

16S V4 region jejunum VS caecum

Anna Jelinkova

2024-09-17

R Markdown

```
path <- "E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files" # CHANGE ME to the directory containing the fastq files
list.files(path)
```

```
## [1] "cec_21_1_R1.fq" "cec_21_1_R2.fq" "cec_21_2_R1.fq" "cec_21_2_R2.fq"
## [5] "cec_21_3_R1.fq" "cec_21_3_R2.fq" "cec_21_4_R1.fq" "cec_21_4_R2.fq"
## [9] "cec_ctrl_1_R1.fq" "cec_ctrl_1_R2.fq" "cec_ctrl_2_R1.fq" "cec_ctrl_2_R2.fq"
## [13] "cec_ctrl_3_R1.fq" "cec_ctrl_3_R2.fq" "cec_ctrl_4_R1.fq" "cec_ctrl_4_R2.fq"
## [17] "cec_wjl_1_R1.fq" "cec_wjl_1_R2.fq" "cec_wjl_2_R1.fq" "cec_wjl_2_R2.fq"
## [21] "cec_wjl_3_R1.fq" "cec_wjl_3_R2.fq" "filtered_F" "filtered_R"
## [25] "jej_21_1_R1.fq" "jej_21_1_R2.fq" "jej_21_2_R1.fq" "jej_21_2_R2.fq"
## [29] "jej_21_3_R1.fq" "jej_21_3_R2.fq" "jej_21_4_R1.fq" "jej_21_4_R2.fq"
## [33] "jej_ctrl_1_R1.fq" "jej_ctrl_1_R2.fq" "jej_ctrl_2_R1.fq" "jej_ctrl_2_R2.fq"
## [37] "jej_ctrl_3_R1.fq" "jej_ctrl_3_R2.fq" "jej_ctrl_4_R1.fq" "jej_ctrl_4_R2.fq"
## [41] "jej_wjl_1_R1.fq" "jej_wjl_1_R2.fq" "jej_wjl_2_R1.fq" "jej_wjl_2_R2.fq"
## [45] "jej_wjl_3_R1.fq" "jej_wjl_3_R2.fq" "mock_com_R1.fq" "mock_com_R2.fq"
## [49] "neg_isol_R1.fq" "neg_isol_R2.fq"
```

Forward and reverse fastq filenames have format: SAMPLENAME_R1.fastq and SAMPLENAME_R2.fastq

```
fnFs <- sort(list.files(path, pattern="_R1.fq", full.names = TRUE))
fnRs <- sort(list.files(path, pattern="_R2.fq", full.names = TRUE))
```

