# Create MRexp and Phyloseq Objects

*Nate Olson*

*2017-03-16*

Generating MRexperiment and phyloseq objects for the three pipelines.

## Metadata

```
meta_df <- sampleSheet %>%
  filter(barcode_lab == "JHU", seq_lab == "JHU") %>%
  select(biosample_id, titration, t_fctr, pcr_16S_plate, pos) %>%
  unite(id, pcr_16S_plate, pos, sep = "-",remove = FALSE)  %>%
  as.data.frame()

rownames(meta_df) <- meta_df$id
```

```
glimpse(meta_df)
```

```
## Observations: 192
## Variables: 6
## $ biosample_id  <chr> "E01JH0004", "E01JH0004", "E01JH0004", "E01JH000...
## $ titration     <dbl> 0, 1, 2, 3, 4, 5, 10, 15, 20, 0, 1, 2, 3, 4, 5, ...
## $ t_fctr        <fctr> 20, 1, 2, 3, 4, 5, 10, 15, 0, 20, 1, 2, 3, 4, 5...
## $ id            <chr> "1-A1", "1-B1", "1-C1", "1-D1", "1-E1", "1-F1", ...
## $ pcr_16S_plate <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ pos           <chr> "A1", "B1", "C1", "D1", "E1", "F1", "G1", "H1", ...
```

## Mothur

### Mothur Phyloseq Object

```
shared_file <- "mothur/mgtst.trim.contigs.good.unique.good.filter.unique.precluster.pick.opti_mcc.unique

constax_file <- "mothur/mgtst.trim.contigs.good.unique.good.filter.unique.precluster.pick.opti_mcc.uniq

mothur_ps <- phyloseq::import_mothur(mothur_shared_file = shared_file,
                        mothur_constaxonomy_file = constax_file)

## Adding sample data
phyloseq::sample_data(mothur_ps) <- meta_df
```

```
mothur_ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 38358 taxa and 192 samples ]
## sample_data() Sample Data:       [ 192 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 38358 taxa by 6 taxonomic ranks ]
```

### Mothur MRexperiment Object

```
mothur_mrexp <- phyloseq_to_metagenomeSeq(mothur_ps)
```

```
mothur_mrexp
```

```
## MRexperiment (storageMode: environment)
## assayData: 38358 features, 192 samples
##   element names: counts
## protocolData: none
## phenoData
##   sampleNames: 1-A1 1-A10 ... 2-H9 (192 total)
##   varLabels: biosample_id titration ... pos (6 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: Otu00001 Otu00002 ... Otu38358 (38358 total)
##   fvarLabels: OTUname Rank1 ... Rank6 (7 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

### Save Mothur Objects

```
saveRDS(mothur_ps, "mothur/mothur_ps.rds")
saveRDS(mothur_mrexp, "mothur/mothur_mrexp.rds")
```

## QIIME

### QIIME Phyloseq Object

```
# qiime_dir <- "qiime/otus_uc_fast_no_chimera/"
qiime_dir <- "qiime/otus_uc_fast/"
biom_file <- file.path(qiime_dir, "otu_table_mc2_w_tax_no_pynast_failures.biom")
seq_file <- file.path(qiime_dir, "rep_set.fna")
tree_file <- file.path(qiime_dir, "rep_set.tre")

qiime_ps <- phyloseq::import_biom(BIOMfilename = biom_file)
```

```
## Warning in strsplit(msg, "\n"): input string 1 is invalid in this locale
```

```
#sample_names(qiime_ps) <- sample_names(qiime_ps) %>% str_replace("centroid=","")

## Adding sample data
phyloseq::sample_data(qiime_ps) <- meta_df
```

```
qiime_ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 11385 taxa and 192 samples ]
## sample_data() Sample Data:       [ 192 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 11385 taxa by 7 taxonomic ranks ]
```

### QIIME MRexperiment Object

```
qiime_mrexp <- phyloseq_to_metagenomeSeq(qiime_ps)
```

```
qiime_mrexp
```

```
## MRexperiment (storageMode: environment)
## assayData: 11385 features, 192 samples
##   element names: counts
## protocolData: none
## phenoData
##   sampleNames: 2-B5 2-E12 ... 2-D12 (192 total)
##   varLabels: biosample_id titration ... pos (6 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: 4333897 1036749 ... New.CleanUp.ReferenceOTU19782
##     (11385 total)
##   fvarLabels: OTUname Rank1 ... Rank7 (8 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

### Save QIIME Objects

```
saveRDS(qiime_ps, "qiime/qiime_ps.rds")
saveRDS(qiime_mrexp, "qiime/qiime_mrexp.rds")
```

### DADA2 Phyloseq Object

```
seqtab <- readRDS("dada2/seqtab_nochim.rds")
otu_tbl <- otu_table(seqtab, taxa_are_rows=FALSE)

## Rep sequences
sv_seqs <- colnames(otu_tbl)
names(sv_seqs) <- paste0("SV",1:ncol(otu_tbl))

## Rename features
colnames(otu_tbl) <- paste0("SV",1:ncol(otu_tbl))

taxa <- readRDS("dada2/taxa.rds")
rownames(taxa) <- names(sv_seqs)[match(sv_seqs, rownames(taxa))]

dada_ps <- phyloseq(otu_tbl,
                    sample_data(meta_df),
                    tax_table(taxa))

## Removing 0 entry samples
dada_samples <- sample_names(dada_ps)
dada_nonzero_sample <- dada_samples[phyloseq::sample_sums(dada_ps) != 0]
dada_ps <- phyloseq::prune_samples(dada_nonzero_sample, dada_ps)
```

```
dada_ps
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 3144 taxa and 191 samples ]
## sample_data() Sample Data:       [ 191 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 3144 taxa by 6 taxonomic ranks ]
```

