Qiime2 workflow for Microbe-mediated Mn-oxidation

Sjöberg et al. submitted

2021-09-06

All files excluding the SRA datafiles are provided at the following figshare: 10.6084/m9.figshare.16574222

Qiime2 installation

```
wget https://data.qiime2.org/distro/core/qiime2-2021.4-py38-linux-conda.yml
conda env create -n qiime2-2021.4 --file qiime2-2021.4-py38-linux-conda.yml
rm qiime2-2021.4-py38-linux-conda.yml
```

Download reads using fastq-dump (SRA tool kit) and convert file names to QIIME2-friendly names.

```
mkdir reads forwardreads
cd reads
for f in SRR9202442 SRR9202456 SRR9202435 SRR9202458 ;
do
  fastq-dump --gzip --split-3 $f ;
done

for x in *_1.fastq.gz ;
do
  g=`basename -s _1.fastq.gz $f` ;
  mv $f ${g}_0_L001_R1_001.fastq.gz ;
done

for x in *_2.fastq.gz ;
do
  g=`basename -s _1.fastq.gz $f` ;
  mv $f ${g}_0_L001_R1_002.fastq.gz ;
done

mv *1.fastq.gz ../forwardreads
```

Import reads into qiime2 artifact `demux-single-end.qza`:

```
source activate qiime2.2021.04
qiime tools import \
  --type 'SampleData[SequencesWithQuality]' \
  --input-path forwardreads/ \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path demux-single-end.qza
```

Denoise reads using dada2 in QIIME2:

```
qiime dada2 denoise-single \
  --i-demultiplexed-seqs demux-single-end.qza \
  --p-trunc-len 275 \
  --p-trim-left 25 \
  --o-representative-sequences rep-seqs-dada2.qza \
  --o-table table-dada2.qza \
  --o-denoising-stats stats-dada2.qza
```

Cluster the features assigned by DADA2 using VSEARCH at the 97% identity level. Where the metadata file is a simple two coloum file with the SRR number and a second column place holder.

```
qiime vsearch cluster-features-de-novo \
  --i-table table-dada2.qza \
  --i-sequences rep-seqs-dada2.qza \
  --p-perc-identity 0.97 \
  --o-clustered-table table-dada2.0.97.qza \
  --o-clustered-sequences rep-seqs-dada2.0.97.qza

qiime feature-table summarize \
  --i-table table-dada2.0.97.qza \
  --o-visualization table-dada2.0.97.qza \
  --m-sample-metadata-file metadata.tsv

qiime feature-table tabulate-seqs \
  --i-data rep-seqs-dada2.0.97.qza \
  --o-visualization rep-seqs-dada2.0.97.qzv

qiime tools export \
  --input-path rep-seqs-dada2.0.97.qza \
  --output-path rep-seqs-dada2.0.97.tsv
```

Calculate the relative frequency of each feature across all samples:

```
qiime feature-table relative-frequency \
  --i-table table-dada2.0.97.qza \
  --output-dir REL_FREQ

qiime tools export \
  --input-path REL_FREQ/relative_frequency_table.qza \
  --output-path REL_FREQ/

biom convert -i REL_FREQ/feature-table.biom -o relative_frequency_table.tsv --to-tsv
```

Perform taxonomy classification using the SILVA 139 silva-138-99-nb-classifier.qza available from https://docs.qiime2.org/2021.4/data-resources/

```
confidence=0.7

feature-classifier classify-sklearn \
      --p-reads-per-batch 10 \
      --p-confidence $confidence \
      --i-classifier silva-138-99-nb-classifier.qza \
      --i-reads rep-seqs-dada2.0.97.qza \
      --verbose \
      --o-classification taxonomy-dada2.confidence.$confidence.qza
qiime taxa barplot \
      --i-table table-dada2.0.97.qza \
      --i-taxonomy taxonomy-dada2.confidence.$confidence.qza  \
      --m-metadata-file metadata.tsv \
      --o-visualization taxonomy-dada2.barplots.confidence.$confidence.qzv

qiime tools export \
      --input-path taxonomy-dada2.confidence.$confidence.qza \
      --output-path summary_data_$confidence/
```

The taxonomy and sequence of each feature was added to the relative_frequency_table.tsv using an in-house python script `QIIME2_combine_relabund_taxonomy_sequence.py`.

```
python3 QIIME2_combine_relabund_taxonomy_sequence.py \
     -i relative_frequency_table.tsv \
```

```
        -j summary_data_$confidence/taxonomy.tsv \
        -f rep-seqs-dada2.0.97.tsv/dna-sequences.fasta \
        -o relative_frequency_table_annotated.$confidence.tsv
```