R Notebook

#telecharge les documents necessaire pour permettre l'utilisation de dada2

```
refdb_folder <- here::here("data", "refdb")
refdb_folder
```

```
## [1] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/refdb"
```

#permet de creer un chemin d'accès a un dosier

```
if (!dir.exists(refdb_folder)) dir.create(refdb_folder, recursive = TRUE)
```

permet de "verifier" que la commande precedente a bien creer le dossier, sinon elle le cree

```
file.copy(from = "course-material-main/data/raw", to = "data", recursive = TRUE)
```

```
## [1] TRUE
```

#copie les donnees de course-material-main/data/raw dans data

```
getOption("timeout")
```

```
## [1] 60
```

```
options(timeout = 1200)

silva_train_set <- file.path(refdb_folder,"silva_nr99_v138.1_train_set.fa.gz")
silva_species_assignment <- file.path(refdb_folder,"silva_species_assignment_v138.1.fa.gz")

if (!file.exists(silva_train_set)) {download.file("https://zenodo.org/record/4587955/files/silva_nr99_v
}

if (!file.exists(silva_species_assignment)) {download.file("https://zenodo.org/record/4587955/files/sil
}
```

#tout d abord on augmente le temps de "calcul" pour permettre de laisser plus de temps au code de tourner
#les lignes de codent vont verifier si les fichiers sont presents et sinon elles vont les telechargers

```
devtools::load_all("/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/course-material-m
```

```
## i Loading ANF_metaB
```

#telecharge tous les outils necessaires pour faire tourner le script

```
path_to_fastqs <- here::here("data", "raw")

print(path_to_fastqs)
```

```
## [1] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw"
```

#copie les donnes de data raw dans le fichier path to fastqs

```
fnFs <- sort(list.files(path_to_fastqs,
                        pattern = "_R1.fastq.gz",
                        full.names = TRUE))
print(fnFs)
```

```
##  [1] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S11B_R1.fastq.gz"
##  [2] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S1B_R1.fastq.gz"
##  [3] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S2B_R1.fastq.gz"
##  [4] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S2S_R1.fastq.gz"
##  [5] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S3B_R1.fastq.gz"
##  [6] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S3S_R1.fastq.gz"
##  [7] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S4B_R1.fastq.gz"
##  [8] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S4S_R1.fastq.gz"
##  [9] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S5B_R1.fastq.gz"
## [10] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S5S_R1.fastq.gz"
## [11] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S6B_R1.fastq.gz"
## [12] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S6S_R1.fastq.gz"
## [13] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S7B_R1.fastq.gz"
## [14] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S7S_R1.fastq.gz"
## [15] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S8B_R1.fastq.gz"
## [16] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S8S_R1.fastq.gz"
## [17] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S9B_R1.fastq.gz"
## [18] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S9S_R1.fastq.gz"
```

#trie les donnees du path to fastqs et mets les R1 dans une nouvelle valeure

```
fnRs <- sort(list.files(path_to_fastqs,
                        pattern = "_R2.fastq.gz",
                        full.names = TRUE))
print(fnRs)
```

```
##  [1] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S11B_R2.fastq.gz"
##  [2] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S1B_R2.fastq.gz"
##  [3] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S2B_R2.fastq.gz"
##  [4] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S2S_R2.fastq.gz"
##  [5] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S3B_R2.fastq.gz"
```

```
##  [6] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S3S_R2.fastq.gz"
##  [7] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S4B_R2.fastq.gz"
##  [8] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S4S_R2.fastq.gz"
##  [9] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S5B_R2.fastq.gz"
## [10] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S5S_R2.fastq.gz"
## [11] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S6B_R2.fastq.gz"
## [12] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S6S_R2.fastq.gz"
## [13] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S7B_R2.fastq.gz"
## [14] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S7S_R2.fastq.gz"
## [15] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S8B_R2.fastq.gz"
## [16] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S8S_R2.fastq.gz"
## [17] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S9B_R2.fastq.gz"
## [18] "/Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tutoriel_dada2/data/raw/S9S_R2.fastq.gz"
```

#trie les donnees du path to fastqs et mets les R2 dans une nouvelle valeure

```
sample_names <- basename(fnFs) |>
  strsplit(split = "_") |>
  sapply(head, 1)
```

