

Homework 3

Rosemary Kinuthia

2/21/2018

```
#proof that I have installed the required packages  
library(dada2); packageVersion("dada2")
```

```
## [1] '1.6.0'
```

```
library(phyloseq); packageVersion("phyloseq");
```

```
## [1] '1.22.3'
```

```
library(ShortRead); packageVersion("shortread");
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##  
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':  
##  
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':  
##  
## IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':  
##  
## anyDuplicated, append, as.data.frame, cbind, colMeans,  
## colnames, colSums, do.call, duplicated, eval, evalq, Filter,  
## Find, get, grep, grepl, intersect, is.unsorted, lapply,  
## lengths, Map, mapply, match, mget, order, paste, pmax,  
## pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce,  
## rowMeans, rownames, rowSums, sapply, setdiff, sort, table,  
## tapply, union, unique, unsplit, which, which.max, which.min
```

```
## Loading required package: BiocParallel
```

```
## Loading required package: Biostrings
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##  
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:base':  
##  
## expand.grid
```

```
## Loading required package: IRanges
```

```
##  
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:phyloseq':  
##  
## distance
```

```
## Loading required package: XVector
```

```
##  
## Attaching package: 'Biostrings'
```

```
## The following object is masked from 'package:base':  
##  
##      strsplit
```

```
## Loading required package: Rsamtools
```

```
## Loading required package: GenomeInfoDb
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: GenomicAlignments
```

```
## Loading required package: SummarizedExperiment
```

```
## Loading required package: Biobase
```

```
## Welcome to Bioconductor  
##  
##      Vignettes contain introductory material; view with  
##      'browseVignettes()'. To cite Bioconductor, see  
##      'citation("Biobase)", and for packages 'citation("pkgname)".
```

```
##  
## Attaching package: 'Biobase'
```

```
## The following object is masked from 'package:phyloseq':  
##  
##      sampleNames
```

```
## Loading required package: DelayedArray
```

```
## Loading required package: matrixStats
```

```
##  
## Attaching package: 'matrixStats'
```

```
## The following objects are masked from 'package:Biobase':  
##  
## anyMissing, rowMedians
```

```
##  
## Attaching package: 'DelayedArray'
```

```
## The following objects are masked from 'package:matrixStats':  
##  
## colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges
```

```
## The following object is masked from 'package:Biostrings':  
##  
## type
```

```
## The following object is masked from 'package:base':  
##  
## apply
```

```
## [1] '1.36.0'
```

```
library(ggplot2); packageVersion("ggplot2");
```

```
## [1] '2.2.1'
```

```
# Proof that I have downloaded MiseqSOP, RDP training set, RDP species on my computer  
path <- "~/Desktop/N741 Big Data/Datasets/MiSeq_SOP"  
fns <- list.files(path)  
fns
```

```
## [1] "F3D0_S188_L001_R1_001.fastq" "F3D0_S188_L001_R2_001.fastq"
## [3] "F3D1_S189_L001_R1_001.fastq" "F3D1_S189_L001_R2_001.fastq"
## [5] "F3D141_S207_L001_R1_001.fastq" "F3D141_S207_L001_R2_001.fastq"
## [7] "F3D142_S208_L001_R1_001.fastq" "F3D142_S208_L001_R2_001.fastq"
## [9] "F3D143_S209_L001_R1_001.fastq" "F3D143_S209_L001_R2_001.fastq"
## [11] "F3D144_S210_L001_R1_001.fastq" "F3D144_S210_L001_R2_001.fastq"
## [13] "F3D145_S211_L001_R1_001.fastq" "F3D145_S211_L001_R2_001.fastq"
## [15] "F3D146_S212_L001_R1_001.fastq" "F3D146_S212_L001_R2_001.fastq"
## [17] "F3D147_S213_L001_R1_001.fastq" "F3D147_S213_L001_R2_001.fastq"
## [19] "F3D148_S214_L001_R1_001.fastq" "F3D148_S214_L001_R2_001.fastq"
## [21] "F3D149_S215_L001_R1_001.fastq" "F3D149_S215_L001_R2_001.fastq"
## [23] "F3D150_S216_L001_R1_001.fastq" "F3D150_S216_L001_R2_001.fastq"
## [25] "F3D2_S190_L001_R1_001.fastq" "F3D2_S190_L001_R2_001.fastq"
## [27] "F3D3_S191_L001_R1_001.fastq" "F3D3_S191_L001_R2_001.fastq"
## [29] "F3D5_S193_L001_R1_001.fastq" "F3D5_S193_L001_R2_001.fastq"
## [31] "F3D6_S194_L001_R1_001.fastq" "F3D6_S194_L001_R2_001.fastq"
## [33] "F3D7_S195_L001_R1_001.fastq" "F3D7_S195_L001_R2_001.fastq"
## [35] "F3D8_S196_L001_R1_001.fastq" "F3D8_S196_L001_R2_001.fastq"
## [37] "F3D9_S197_L001_R1_001.fastq" "F3D9_S197_L001_R2_001.fastq"
## [39] "HMP MOCK.v35.fasta" "Mock_S280_L001_R1_001.fastq"
## [41] "Mock_S280_L001_R2_001.fastq" "mouse.dpw.metadata"
## [43] "mouse.time.design" "silva_nr_v132_train_set.fa"
## [45] "silva_species_assignment_v132.fa" "stability.batch"
## [47] "stability.files"
```

#The Github repository for this assignment can be accessed via this link <https://github.com/RosemaryKinuthia/N741-Homework-3.git>