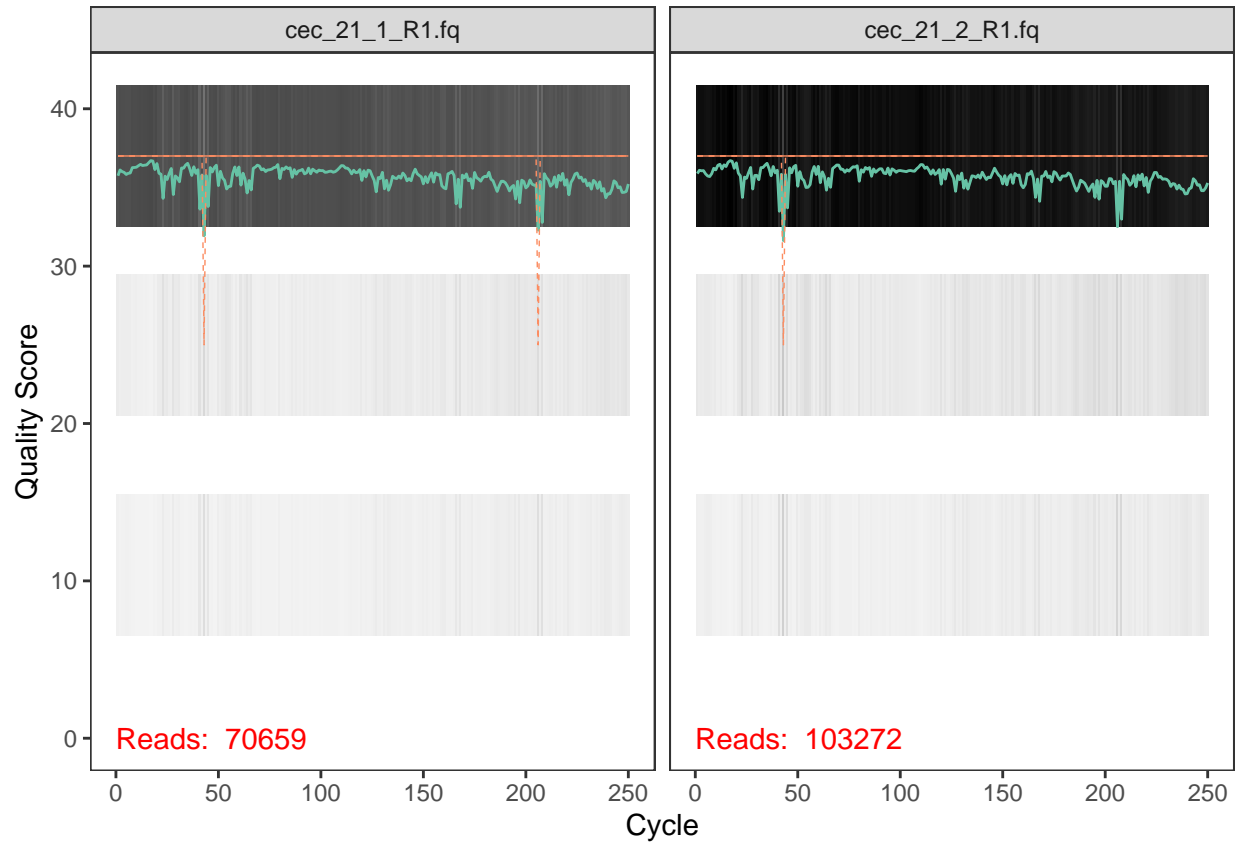
Extract sample names, assuming filenames have format: SAMPLENAME_XXX.fastq

