-stats deblur-stats.qza --p-no-hashed-feature-ids
```

2. Create a phylogenetic tree for your data (needed for qiime2 alpha/beta diversity)

```
qiime phylogeny align-to-tree-mafft-fasttree --i-sequences rep-seqs.qza --o-alignment aligned-rep-seqs.qza --o-masked-alignment masked-aligned-rep-seqs.qza --o-tree unrooted-tree.qza --o-rooted-tree rooted-tree.qza
```

NOTE: it is preferable to use SEPP for the tree creation, but it is more computationally intensive (will discuss more in the workshop).

3. (optional) In the workshop we will use EZCalour (the GUI version of Calour). If you would like to use the python version (for jupyter notebook, etc.) please follow the Calour installation instruction here.

Preparation for using the Borenstein lab tools (FishTaco, MIMOSA)

Completing these installations in advance will help ensure that you are able to spend most of the workshop time interacting with data and results. But don't worry if you are not able to complete these steps or if you encounter any difficulties - we can help you address them during the sandbox portion of the workshop.

1) Make sure that you have R and RStudio installed on your laptop.
Installation instructions for R can be found here: https://cran.r-project.org
Installation instructions for Rstudio here: https://www.rstudio.com/products/rstudio/download/

2) Install several R packages, which are dependencies for MIMOSA.
From CRAN: devtools, data.table, and Rcpp
From Bioconductor: KEGGREST and qvalue

Open RStudio and run the code below in the console to install all of these:

```
cran_packages = c("devtools", "data.table", "Rcpp", "BiocManager")
for (package in cran_packages){
  install.packages(package)
}

bioc_packages = c("qvalue", "KEGGREST")
for (package in bioc_packages){
  BiocManager::install(package)
}
```

3) Install a Python conda environment containing the MUSiCC and FishTaco packages. To set up the environment, download and save this .yml file to your computer. Then run the following command in a command line shell:

```
conda env create -f fishtaco_1-1-1.yml
```

Activate the environment by running:

```
conda activate fishtaco
```

Once the environment is activated, test the installation:

```
test_fishtaco.py
```

Deactivate the environment after testing:

```
conda deactivate
```

If you do not want to use Anaconda or if you run into problems, you can also install FishTaco and its dependencies individually by following the instructions here: https://borenstein-lab.github.io/fishtaco/installation.html

4) If you would like to analyze your own 16S rRNA data, processed with qiime2 for the first part of the workshop, you can perform an additional step in your qiime2 analysis to obtain a closed-reference Greengenes OTU table.

First, download and unzip the Greengenes 13.8 database from here:
https://docs.qiime2.org/2019.1/data-resources/#marker-gene-reference-databases
You will use the reference file located in gg_13_8_otus/rep_set/97_otus.fasta , which contains representative 16S rRNA sequences for 97% Operational Taxonomic Units.

Run the following commands in your qiime2 environment to map the sequences in your Deblur or DADA2 table to Greengenes reference OTUs:

```
# Import Greengenes reference OTU sequences
qiime tools import \
  --input-path 97_otus.fasta \
  --output-path 97_otus.qza \
  --type 'FeatureData[Sequence]'

## Here table.qza and rep-seqs.qza are the processed outputs from your
Deblur or DADA2 analysis
qiime vsearch cluster-features-closed-reference \
  --i-table table.qza \
  --i-sequences rep-seqs.qza \
  --i-reference-sequences 97_otus.qza \
  --p-perc-identity 0.97 \
  --o-clustered-table table-cr-97.qza \
  --o-clustered-sequences rep-seqs-cr-97.qza \
  --o-unmatched-sequences unmatched-cr-97.qza

## Export the table from qiime, and convert the biom table to a tab-
delimited table
```

```
qiime tools export \
  --input-path table-cr-97.qza \
  --output-path greengenes-cr-feature-table

biom convert -i greengenes-cr-feature-table/feature-table.biom -o
otu_table.txt --to-tsv
```

The resulting OTU table can be used as input for BURRITO and FishTaco.

Any questions, please email:
Microbiome.Center@sheba.health.gov.il and/or elbo@tauex.tau.ac.il