#permet d'extraire les noms des fchiers avant le "_" et de les regrouper dans sample_names

```
basename(fnFs) |>
  head()
```

```
## [1] "S11B_R1.fastq.gz" "S1B_R1.fastq.gz"  "S2B_R1.fastq.gz"  "S2S_R1.fastq.gz"
## [5] "S3B_R1.fastq.gz"  "S3S_R1.fastq.gz"
```

#montre les premières ligne du fichier

```
basename(fnFs) |>
  strsplit(split = "_") |>
  head()
```

```
## [[1]]
## [1] "S11B"        "R1.fastq.gz"
##
## [[2]]
## [1] "S1B"         "R1.fastq.gz"
##
## [[3]]
## [1] "S2B"         "R1.fastq.gz"
##
## [[4]]
## [1] "S2S"         "R1.fastq.gz"
##
## [[5]]
## [1] "S3B"         "R1.fastq.gz"
##
## [[6]]
## [1] "S3S"         "R1.fastq.gz"
```

#permet de mieux separer les noms des fichers dans le chemin fnFs afin de pouvoir les utiliser de façon optimal après

```
basename(fnFs) |>
  strsplit(split = "_") |>
  sapply(head, 1) |>
  head()
```

```
## [1] "S11B" "S1B"  "S2B"  "S2S"  "S3B"  "S3S"
```

#pareil que au dessus

```
gsub("^.+/|_.+$", "", fnFs) |> head()
```

```
## [1] "S11B" "S1B"  "S2B"  "S2S"  "S3B"  "S3S"
```

#"nettoie" encore plus les noms

```
quality_folder <- here::here("outputs","dada2","quality_plots")

if (!dir.exists(quality_folder)) {dir.create(quality_folder, recursive = TRUE)
}

qualityprofile(fnFs,fnRs,file.path(quality_folder,"quality_plots.pdf"))
```

```
## pdf
##   2
```

#genere un pdf regroupant les graphiques

```
path_to_trimmed_reads <- here::here("outputs","dada2","trimmed")

if (!dir.exists(path_to_trimmed_reads)) dir.create(path_to_trimmed_reads, recursive = TRUE)
```

verifie si un dossier existe, si il existe ne fait rien sinon il le creer

```
primer_fwd  <- "CCTACGGGNBGCASCAG"
primer_rev  <- "GACTACNVGGGTATCTAAT"
```

#assigne au nom des primers le code "ATCG" correspondant

```
Biostrings::readDNAStringSet(fnFs[1],format = "fastq",nrec = 10)
```

```
## DNAStringSet object of length 10:
##      width seq                                               names
##  [1]   293 CCTACGGGGGGCAGCAGTAGGGA...ACATCGGCTTAACCGATGAAGT M01522:260:000000...
##  [2]   293 CCTACGGGTGGCACCAGTAGGGA...CGGGGCTTAACCTCGGAACTGC M01522:260:000000...
```

```
##  [3]   292 CCTACGGGGCGCAGCAGGCGCGA...GGGACCGGGAGAGGTGTGAGGT M01522:260:000000...
##  [4]   293 CCTACGGGGTGCAGCAGTAGGGA...TCAAAACTCCCAGTCTAGAGTT M01522:260:000000...
##  [5]   291 CCTACGGGTGGCAGCAGTGGGGA...GCAGTGGAAACTGTTGGGCTTG M01522:260:000000...
##  [6]   293 CCTACGGGATGCAGCAGGCGCGA...GGGACCGGGAGAGGTGTGGGGG M01522:260:000000...
##  [7]   292 CCTACGGGATGCAGCAGTGGGGA...TTTAATCCTGATGAGCTAGAAA M01522:260:000000...
##  [8]   293 CCTACGGGGCGCAGCAGTAGGGA...TTAAAACTTTTGTTCTGGAATT M01522:260:000000...
##  [9]   292 CCTACGGGTTGCAGCAGTGGGGA...ATTAAAACTTTTCAGCTAGAGT M01522:260:000000...
## [10]   293 CCTACGGGAGGCAGCAGTGGGGA...CCCGGGCTCAACCTGGGAACGG M01522:260:000000...
```

```
Biostrings::readDNAStringSet(fnRs[1],format = "fastq", nrec = 10)
```

```
## DNAStringSet object of length 10:
##      width seq                                               names
##  [1]   301 GACTACCAGGGTATCTAATCCTG...GGCTGCTGGCACGAAGTTCGCC M01522:260:000000...
##  [2]   301 GACTACCGGGGTATCTAATCCTG...GGCTGCTGGCACGGAGTTAGCC M01522:260:000000...
##  [3]   300 AATCCGGTTCGTGCCCCTAGGCT...TCTTTCCCAGCCCTTATTCCAA M01522:260:000000...
##  [4]   301 GACTACCGGGGTATCTAATCCTG...GGCTGCTGGCACGGAGTTAGCC M01522:260:000000...
##  [5]   301 GACTACCGGGGTATCTAATCCCT...GGCTGCTGGCCCGGAATTAGCC M01522:260:000000...
##  [6]   301 GGTATCTAATCCGGTTCGTGCCC...CACCGTCCTTACCCCCCCCTTT M01522:260:000000...
##  [7]   301 GGTATCTAATCTTGTTTGCTCCC...CCCGACGTTAGCCGGGGCTTCT M01522:260:000000...
##  [8]   301 GACTACGAGGGTATCTAATCCCG...GGCTGCTGGCACGGAATTAGCC M01522:260:000000...
##  [9]   301 GGTATCTAATCCTCTTCGCTACC...CACGAAGTTAGCCGGACCTTCT M01522:260:000000...
## [10]   301 GACTACGGGGGTATCTAATCCTG...GGCTGCCGGCACGGGGTTAGCC M01522:260:000000...
```