Place filtered files in filtered/ subdirectory

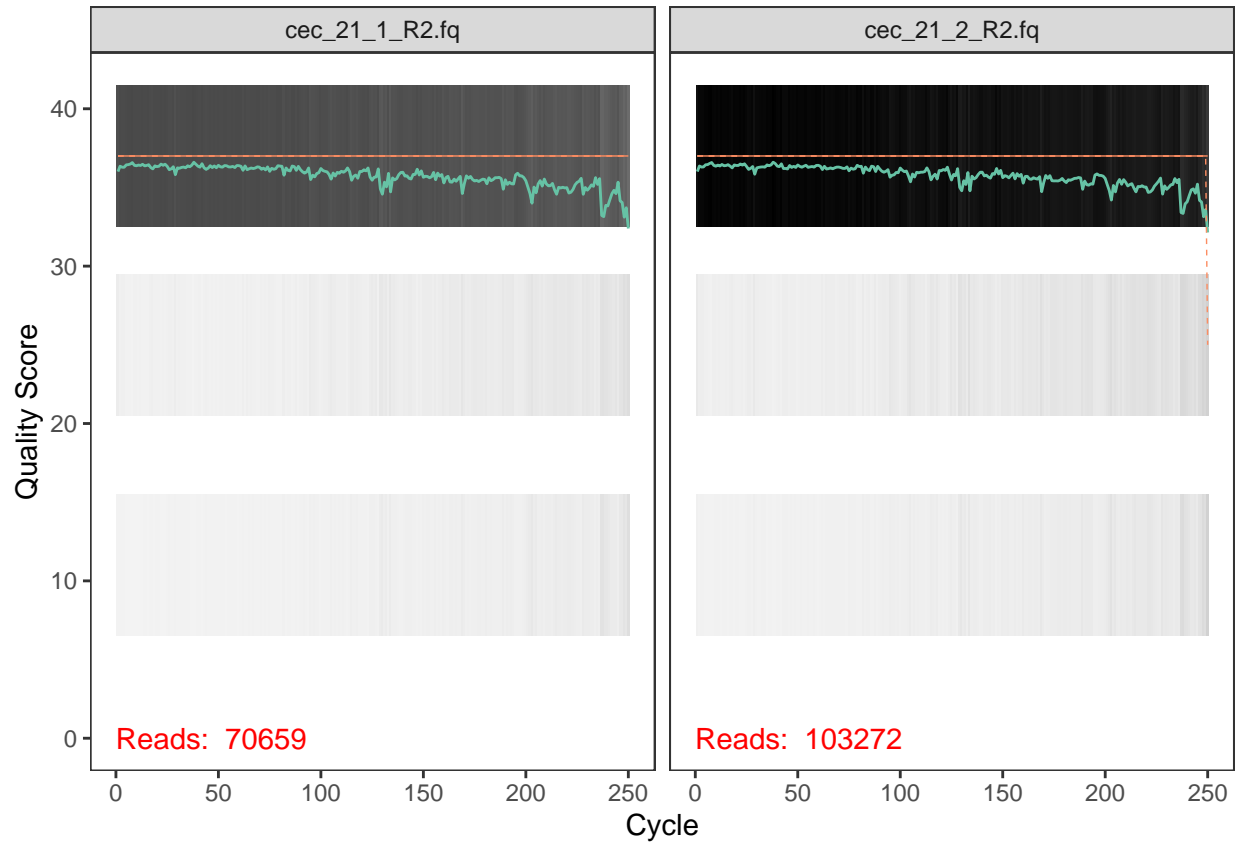
```
filtFs <- file.path(path, "filtered_F", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path, "filtered_R", paste0(sample.names, "_R_filt.fastq.gz"))
```

Plot first two samples- forward reads and reverse reads

```
plotQualityProfile(fnFs[1:2])
```



```
plotQualityProfile(fnRs[1:2])
```



Filter and trim reads:

```
outF <- filterAndTrim(fnFs, filtFs, truncLen=250, maxN=0, maxEE=2, truncQ=2, rm.phix=TRUE, compress=TRUE)
outR <- filterAndTrim(fnRs, filtRs, truncLen=250, maxN=0, maxEE=2, truncQ=2, rm.phix=TRUE, compress=TRUE)
outF
```

##	reads.in	reads.out
## cec_21_1_R1.fq	70659	64954
## cec_21_2_R1.fq	103272	95506
## cec_21_3_R1.fq	44413	40942
## cec_21_4_R1.fq	57326	52625
## cec_ctrl_1_R1.fq	100400	92891
## cec_ctrl_2_R1.fq	64417	59330
## cec_ctrl_3_R1.fq	78310	71781
## cec_ctrl_4_R1.fq	74441	68393
## cec_wjl_1_R1.fq	72629	66787
## cec_wjl_2_R1.fq	69233	63538
## cec_wjl_3_R1.fq	53434	49296
## jej_21_1_R1.fq	60289	54972
## jej_21_2_R1.fq	73602	67246
## jej_21_3_R1.fq	59620	54914
## jej_21_4_R1.fq	72559	66557
## jej_ctrl_1_R1.fq	53040	48413
## jej_ctrl_2_R1.fq	51925	47660
## jej_ctrl_3_R1.fq	63385	58167
## jej_ctrl_4_R1.fq	64040	59112
## jej_wjl_1_R1.fq	51851	47773

```
## jej_wjl_2_R1.fq      61381      56142
## jej_wjl_3_R1.fq      58765      53626
## mock_com_R1.fq       52352      48173
## neg_isol_R1.fq       319         288
```

```
outR
```

```
##                reads.in reads.out
## cec_21_1_R2.fq      70659      65966
## cec_21_2_R2.fq     103272      96783
## cec_21_3_R2.fq      44413      41474
## cec_21_4_R2.fq      57326      53320
## cec_ctrl_1_R2.fq   100400      93888
## cec_ctrl_2_R2.fq    64417      60032
## cec_ctrl_3_R2.fq    78310      72617
## cec_ctrl_4_R2.fq    74441      68984
## cec_wjl_1_R2.fq     72629      67379
## cec_wjl_2_R2.fq     69233      64381
## cec_wjl_3_R2.fq     53434      49884
## jej_21_1_R2.fq      60289      56018
## jej_21_2_R2.fq      73602      68008
## jej_21_3_R2.fq      59620      55254
## jej_21_4_R2.fq      72559      67075
## jej_ctrl_1_R2.fq    53040      49250
## jej_ctrl_2_R2.fq    51925      48308
## jej_ctrl_3_R2.fq    63385      58647
## jej_ctrl_4_R2.fq    64040      59497
## jej_wjl_1_R2.fq     51851      48361
## jej_wjl_2_R2.fq     61381      56561
## jej_wjl_3_R2.fq     58765      54316
## mock_com_R2.fq      52352      48449
## neg_isol_R2.fq       319         295
```

