H3ABioNet

Pan African Bioinformatics Network for H3Africa

16SrRNA Intermediate Bioinformatics Online Course: Int_BT_2019

16S analysis pipeline

QC and ASV picking using the dada2 pipeline

Practical

Practical

The practical is available here:

https://iallali.github.io/DADA2_pipeline/16SrRNA_DADA2_pipeline.html

Practical

iallali.github.io

16S Tutorial | General Information | Tutorial

Getting Ready

First, we load the dada2 package on your RStudio. if you do not already have it, see the dada2 installation instructions.

```
library(dada2); packageVersion("dada2")
```

```
## [1] '1.13.1'
```

We set the path so that it points to the extracted directory of the dataset named "dog_samples" on your computer or cluster:

```
MY_HOME <- Sys.getenv("HOME")
data <- paste(MY_HOME, "/dada2_tutorial_dog/dog_samples", sep='')  # change the path
list.files(data)
```

```
##  [1] "Dog1_R1.fastq"  "Dog1_R2.fastq"  "Dog10_R1.fastq" "Dog10_R2.fastq"
##  [5] "Dog15_R1.fastq" "Dog15_R2.fastq" "Dog16_R1.fastq" "Dog16_R2.fastq"
##  [9] "Dog17_R1.fastq" "Dog17_R2.fastq" "Dog2_R1.fastq"  "Dog2_R2.fastq"
## [13] "Dog22_R1.fastq" "Dog22_R2.fastq" "Dog23_R1.fastq" "Dog23_R2.fastq"
## [17] "Dog24_R1.fastq" "Dog24_R2.fastq" "Dog29_R1.fastq" "Dog29_R2.fastq"
## [21] "Dog3_R1.fastq"  "Dog3_R2.fastq"  "Dog30_R1.fastq" "Dog30_R2.fastq"
## [25] "Dog31_R1.fastq" "Dog31_R2.fastq" "Dog8_R1.fastq"  "Dog8_R2.fastq"
## [29] "Dog9_R1.fastq"  "Dog9_R2.fastq"  "filtered"
```

If your listed files match those here, you can start running the DADA2 pipeline.

Now, we read in the names of the fastq files and we sort them by forward and reverse. Then, we perform some string manipulation to extract a list of the sample names.

```
# Forward and reverse fastq filenames have format: SAMPLENAME_R1.fastq and SAMPLENAME_R2.fastq
dataF <- sort(list.files(data, pattern="_R1.fastq", full.names = TRUE))
dataR <- sort(list.files(data, pattern="_R2.fastq", full.names = TRUE))

# Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq
list.sample.names <- sapply(strsplit(basename(dataF), "_"), `[`, 1)
list.sample.names
```

```
##  [1] "Dog1"  "Dog10" "Dog15" "Dog16" "Dog17" "Dog2"  "Dog22" "Dog23"
##  [9] "Dog24" "Dog29" "Dog3"  "Dog30" "Dog31" "Dog8"  "Dog9"
```