### DADA2 MRexperiment Object

```
dada_mrexp <- phyloseq_to_metagenomeSeq(dada_ps)
```

```
dada_mrexp
```

```
## MRexperiment (storageMode: environment)
## assayData: 3144 features, 191 samples
##   element names: counts
## protocolData: none
## phenoData
##   sampleNames: 1-A1 1-A10 ... 2-H9 (191 total)
##   varLabels: biosample_id titration ... pos (6 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: SV1 SV2 ... SV3144 (3144 total)
##   fvarLabels: OTUname Kingdom ... Genus (7 total)
##   fvarMetadata: labelDescription
## experimentData: use 'experimentData(object)'
## Annotation:
```

### Save DADA2 Objects

```
saveRDS(dada_ps, "dada2/dada_ps.rds")
saveRDS(dada_mrexp, "dada2/dada_mrexp.rds")
DNAStringSet(sv_seqs) %>% writeXStringSet("dada2/sv_seqs.rds")
```

## Session information

```
s_info <- devtools::session_info()
print(s_info$platform)
```

```
##  setting  value
##  version  R version 3.3.2 (2016-10-31)
##  system   x86_64, darwin15.6.0
##  ui       unknown
##  language (EN)
##  collate  en_US.UTF-8
##  tz       America/New_York
##  date     2017-03-16
```

```
s_info$packages %>% filter(`*` == "*") %>% select(-`*`) %>%
      knitr::kable()
```

| package | version | date | source |
|---|---|---|---|
| bbmle | 1.0.18 | 2016-02-11 | CRAN (R 3.3.2) |
| Biobase | 2.34.0 | 2016-11-07 | Bioconductor |
| BiocGenerics | 0.20.0 | 2016-11-07 | Bioconductor |
| BiocParallel | 1.8.1 | 2016-11-07 | Bioconductor |
| Biostrings | 2.42.1 | 2016-12-19 | Bioconductor |
| DESeq | 1.26.0 | 2016-11-28 | Bioconductor |
| DESeq2 | 1.15.28 | 2017-02-02 | bioc (readonly/DESeq2@125913) |
| dplyr | 0.5.0 | 2016-06-24 | CRAN (R 3.3.2) |
| edgeR | 3.16.5 | 2017-02-02 | Bioconductor |
| forcats | 0.2.0 | 2017-01-23 | CRAN (R 3.3.2) |
| foreach | 1.4.3 | 2015-10-13 | CRAN (R 3.3.1) |
| GenomeInfoDb | 1.10.2 | 2017-01-04 | Bioconductor |
| GenomicAlignments | 1.10.0 | 2016-11-07 | Bioconductor |
| GenomicRanges | 1.26.2 | 2017-01-04 | Bioconductor |
| ggplot2 | 2.2.1 | 2016-12-30 | CRAN (R 3.3.2) |
| glmnet | 2.0-5 | 2016-03-17 | CRAN (R 3.3.1) |
| IRanges | 2.8.1 | 2016-11-18 | Bioconductor |
| knitr | 1.15.1 | 2016-11-22 | CRAN (R 3.3.2) |
| lattice | 0.20-34 | 2016-09-06 | CRAN (R 3.3.2) |
| limma | 3.30.9 | 2017-02-02 | Bioconductor |
| locfit | 1.5-9.1 | 2013-04-20 | CRAN (R 3.3.1) |
| Matrix | 1.2-8 | 2017-01-20 | CRAN (R 3.3.2) |
| metagenomeSeq | 1.16.0 | 2016-11-07 | Bioconductor |
| modelr | 0.1.0 | 2016-08-31 | cran (@0.1.0) |
| permute | 0.9-4 | 2016-09-09 | CRAN (R 3.3.1) |
| phyloseq | 1.19.1 | 2017-01-04 | Bioconductor |
| ProjectTemplate | 0.7 | 2016-08-11 | CRAN (R 3.3.1) |
| purrr | 0.2.2 | 2016-06-18 | CRAN (R 3.3.1) |
| RColorBrewer | 1.1-2 | 2014-12-07 | CRAN (R 3.3.1) |
| readr | 1.0.0 | 2016-08-03 | CRAN (R 3.3.1) |
| readxl | 0.1.1 | 2016-03-28 | cran (@0.1.1) |
| Rqc | 1.8.0 | 2016-11-07 | Bioconductor |
| Rsamtools | 1.26.1 | 2016-11-07 | Bioconductor |
| S4Vectors | 0.12.1 | 2016-12-19 | Bioconductor |
| sads | 0.3.1 | 2016-05-13 | CRAN (R 3.3.2) |
| savR | 1.12.0 | 2016-11-07 | Bioconductor |
| ShortRead | 1.32.0 | 2016-11-07 | Bioconductor |
| stringr | 1.1.0 | 2016-08-19 | CRAN (R 3.3.1) |
| SummarizedExperiment | 1.4.0 | 2016-11-07 | Bioconductor |
| tibble | 1.2 | 2016-08-26 | CRAN (R 3.3.1) |
| tidyr | 0.6.1 | 2017-01-10 | CRAN (R 3.3.2) |
| tidyverse | 1.1.1 | 2017-01-27 | CRAN (R 3.3.2) |
| vegan | 2.4-2 | 2017-01-17 | CRAN (R 3.3.2) |
| XVector | 0.14.0 | 2016-11-07 | Bioconductor |