Dereplication step:

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

```
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.
```

```
## Encountered 51662 unique sequences from 64954 total sequences read.
```

```
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.
```

```
## Encountered 73491 unique sequences from 95506 total sequences read.
```

```
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.
```

```
## Encountered 33317 unique sequences from 40942 total sequences read.
```

```
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.
```

Encountered 42758 unique sequences from 52625 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 73627 unique sequences from 92891 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 48554 unique sequences from 59330 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 56928 unique sequences from 71781 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 54935 unique sequences from 68393 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 52849 unique sequences from 66787 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 49433 unique sequences from 63538 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/cec.

Encountered 38871 unique sequences from 49296 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jej.

Encountered 39621 unique sequences from 54972 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jej.

Encountered 49600 unique sequences from 67246 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jej.

Encountered 39955 unique sequences from 54914 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jej.

Encountered 45632 unique sequences from 66557 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jej.

```
## Encountered 33468 unique sequences from 48413 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 34007 unique sequences from 47660 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 40309 unique sequences from 58167 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 40041 unique sequences from 59112 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 33384 unique sequences from 47773 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 35866 unique sequences from 56142 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/jeje
## Encountered 34666 unique sequences from 53626 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/mocj
## Encountered 33482 unique sequences from 48173 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_F/negj
## Encountered 260 unique sequences from 288 total sequences read.
```

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

```
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cece
## Encountered 55392 unique sequences from 65966 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cece
## Encountered 80223 unique sequences from 96783 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cece
## Encountered 35262 unique sequences from 41474 total sequences read.
```

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 45088 unique sequences from 53320 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 78980 unique sequences from 93888 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 51571 unique sequences from 60032 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 61518 unique sequences from 72617 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 58571 unique sequences from 68984 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 57171 unique sequences from 67379 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 53623 unique sequences from 64381 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/cec.
Encountered 41615 unique sequences from 49884 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jej.
Encountered 44784 unique sequences from 56018 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jej.
Encountered 56260 unique sequences from 68008 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jej.
Encountered 44912 unique sequences from 55254 total sequences read.

Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jej.
Encountered 52803 unique sequences from 67075 total sequences read.

```

## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 38505 unique sequences from 49250 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 37982 unique sequences from 48308 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 46296 unique sequences from 58647 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 46572 unique sequences from 59497 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 38808 unique sequences from 48361 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 43293 unique sequences from 56561 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/jeje
## Encountered 41872 unique sequences from 54316 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/mocj
## Encountered 39328 unique sequences from 48449 total sequences read.
## Dereplicating sequence entries in Fastq file: E:/MBU/SEQMe_LACTOBACILLUS_16S/fq_files/filtered_R/negj
## Encountered 264 unique sequences from 295 total sequences read.

