

[PacificBiosciences/pb-16S-nf：Nextflow管道分析PacBio HiFi全长16S数据 (github.com)](https://github.com/PacificBiosciences/pb-16S-nf)

Install

默认情况下，所有软件依赖关系都通过 进行管理。Nextflow 将用于构建 所需的环境，因此您无需手动构建任何环境。

conda install -c bioconda nextflow

# If this is your first time using conda

conda init

安装 Nextflow 后，克隆存储库并 使用以下命令下载数据库。

git clone https://github.com/PacificBiosciences/pb-16S-nf.git

cd pb-16S-nf

nextflow run main.nf --download\_db

# With docker (If you use docker, add -profile docker to all Nextflow-related command)

nextflow run main.nf --download\_db -profile docker

下载数据库后，在克隆的文件夹中运行以下命令 若要查看管道的选项，请执行以下操作：

nextflow run main.nf --help

Usage:

This pipeline takes in the standard sample manifest and metadata file used in

QIIME 2 and produces QC summary, taxonomy classification results and visualization.

For samples TSV, two columns named "sample-id" and "absolute-filepath" are

required. For metadata TSV file, at least two columns named "sample\_name" and

"condition" to separate samples into different groups.

nextflow run main.nf --input samples.tsv --metadata metadata.tsv \\

--dada2\_cpu 8 --vsearch\_cpu 8

By default, sequences are first trimmed with cutadapt. If adapters are already trimmed, you can skip

cutadapt by specifying "--skip\_primer\_trim".

Other important options:

--front\_p Forward primer sequence. Default to F27. (default: AGRGTTYGATYMTGGCTCAG)

--adapter\_p Reverse primer sequence. Default to R1492. (default: AAGTCGTAACAAGGTARCY)

F27 <- "AGRGTTYGATYMTGGCTCAG"

R1492 <- "RGYTACCTTGTTACGACTT"

--filterQ Filter input reads above this Q value (default: 20).

--downsample Limit reads to a maximum of N reads if there are more than N reads (default: off)

--max\_ee DADA2 max\_EE parameter. Reads with number of expected errors higher than

this value will be discarded (default: 2)

--minQ DADA2 minQ parameter. Reads with any base lower than this score

will be removed (default: 0)

--min\_len Minimum length of sequences to keep (default: 1000)

--max\_len Maximum length of sequences to keep (default: 1600)

--pooling\_method QIIME 2 pooling method for DADA2 denoise see QIIME 2

documentation for more details (default: "pseudo", alternative: "independent")

--maxreject max-reject parameter for VSEARCH taxonomy classification method in QIIME 2

(default: 100)

--maxaccept max-accept parameter for VSEARCH taxonomy classification method in QIIME 2

(default: 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)

--min\_asv\_sample ASV must exist in at least min\_asv\_sample to be retained.

Set this to 0 to disable. (default 1)

--vsearch\_identity Minimum identity to be considered as hit (default 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)

--dada2\_cpu Number of threads for DADA2 denoising (default: 8)

--vsearch\_cpu Number of threads for VSEARCH taxonomy classification (default: 8)

--cutadapt\_cpu Number of threads for primer removal using cutadapt (default: 16)

--outdir Output directory name (default: "results")

--vsearch\_db Location of VSEARCH database (e.g. silva-138-99-seqs.qza can be

downloaded from QIIME database)

--vsearch\_tax Location of VSEARCH database taxonomy (e.g. silva-138-99-tax.qza can be

downloaded from QIIME database)

--silva\_db Location of Silva 138 database for taxonomy classification

--gtdb\_db Location of GTDB r202 for taxonomy classification

--refseq\_db Location of RefSeq+RDP database for taxonomy classification

--skip\_primer\_trim Skip all primers trimming (switch off cutadapt and DADA2 primers

removal) (default: trim with cutadapt)

--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)

--run\_picrust2 Run PICRUSt2 pipeline. Note that pathway inference with 16S using PICRUSt2

has not been tested systematically (default: false)

--download\_db Download databases needed for taxonomy classification only. Will not

run the pipeline. Databases will be downloaded to a folder "databases"

in the Nextflow pipeline directory.

--publish\_dir\_mode Outputs mode based on Nextflow "publishDir" directive. Specify "copy"

if requires hard copies. (default: symlink)

--version Output version

若要测试管道，请运行以下示例。请注意，数据库的路径需要 如果不同，则更改为服务器上的相应位置（请参阅上面的参数）。如果你 按照上面的命令，数据库将被下载到文件夹中的文件夹中 并且您不需要指定路径。默认情况下，Conda 环境将在文件夹中创建，除非文件中发生更改。

databasespb-16S-nf$HOME/nf\_condanextflow.config

# Create sample TSV for testing

echo -e "sample-id\tabsolute-filepath\ntest\_data\t$(readlink -f test\_data/test\_1000\_reads.fastq.gz)" > test\_data/test\_sample.tsv

nextflow run main.nf --input test\_data/test\_sample.tsv \

--metadata test\_data/test\_metadata.tsv -profile conda \

--outdir results

# To test using Singularity or docker (change singularity to docker)

nextflow run main.nf --input test\_data/test\_sample.tsv \

--metadata test\_data/test\_metadata.tsv -profile singularity \

--outdir results