#pareil pour les deux lignes: lit les dix premères lignes du codes

```
(primer_log <- primer_trim(
  forward_files = fnFs,
  reverse_files = fnRs,
  primer_fwd = primer_fwd,
  primer_rev = primer_rev,
  output_dir = path_to_trimmed_reads,
  min_size = 200
))
```

```
##    sample status in_reads   in_bp too_short too_long too_many_n out_reads
## 1    S11B     OK     2000 1186767         0        0          0      1870
## 2     S1B     OK     2000 1186613         1        0          0      1857
## 3     S2B     OK     2000 1186942         0        0          0      1847
## 4     S2S     OK     2000 1186868         0        0          0      1839
## 5     S3B     OK     2000 1186650         0        0          0      1862
## 6     S3S     OK     2000 1186475         1        0          0      1885
## 7     S4B     OK     2000 1186331         2        0          0      1871
## 8     S4S     OK     2000 1186681         0        0          0      1882
## 9     S5B     OK     2000 1186386         1        0          0      1847
## 10    S5S     OK     2000 1186501         1        0          0      1866
## 11    S6B     OK     2000 1186261         2        0          0      1844
## 12    S6S     OK     2000 1187078         1        0          0      1844
## 13    S7B     OK     2000 1186888         0        0          0      1832
## 14    S7S     OK     2000 1186299         3        0          0      1849
## 15    S8B     OK     2000 1186354         3        0          0      1849
## 16    S8S     OK     2000 1186610         1        0          0      1854
## 17    S9B     OK     2000 1187038         0        0          0      1840
```

```
## 18    S9S     OK     2000 1186867         0        0          0      1838
##    w/adapters qualtrim_bp out_bp w/adapters2 qualtrim2_bp out2_bp
## 1        1986           0 515080        1883            0  530630
## 2        1975           0 511648        1879            0  526475
## 3        1987           0 508866        1858            0  523698
## 4        1989           0 506650        1849            0  521725
## 5        1989           0 512875        1872            0  528100
## 6        1989           0 518974        1896            0  534213
## 7        1980           0 515444        1888            0  530542
## 8        1987           0 518269        1894            0  533465
## 9        1984           0 508624        1862            0  523759
## 10       1991           0 513915        1874            0  529047
## 11       1981           0 507956        1863            0  523242
## 12       1982           0 508408        1860            0  523180
## 13       1987           0 504963        1843            0  519967
## 14       1987           0 509627        1861            0  524203
## 15       1993           0 509662        1856            0  524756
## 16       1982           0 510830        1871            0  525831
## 17       1983           0 507075        1857            0  522451
## 18       1979           0 506347        1856            0  520976
```