```

DADA OPTION 1: Skipping pre-training error rates

```
dadaFs <- dada(derep=filtFs, err = NULL, selfConsist = TRUE, pool = FALSE, verbose = TRUE)
```

```

## Initializing error rates to maximum possible estimate.
## selfConsist step 1 .....
##   selfConsist step 2
##   selfConsist step 3
##   selfConsist step 4
##   selfConsist step 5
##   selfConsist step 6
##   selfConsist step 7
##   selfConsist step 8
##   selfConsist step 9
## Convergence after 9 rounds.

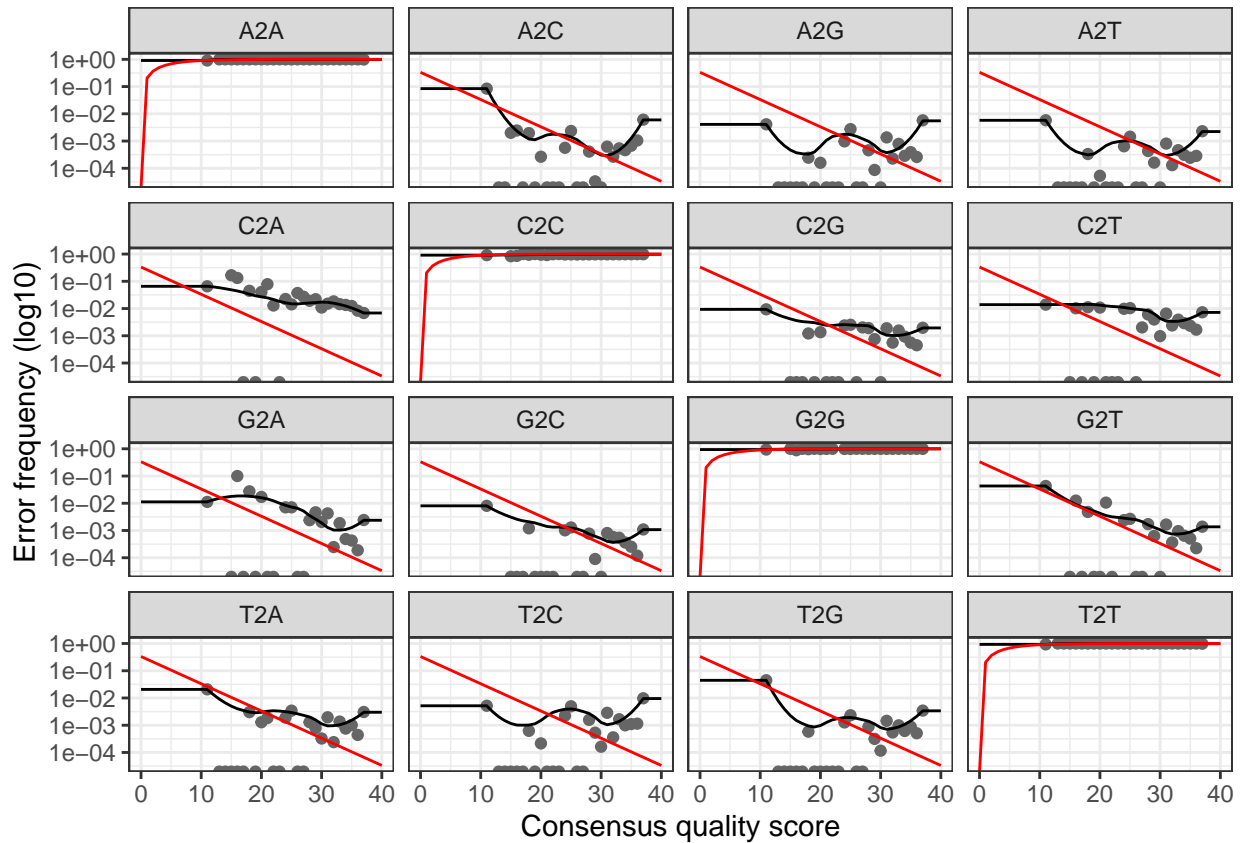
```



