Jul 07, 2020

Taxonomic classification of IonTorrent-sequenced 16S amplicon sequences

Laura Espina[1]

[1]Cardiff University

1 Works for me dx.doi.org/10.17504/protocols.io.bh8pj9vn

Laura Espina

DOI

dx.doi.org/10.17504/protocols.io.bh8pj9vn

PROTOCOL CITATION

Laura Espina 2020. Taxonomic classification of IonTorrent-sequenced 16S amplicon sequences. **protocols.io** dx.doi.org/10.17504/protocols.io.bh8pj9vn

CREATED

Jul 04, 2020

LAST MODIFIED

Jul 07, 2020

PROTOCOL INTEGER ID

38895

DISCLAIMER:

Protocol using QIIME2 pipeline, v2019-4 (doi: 10.1038/s41587-019-0209-9) and the Naïve Bayes classifier implemented in the q2-feature-classifier plugin (doi: 10.1186/s40168-018-0470-z). This classifier was trained on the Greengenes 13_8 99% OTU database (doi: 10.1038/ismej.2011.139). R v3.6.1 environment was used with Phyloseq v1.29.0 (doi: 10.1371/journal.pone.0061217).

Pre-processing

1

Torrent Suite 5.4
by Thermo Fisher Scientific

Demultiplex sequences. Quality-filter sequences. Tim barcodes and adapters. Export as fastq files (1 fastq file per sample).

Example of fastq files at this point:

fcm1_1_L001_R2_001.fastq

fcm2_2_L001_R2_001.fastq

2 Import fastq files.
Trim primers.
Remove sequences shorter than 100 bp.
Suggested software:

Geneious 11.1.5
by Biomatters Ltd

(or python scripts).

3 (Optional) Check the Phred quality score of the sequences.
Suggested software:

FastQC 0.11.9
by Simon Andrews

On Qiime2

4 **To train a Naïve Bayes classifier**: Download the Greengenes 13_8 99% OTU database. Upload the files "99_otus.fasta" and "99_otu_taxonomy.txt" onto a folder accesible by the Qiime2 environment. With Qiime2 activated and inside the folder containing those files, type in the command line:

4.1 >qiime tools import --type 'FeatureData[Sequence]' --input-path 99_otus.fasta --output-path otus.qza

4.2 >qiime tools import --type 'FeatureData[Taxonomy]' --input-format HeaderlessTSVTaxonomyFormat --input-path 99_otu_taxonomy.txt --output-path taxonomy.qza

4.3 >qiime feature-classifier extract-reads --i-sequences otus.qza --p-f-primer AGAGTTTGATCMTGGCTCAG --p-r-primer CYNACTGCTGCCTCCCGTAG --o-reads ref-seqs.qza

4.4 >qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads ref-seqs.qza --i-reference-taxonomy taxonomy.qza --o-classifier classifierp.qza

The artifact "classifier.qza" can be copied to any other folder for the following steps.
Example of **classifier trained on the Greengenes 99% OTU database for the V1-V2 region of the 16S rRNA gene:**

classifier.qza

Citation: Laura Espina (07/07/2020). Taxonomic classification of IonTorrent-sequenced 16S amplicon sequences. https://dx.doi.org/10.17504/protocols.io.bh8pj9vn

5 **To perform taxonomy classification**: In a new folder, upload the artifact "classifier.qza" and the file with the metadata corresponding to the fastq files in the correct format. Example of file:

metadata.tsv

Also create a subfolder "fastq" containing the compressed fastq files in the correct nominal format (eg. "fcm1_1_L001_R1_001.fastq.qz" and "fcm2_2_L001_R1_001.fastq.qz") and the classifier artifact.
In the command line, type in:

5.1 >qiime tools import --type 'SampleData[SequencesWithQuality]' --input-path fastq --input-format CasavaOneEightSingleLanePerSampleDirFmt --output-path demux-single-end.qza

5.2 >qiime vsearch dereplicate-sequences --i-sequences demux-single-end.qza --o-dereplicated-table dereplicatedtable.qza --o-dereplicated-sequences rep-seqs.qza

5.3 >qiime feature-classifier classify-sklearn --i-classifier classifier.qza --i-reads rep-seqs.qza --o-classification taxonomy.qza

5.4 >qiime taxa barplot --i-table dereplicatedtable.qza --i-taxonomy taxonomy.qza --m-metadata-file sample-metadata.tsv --o-visualization taxa-bar-plots.qzv

The artifact "taxa-bar-plots.qzv" can be interactively visualized in https://view.qiime2.org/
Example of artifact:

taxa-bar-plots.qzv

5.5 The components of the biom table (the OTU abundance table and the taxonomy table) can be obtained from the "taxa-bar-plots.qzv" artifact or with the following commands:
>qiime tools export --input-path dereplicatedtable.qza --output-path exported-feature-table
>cd exported-feature-table
>biom convert -i feature-table.biom -o otutable.tsv --to-tsv
>qiime tools export --input-path taxonomy.qza --output-path taxonomy

The files "taxonomy.tsv" and "otutable.tsv" should be downloaded and can be manually converted into .csv files and adapted for the analysis with R. The additional file "metadata.csv" is also necessary.
Example of files:

taxonomy.csv otutable.csv metadata.csv

Analysis with R

6 Within R and activating the *phyloseq* package, different types of analysis can be easily performed.

Citation: Laura Espina (07/07/2020). Taxonomic classification of IonTorrent-sequenced 16S amplicon sequences. https://dx.doi.org/10.17504/protocols.io.bh8pj9vn

6.1 The first step is the creation of the phyloseq object using the type of files shown in step 5.5:

```
> otu_table <- read.csv("otutable.csv",sep=",",row.names=1)
> tax_table <- read.csv("taxonomy.csv",sep=",",row.names=1)
> OTU <- otu_table(otu_table,taxa_are_rows=TRUE)
> tax_matrix<- as.matrix(tax_table)
> TAX <- tax_table(tax_matrix)
> metadata <- read.csv("metadata.csv",sep=",",row.names=1)
> meta <- sample_data(metadata)
> fcm <- phyloseq(OTU, TAX, meta)
```

6.2 Basic operations such as tax_glom are used to prepare the data for further analyses. For example, with the following commands a table containing the taxonomic abundance for the families present in each sample is created:

```
> fcm.family <- tax_glom(fcm,"Family", NArm= FALSE)
> fcm.family.m <- psmelt(fcm.family)
> write.table(x=fcm.family.m, sep = ",", file = "fcm_family.csv")
```

At this point a table with the taxonomic abundance of each family is produced. Example of file:

fcm_family.csv

Citation: Laura Espina (07/07/2020). Taxonomic classification of IonTorrent-sequenced 16S amplicon sequences. https://dx.doi.org/10.17504/protocols.io.bh8pj9vn