#permet de retirer les amorces des séquences

```
nopFw <- sort(list.files(path_to_trimmed_reads, pattern = "R1", full.names = TRUE))
nopRv <- sort(list.files(path_to_trimmed_reads, pattern = "R2", full.names = TRUE))
```

#permet de creer deux valeurs contenant les R1 et R2 des sequences

```
path_to_filtered_reads <- here::here("outputs", "dada2", "filtered")
if (!dir.exists(path_to_filtered_reads)) dir.create(path_to_filtered_reads, recursive = TRUE)
```

#verifie si le fichier existe bien sinon il le creer

```
filtFs <- file.path(path_to_filtered_reads, basename(fnFs))
filtRs <- file.path(path_to_filtered_reads, basename(fnRs))
```

#creer deux valeurs avec les reads filtres

```
names(filtFs) <- sample_names
names(filtRs) <- sample_names
```

#creer deux dossier pour faciliter l'utlisation des donnees presentent

```
(out <- dada2::filterAndTrim(
  fwd = nopFw,
  filt = filtFs,
  rev = nopRv,
  filt.rev = filtRs,
  minLen = 150,
  matchIDs = TRUE,
  maxN = 0,
  maxEE = c(3, 3),
  truncQ = 2
))
```

```
##                   reads.in reads.out
## S11B_R1.fastq.gz      1870      1202
## S1B_R1.fastq.gz       1857      1251
## S2B_R1.fastq.gz       1847      1257
## S2S_R1.fastq.gz       1839      1245
## S3B_R1.fastq.gz       1862      1245
## S3S_R1.fastq.gz       1885      1313
## S4B_R1.fastq.gz       1871      1262
## S4S_R1.fastq.gz       1882      1331
## S5B_R1.fastq.gz       1847      1256
## S5S_R1.fastq.gz       1866      1245
## S6B_R1.fastq.gz       1844      1253
## S6S_R1.fastq.gz       1844      1243
## S7B_R1.fastq.gz       1832      1205
## S7S_R1.fastq.gz       1849      1184
## S8B_R1.fastq.gz       1849      1171
## S8S_R1.fastq.gz       1854      1269
## S9B_R1.fastq.gz       1840      1196
## S9S_R1.fastq.gz       1838      1250
```

#permet de filtrer les donnees a l'aide de dada2, elimine les sequences de mauvaises qualitees, aide a faire correspondre R1 R2 afin d'obtenir des donnees de qualitees pour les manipulations a suivre

```
errF <- dada2::learnErrors(filtFs,randomize = TRUE,multithread = TRUE)
```

```
## 6164786 total bases in 22378 reads from 18 samples will be used for learning the error rates.
```

