

All following code was executed in Ubuntu 20.04 for Windows System Linux, using dedicated Conda v4.9.2 environments.

Increase shell available resources to handle large number of files.

```
ulimit -S -n 4096
```

Demultiplexing paired-end reads with combinatorial dual indexes.

```
cutadapt -j 15 -e 0.15 --no-indels -g file:fw_adapters.fasta -G file:rv_adapters.fasta -o ./biobank_{name1}-{name2}.R1.fastq.gz -p ./biobank_{name1}-{name2}.R2.fastq.gz biobank_R1.fastq.gz biobank_R2.fastq.gz
```

Import sequences in QIIME2.

```
qiime tools import \  
  --type 'SampleData[PairedEndSequencesWithQuality]' \  
  --input-path biobank_demult_qiime2_manifest_file.tsv \  
  --output-path biobank.qza \  
  --input-format PairedEndFastqManifestPhred33V2
```

Summarize artifact: statistics about the sample reads including average quality.

```
qiime demux summarize \  
  --i-data biobank.qza \  
  --o-visualization biobank.qza_summary.qzv
```

Sequence quality control and feature table construction.

```
qiime dada2 denoise-paired \  
  --p-n-threads 15 \  
  --i-demultiplexed-seqs biobank.qza \  
  --p-trim-left-f 0 \  
  --p-trim-left-r 0 \  
  --p-trunc-len-f 260 \  
  --p-trunc-len-r 210 \  
  --o-representative-sequences biobank_rep-seqs-dada2.qza \  
  --o-table biobank_table-dada2.qza \  
  --o-denoising-stats biobank_stats-dada2.qza
```

Taxonomy assignment using Silva 138 full length sequences Naive Bayes classified database. Pre-trained classifier dataset available at: <https://data.qiime2.org/2022.4/common/silva-138-99-nb-classifier.qza>

```
qiime feature-classifier classify-sklearn \  
  --i-classifier silva-138-99-nb-classifier.qza \  
  --i-reads biobank_rep-seqs-dada2.qza \  
  --o-classification biobank_dada2_taxonomy_silva_138_full_nb.qza
```