MICROBIAL ITS PROFILING PACKAGE

User Manual

Installation

The Microbial ITS profiling analysis package can be downloaded from the website "http://probiogenomics.unipr.it/pbi/" as a .zip file containing:

- Bacterial_ITS_annotator.sh: the bash script that performs the Microbial ITS profiling analysis.
- "database" folder: folder containing the Microbial ITS database.
- "fastq_input" folder: folder where input fastq files needs to be placed.

These three elements must be placed in the same folder in a linux environment with Qiime2 installed. All the dependencies/software needed to run the Bacterial_ITS_annotator.sh bash script are included in Qiime2.

Usage

In order to perform the Bacterial ITS profiling analysis, a single .fastq files for each sample must be placed in the "fastq_input" folder. In case paired-end sequencing was used, the .fastq file provided in the "fastq_input" folder must correspond to reads starting with primer UNI_ITS_fw (5'-KRGGRYKAAGTCGTAACAAG-3') covering the 3'-end of the 16S rRNA sequence.

The script can be executed in the linux console simply by moving to the folder path where the script is located and by typing "./Bacterial_ITS_annotator.sh".

Console text will notify the user that the "Fasting_Map.txt" and "sample-metadata.tsv" file has been created and allows the user to add to this file additional columns with samples' metadata that

will be used for downstream analyses (e.g. PCoA representation of beta-diversity) before continue the execution of the script.

Console messages will report the ongoing analyses performed by the script.

A log file named “Qiime2_log.txt” is also created.

Results

Upon completion of the pipeline by the Bacterial_ITS_annotator.sh script, the filtered .fastq files will be placed in the “fastq_input_filtered” folder and all temporary files generated along the pipeline are moved in the “data_analysis” folder.

Furthermore, results will be placed in the “results” folder. In detail the “results” folder contains:

- “alpha_diversity” folder: results of the alpha diversity analysis using the Simpson, Shannon, Faith, Chao1 and Observed OTUs indexes.
- “beta_diversity” folder: results of the beta diversity analysis using the unweighted UniFrac, weighted UniFrac and Bray Curtis indexes.
- “taxonomy” folder: results of the taxonomic profiling at Phylum, Family, Genus and Species level.
- “OTUs” folder: results of the OTUs analysis encompassing the reference sequence of each OTUS (OTUs_rep_seqs.fasta), the taxonomy attributed at each OTUs (OTUs_taxonomy.txt) and the table of OTUs abundance for each sample (OTUs_table.txt).
- “forward_primer_screening_stats.txt” file: results of the UNI_ITS_fw primer screening through cutadapt.
- “filtering_table.txt” file: filtering table of all the analysed samples.

Notes

- If no metadata is added to the “sample-metadata.tsv” files, the script will be completed but few analyses will be skipped.
- Alpha-diversity analysis is set to reach a maximum subsampling of 30.000 reads. In case no sample reaches this depth, the analysis will be skipped. The user can lower this value at line 151 of the Bacterial_ITS_annotator.sh script, by changing the option “--p-max-depth 30000”.
- Sensitivity and Specificity of the taxonomic classification can be adjusted by the user by editing command options at line 156 of the Bacterial_ITS_annotator.sh script. Details are reported in the Qiime2 manual (https://docs.qiime2.org) at the “qiime feature-classifier” section.