#permet d'estimer les erreurs de sequençages

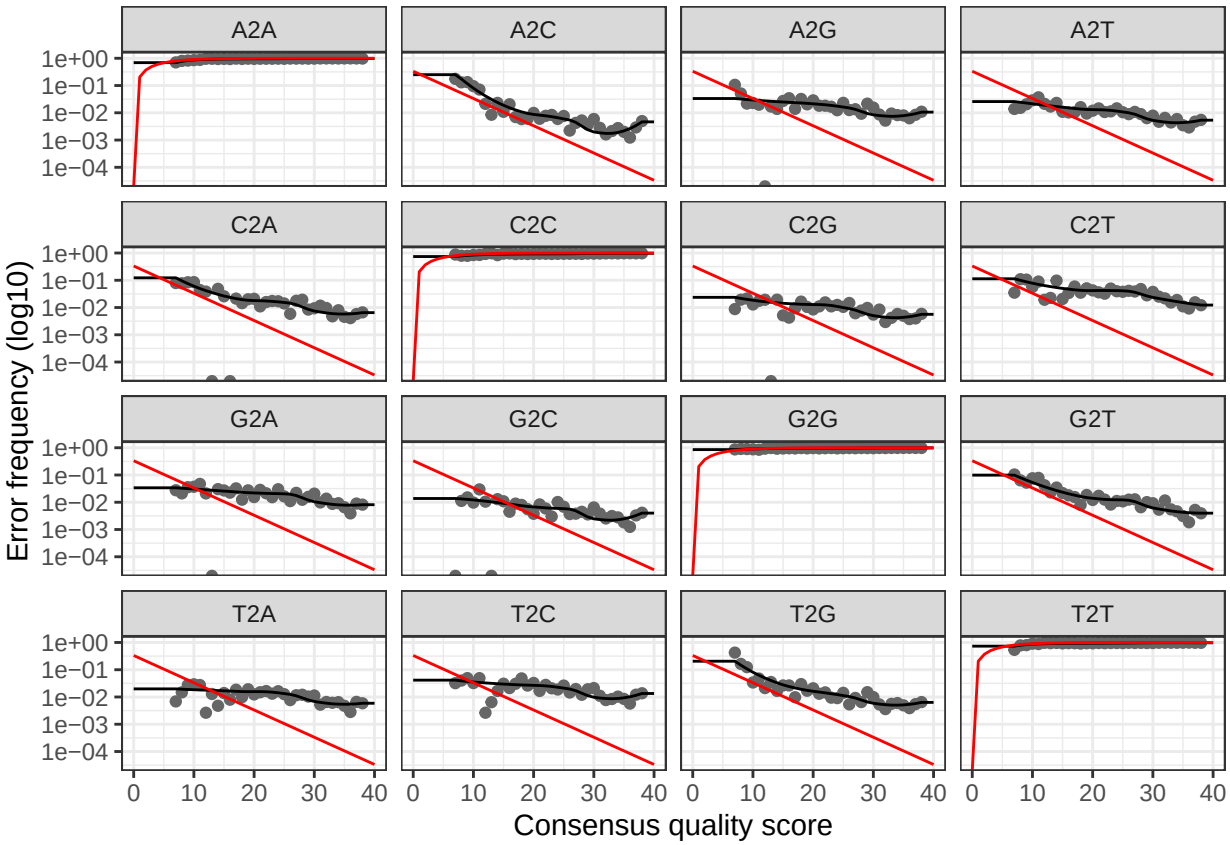
```
errR <- dada2::learnErrors(filtRs,randomize = TRUE,multithread = TRUE)
```

```
## 6345784 total bases in 22378 reads from 18 samples will be used for learning the error rates.
```

#pareil

```
dada2::plotErrors(errF, nominalQ=TRUE)
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
## Transformation introduced infinite values in continuous y-axis
```



#genere des graphiques montrant les erreurs après filtrage

```
derepFs <- dada2::derepFastq(filtFs, verbose = TRUE)
```

```
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 755 unique sequences from 1202 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 779 unique sequences from 1251 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 791 unique sequences from 1257 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 762 unique sequences from 1245 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 773 unique sequences from 1245 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
```

```
## Encountered 763 unique sequences from 1313 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 738 unique sequences from 1262 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 641 unique sequences from 1331 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 783 unique sequences from 1256 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 664 unique sequences from 1245 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 697 unique sequences from 1253 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 658 unique sequences from 1243 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 693 unique sequences from 1205 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 676 unique sequences from 1184 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 698 unique sequences from 1171 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 714 unique sequences from 1269 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 685 unique sequences from 1196 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 677 unique sequences from 1250 total sequences read.
```

```r
derepRs <- dada2::derepFastq(filtRs, verbose = TRUE)
```

```
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 930 unique sequences from 1202 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 948 unique sequences from 1251 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 970 unique sequences from 1257 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 926 unique sequences from 1245 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 949 unique sequences from 1245 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 968 unique sequences from 1313 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 953 unique sequences from 1262 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 906 unique sequences from 1331 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 976 unique sequences from 1256 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 888 unique sequences from 1245 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
## Encountered 916 unique sequences from 1253 total sequences read.
## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu
```

```
## Encountered 848 unique sequences from 1243 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 883 unique sequences from 1205 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 876 unique sequences from 1184 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 880 unique sequences from 1171 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 969 unique sequences from 1269 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 893 unique sequences from 1196 total sequences read.

## Dereplicating sequence entries in Fastq file: /Users/yvanohemartinier/Documents/Fac/Cours/ADM/ADM_tu

## Encountered 912 unique sequences from 1250 total sequences read.
```