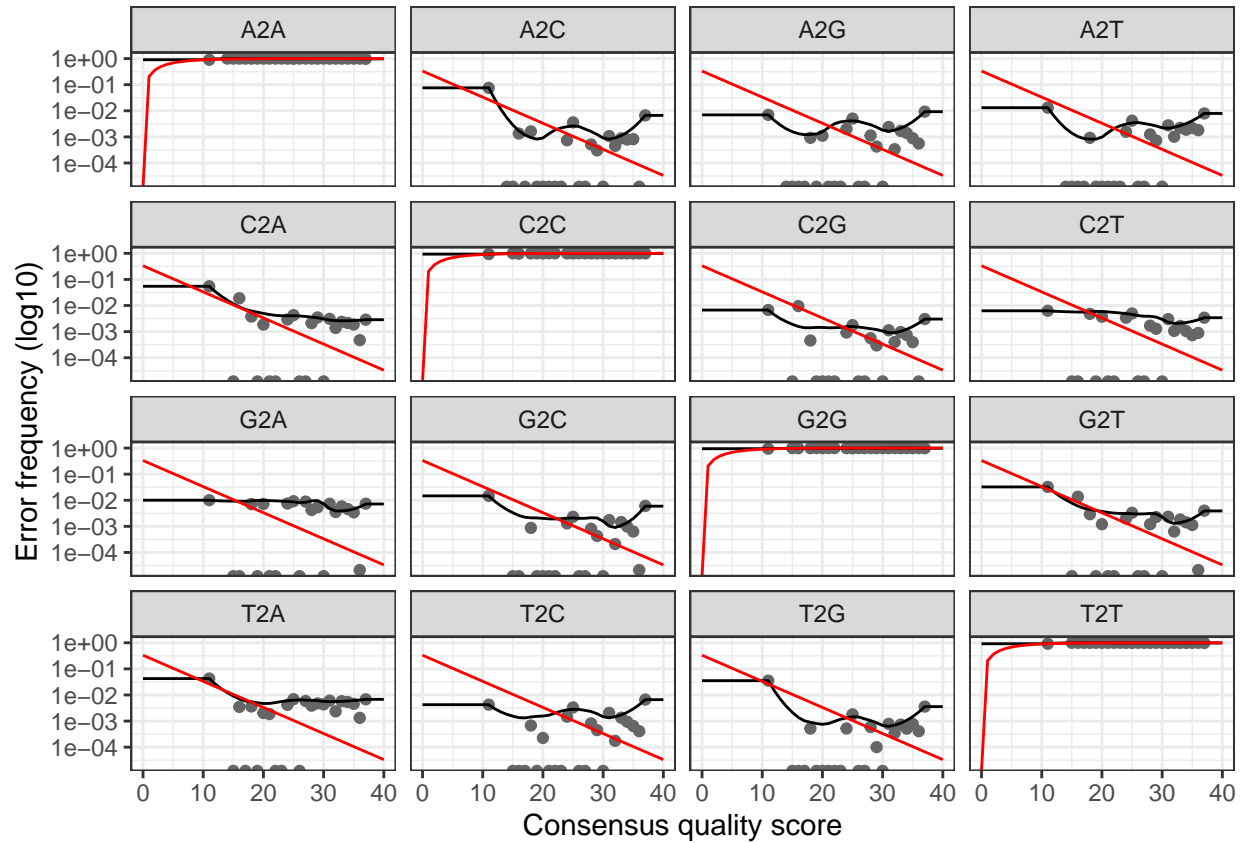
```
dadaRs <- dada(derep=filtRs, err = NULL, selfConsist = TRUE, pool = FALSE, verbose = TRUE)
```

```
## Initializing error rates to maximum possible estimate.
## selfConsist step 1 .....
## selfConsist step 2
## selfConsist step 3
## selfConsist step 4
## selfConsist step 5
## selfConsist step 6
## selfConsist step 7
## selfConsist step 8
## Convergence after 8 rounds.
```

```
plotErrors(dadaFs, nominalQ=TRUE)
```



```
plotErrors(dadaRs, nominalQ=TRUE)
```

