PacBio **pb-16S-nf** pipeline

[VIB - Nucleomics Core, nucleomics@vib.be]

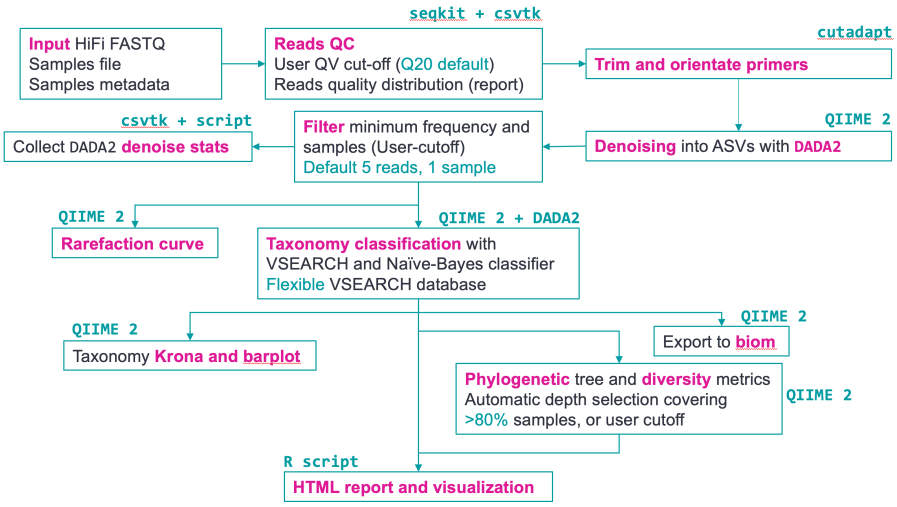
*FriJun*07, 2024 - version 1.0

General Information

Analyse demultiplexed data obtained from one or more Sequel-IIe 16S amplicon sequencing runs and return user-readable summary data and data objects that the end-user can evaluate further using the QIIME2 toolbox.

The analysis was performed using default pipeline parameters except for the Rarefaction which was set to **'auto'** in order to better compare separate run experiments (this value is generally optimal for a regular run where most of the samples return more than 8000-10000 HiFi reads). Please refer to the separate **barcode_QC_v11.html** document to find out about your sample depth and see which samples were below that arbitrary limit.

Note: Some of the default options might not be ideal for every biological question and project. If a new analysis needs be done with other parameters, the user can contact us for a quote or use the pb-16S-nf tool on their own bioinformatics infrastructure [1].



(https://github.com/PacificBiosciences/HiFi-16S-workflow)

[1] https://github.com/PacificBiosciences/HiFi-16S-workflow

Analysis settings

The version of the **pb-16S-nf** pipeline used here was **v0.6** [2].

Table 1: list of analysis arguments accepted by the pb-16S-nf pipeline

Other important options:	SET
–front_p Forward primer sequence. Default to F27. (default: AGRGTTYGATYMTGGCTCAG)	AGRGTTYGATYMTGGCTCAG
–adapter_p Reverse primer sequence. Default to R1492. (default: AAGTCGTAACAAGGTARCY)	AAGTCGTAACAAGGTARCY
–filterQ Filter input reads above this Q value (default: 20).	20
–downsample Limit reads to a maximum of N reads if there are more than N reads (default: off)	OFF
–max_ee DADA2 max_EE parameter. Reads with number of expected errors higher than this value will be discarded (default: 2)	2
–minQ DADA2 minQ parameter. Reads with any base lower than this score will be removed (default: 0)	0
–min_len Minimum length of sequences to keep (default: 1000)	1000
–max_len Maximum length of sequences to keep (default: 1600)	1600
–pooling_method QIIME 2 pooling method for DADA2 denoise see QIIME 2 documentation for more details (default: “pseudo”, alternative: “independent”)	pseudo
–maxreject max-reject parameter for VSEARCH taxonomy classification method in QIIME 2 (default: 100)	100
–maxaccept max-accept parameter for VSEARCH taxonomy classification method in QIIME 2 (default: 100)	100
–min_asv_totalfreq Total frequency of any ASV must be above this threshold across all samples to be retained. Set this to 0 to disable filtering (default 5)	5
–min_asv_sample ASV must exist in at least min_asv_sample to be retained. Set this to 0 to disable. (default 1)	1
–vsearch_identity Minimum identity to be considered as hit (default 0.97)	0.97
–rarefaction_depth Rarefaction curve “max-depth” parameter. By default the pipeline automatically select a cut-off above the minimum of the denoised reads for >80% of the samples. This cut-off is stored in a file called “rarefaction_depth_suggested.txt” file in the results folder (default: null)	‘auto’ (*)
–skip_primer_trim Skip all primers trimming (switch off cutadapt and DADA2 primers removal) (default: trim with cutadapt)	trim
–skip_nb Skip Naive-Bayes classification (only uses VSEARCH) (default: false)	
–colorby Columns in metadata TSV file to use for coloring the MDS plot in HTML report (default: condition)	condition=run#
–run_picrust2 Run PICRUSt2 pipeline. Note that pathway inference with 16S using PICRUSt2 has not been tested systematically (default: false)	FALSE

NOTE: We used mostly default values unless noted by (*). The list of applied parameters can be found in **nextflow_reports/parameters.txt**.

Zymo Control sample

When performing the amplification of the 16S V1V9 amplicon for the customer, we usually include a negative control (buffer) and a positive control from the Zymo mock community [3] (ref:**D6305**). These samples are labelled ‘Neg_ctrl’ and ‘Pos_ctrl’ in our results and will correspond to different barcode pairs in each experiment (amplified alongside with the customer samples).

A PDF file found on the Zymo site describes the mock sample in details [4]

last edits: $FriJun07, 2024$

[2] https://github.com/PacificBiosciences/HiFi-16S-workflow

[3] https://zymoresearch.eu/collections/zymobiomics-microbial-community-standards/products/zymobiomics-microbial-community-dna-standard

[4] https://files.zymoresearch.com/datasheets/ds1706_zymobiomics_microbial_community_standards_data_sheet.pdf