#contient les donnees apres dereplication ce qui va simplifier leurs traitements

```
dadaFs <- dada2::dada(derepFs, err = errF, multithread = TRUE)
```

```
## Sample 1 - 1202 reads in 755 unique sequences.
## Sample 2 - 1251 reads in 779 unique sequences.
## Sample 3 - 1257 reads in 791 unique sequences.
## Sample 4 - 1245 reads in 762 unique sequences.
## Sample 5 - 1245 reads in 773 unique sequences.
## Sample 6 - 1313 reads in 763 unique sequences.
## Sample 7 - 1262 reads in 738 unique sequences.
## Sample 8 - 1331 reads in 641 unique sequences.
## Sample 9 - 1256 reads in 783 unique sequences.
## Sample 10 - 1245 reads in 664 unique sequences.
## Sample 11 - 1253 reads in 697 unique sequences.
## Sample 12 - 1243 reads in 658 unique sequences.
## Sample 13 - 1205 reads in 693 unique sequences.
## Sample 14 - 1184 reads in 676 unique sequences.
## Sample 15 - 1171 reads in 698 unique sequences.
## Sample 16 - 1269 reads in 714 unique sequences.
## Sample 17 - 1196 reads in 685 unique sequences.
## Sample 18 - 1250 reads in 677 unique sequences.
```

```r
dadaRs <- dada2::dada(derepRs, err = errR, multithread = TRUE)
```

```
## Sample 1 - 1202 reads in 930 unique sequences.
## Sample 2 - 1251 reads in 948 unique sequences.
## Sample 3 - 1257 reads in 970 unique sequences.
## Sample 4 - 1245 reads in 926 unique sequences.
## Sample 5 - 1245 reads in 949 unique sequences.
## Sample 6 - 1313 reads in 968 unique sequences.
## Sample 7 - 1262 reads in 953 unique sequences.
## Sample 8 - 1331 reads in 906 unique sequences.
## Sample 9 - 1256 reads in 976 unique sequences.
## Sample 10 - 1245 reads in 888 unique sequences.
## Sample 11 - 1253 reads in 916 unique sequences.
## Sample 12 - 1243 reads in 848 unique sequences.
## Sample 13 - 1205 reads in 883 unique sequences.
## Sample 14 - 1184 reads in 876 unique sequences.
## Sample 15 - 1171 reads in 880 unique sequences.
## Sample 16 - 1269 reads in 969 unique sequences.
## Sample 17 - 1196 reads in 893 unique sequences.
## Sample 18 - 1250 reads in 912 unique sequences.
```

```r
mergers <- dada2::mergePairs(
  dadaF = dadaFs,
  derepF = derepFs,
  dadaR = dadaRs,
  derepR = derepRs,
  maxMismatch = 0,
  verbose = TRUE
)
```

```
## 881 paired-reads (in 28 unique pairings) successfully merged out of 972 (in 51 pairings) input.

## 835 paired-reads (in 33 unique pairings) successfully merged out of 943 (in 63 pairings) input.

## 785 paired-reads (in 30 unique pairings) successfully merged out of 946 (in 59 pairings) input.

## 930 paired-reads (in 32 unique pairings) successfully merged out of 1041 (in 59 pairings) input.

## 787 paired-reads (in 26 unique pairings) successfully merged out of 928 (in 60 pairings) input.

## 920 paired-reads (in 36 unique pairings) successfully merged out of 1040 (in 60 pairings) input.

## 808 paired-reads (in 29 unique pairings) successfully merged out of 971 (in 62 pairings) input.

## 1052 paired-reads (in 32 unique pairings) successfully merged out of 1133 (in 56 pairings) input.

## 906 paired-reads (in 24 unique pairings) successfully merged out of 1037 (in 40 pairings) input.

## 898 paired-reads (in 27 unique pairings) successfully merged out of 1039 (in 56 pairings) input.
```

```
## 971 paired-reads (in 31 unique pairings) successfully merged out of 1063 (in 51 pairings) input.

## 904 paired-reads (in 23 unique pairings) successfully merged out of 1066 (in 62 pairings) input.

## 824 paired-reads (in 31 unique pairings) successfully merged out of 990 (in 67 pairings) input.

## 854 paired-reads (in 30 unique pairings) successfully merged out of 970 (in 48 pairings) input.

## 844 paired-reads (in 26 unique pairings) successfully merged out of 946 (in 58 pairings) input.

## 851 paired-reads (in 31 unique pairings) successfully merged out of 1033 (in 62 pairings) input.

## 788 paired-reads (in 25 unique pairings) successfully merged out of 977 (in 55 pairings) input.

## 873 paired-reads (in 29 unique pairings) successfully merged out of 1045 (in 57 pairings) input.
```