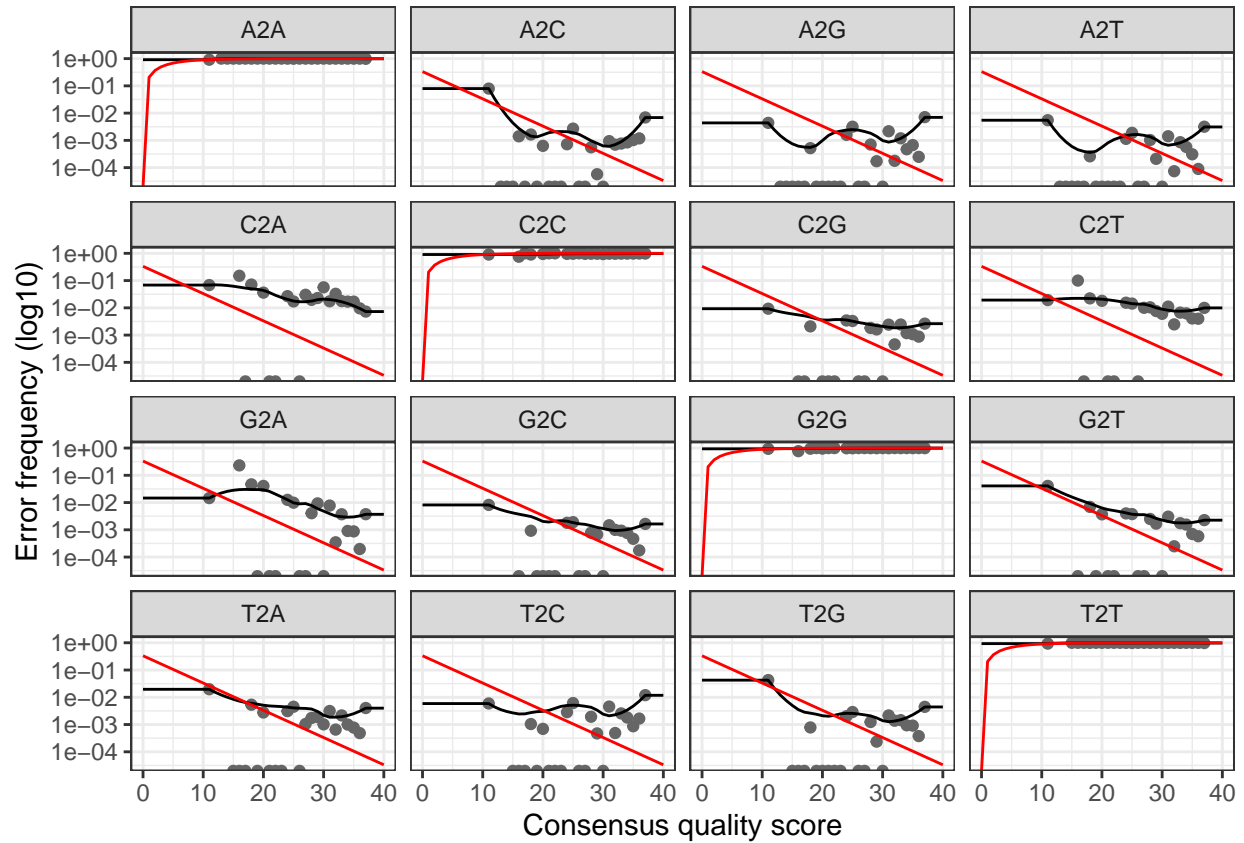


DADA OPTION 2: Learn Errors for forward reads and run DADA with pre-trained model

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

101562000 total bases in 406248 reads from 6 samples will be used for learning the error rates.

