DATARMOR | **R** | **DATARMOR**

1. DATA PREPROCESSING: *remove primers, re-pair reads, produce tar archive*
a. ligation-preprocess for ligation data
b. cleaning_process for double-PCR data

2. dada_main _process *To process single runs with DADA2*

Output: *ASV counts, taxonomy (RDP), and fasta files*

2. frogs_only *Activates FROGS pipeline on tar archive: FROGS preprocessing (~contig length filtering), swarm clustering, FROGS chimera removal, taxo. assignment (RDP+BLAST)*

3. extract_amplicons process**

2bis.dada_mergeasv_process*

3. blast_assignation process *Taxonomic assignment of ASVs with BLAST* COI: Genbank COI extraction, MIDORI SSU: Silva, Greengenes, PR2 Look into: **/home/ref-bioinfo/tools_data**

2bis.dada_mergeasv_process*

3. blast_assignation process *Taxonomic assignment of ASVs with BLAST*

4. frogs_process *swarm clustering, FROGS chimera removal, taxo. assignment (RDP+BLAST)*

5. extract_amplicons process**

ASV table: 10_ASVs_counts.tsv 10_ASVs_taxonomy.tsv ASVs_rdp_blast_taxonomy.tsv

ASV sequences: 10_ASVs.fasta

OTU table: swarm_5_affiliation4.tsv swarm_5_affiliation4_multi.tsv **OTU representative sequences:** swarm_4_filters4.fasta

Decontam + possible taxonomic filtration#

6. tag_switching_ norm_process *Abundance-renormalization using owi_renormalize.R*

7. lulu_process *LULU curation using lulu.R*

Curated ASVs (which are now OTUs)

Curated OTUs

*OPTIONAL: Merge several _asv.rds files of different runs analysed together, and produce DADA2 output files of the merged data. Ifonly 1 run is analysed you can directly use the output of dada_main_process.

** If one wants to work on OTU variants (individual ASVs making up the OTUs) at some point in the data analysis, activate this process. Only works in FROGS.

#The step of decontam and filtering in R is optionally inserted within the flow of the pipe. It may be skipped and run after the pipeline is completed BUT this results in much longer runs of LULU, which duration increases exponentially with the number of ASV/OTU being treated.

Clarification on databases and assignments methods:
As of 04/2020,

RDP can be used on ASVs using database files in: **/home/ref-bioinfo/tools_data/dada2**

BLAST can be used on any fasta file in **/home/ref-bioinfo/tools_data/frogs**

Within FROGS, the scripts are written so that BLAST and RDP assignments occur in parallel (--rdp option).
This works ONLY if databases are in the FROGS format.
In: **/home/ref-bioinfo/tools_data/frogs**, databases in the functional FROGS format are:
16S_greengenes_13.5; 16S_silva132; 16S_silva132-abyss-seq; 18S_pr2 (v 4.5)
BOLD_COI-5P_2019-02-01; midori (full and marine versions)

For the moment being, with Genbank COI database only BLAST works (remove --rdp in your personal copy of the abyss-pipeline frogs.pbs script line 116)

For more FROGS-formatted databases, see:
http://genoweb.toulouse.inra.fr/frogs_databanks/assignation/
Or contact
frogs-support@inra.fr