#cree une valeure avec les sequences fusionnees apres le denovo-assemblage

```
seqtab <- dada2::makeSequenceTable(mergers)
```

#cree une valeur qui sert de table de comptage pour les sequences dans les echantillons

```
seqtab_nochim <- dada2::removeBimeraDenovo(seqtab, method = "consensus",multithread = TRUE,verbose = TRU
```

```
## Identified 2 bimeras out of 162 input sequences.
```

#creer une valeure qui regroupent les sequences apres l'elimination des chimeres

```
taxonomy <- dada2::assignTaxonomy(
  seqs = seqtab_nochim,
  refFasta = silva_train_set,
  taxLevels = c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus","Species"),
  multithread = TRUE,
  minBoot = 60)
```

#cree une valeure qui contient les attributions taxonomiques

```
taxonomy <- dada2::addSpecies(taxonomy,silva_species_assignment,allowMultiple = FALSE)
```

#met a jour la valeure precedente au niveau des especes avec l'outil silva

```
export_folder <- here::here("outputs", "dada2", "asv_table")

if (!dir.exists(export_folder)) dir.create(export_folder, recursive = TRUE)

saveRDS(object = seqtab_nochim,
        file = file.path(export_folder, "seqtab_nochim.rds"))

saveRDS(object = taxonomy,
        file = file.path(export_folder, "taxonomy.rds"))
```

#verifie que un dossier existe sinon il le cree

```
asv_seq <- colnames(seqtab_nochim)
```

#stocke les ASV dans la valeure asv_seq

```
ndigits <- nchar(length(asv_seq))
asv_id <- sprintf(paste0("ASV_%0", ndigits, "d"), seq_along(asv_seq))
```

#permet de generer des identifiants pour les asv

```
row.names(taxonomy) <- colnames(seqtab_nochim) <- names(asv_seq) <- asv_id
```

#affecter des identifiants au asv

```
taxonomy_export <- df_export(taxonomy, new_rn = "asv")
seqtab_nochim_export <- t(seqtab_nochim)
seqtab_nochim_export <- df_export(seqtab_nochim_export, new_rn = "asv")
```

#exporte les donnees de taxonomy et seqtab_nochim en s'assurant que les noms sont bien rangés

```
write.table(taxonomy_export, file = file.path(export_folder, "taxonomy.tsv"),quote = FALSE,sep = "\t",r
```

```
write.table(seqtab_nochim_export,file = file.path(export_folder, "asv_table.tsv"),quote = FALSE,sep = "
```

#ecrit les donnes de taxonomy_export dans un fichier TSV, permettant de partager ou archiver les donnees taxonommiques

```
cat(paste0(">", names(asv_seq), "\n", asv_seq), sep = "\n", file = file.path(export_folder, "asv.fasta"
```

#ecrit les ASV et leurs noms dans fichier fasta

```
getN <- function(x) sum(dada2::getUniques(x))

log_table <- data.frame(
  input = primer_log$in_reads,
  with_fwd_primer = primer_log$`w/adapters`,
  with_rev_primer = primer_log$`w/adapters2` ,
  with_both_primers = out[, 1],
  filtered = out[, 2],
  denoisedF = sapply(dadaFs, getN),
  denoisedR = sapply(dadaRs, getN),
  merged = sapply(mergers, getN),
  nonchim = rowSums(seqtab_nochim),
  perc_retained = rowSums(seqtab_nochim) / out[, 1] * 100
)

rownames(log_table) <- sample_names
```

#permet d'evaluer et de quantifier la quantitee de sequences a chaque etape de l'analyse

```r
df_export(log_table, new_rn = "sample") |>
  write.table(file = file.path(export_folder, "log_table.tsv"),quote = FALSE,sep = "\t",row.names = FALS
```

#exporte log_table vers un fichier TSV, recapitulant les differentes etapes du traitement des donnees

#version fonctionnant sur ma machine perso.