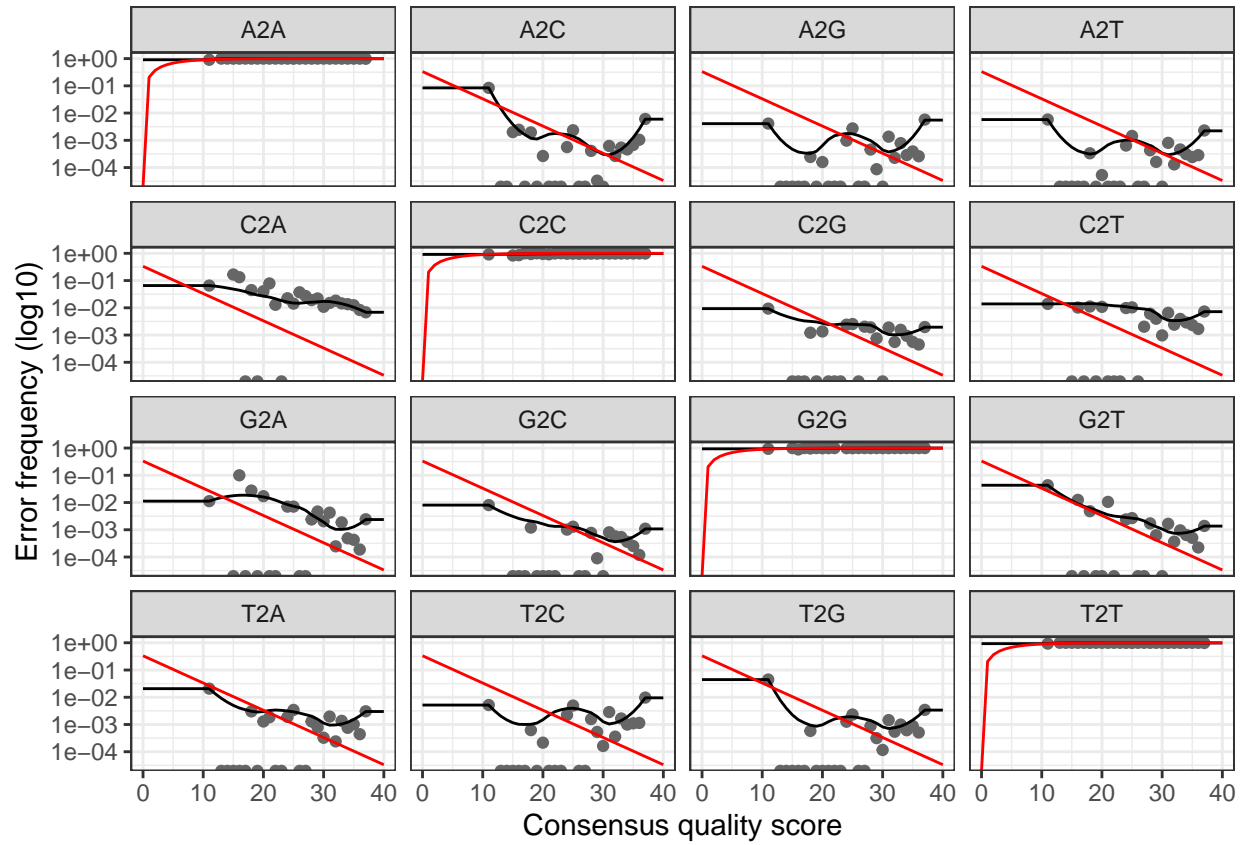
```
plotErrors(errF, nominalQ=TRUE)
```



```
dadaFs_errF <- dada(derep=filtFs, err = errF ,selfConsist = TRUE, pool = FALSE, verbose = TRUE)
```

```
## selfConsist step 1 .....
##   selfConsist step 2
##   selfConsist step 3
##   selfConsist step 4
##   selfConsist step 5
##   selfConsist step 6
##   selfConsist step 7
## Convergence after 7 rounds.
```

```
plotErrors(dadaFs_errF, nominalQ=TRUE)
```



```
#Learn Errors for reverse reads -does not work! #{r, include=TRUE, echo=TRUE, warning=FALSE}
#errR <- learnErrors(filtRs, multithread = TRUE, verbose = TRUE) #plotErrors(errR, nominalQ=TRUE)
#
```