Titration Abundance Expectation Calculation

Nate Olson

2017-04-04

Objective

Generate a data_frame with observed and predicted count values for individual PCR replicates. Expected values are calculated using unmixed PCR count table values from the same set of PCR replicates, same half of one of the replicate 96 well plates.

```
## Extracting a tidy dataframe with count values from MRexpiment objects
get_count_df <- function(mrobj){
      mrobj <- cumNorm(mrobj, p = 0.75)
      mrobj %>%
            # not sure whether or not to normalize counts prior to analysis
            MRcounts(norm = TRUE, log = FALSE, sl = 1000) %>%
            #MRcounts(sl = 1) %>%
            as.data.frame() %>%
            rownames_to_column(var = "feature_id") %>%
            gather("id","count", -feature_id)
}


count_df <- mrexp %>% map_df(get_count_df, .id = "pipe")


## Annotate with meta data and exclude NTCs
count_df <- pData(mrexp$dada2) %>% right_join(count_df) %>%
      filter(biosample_id != "NTC")
```

```
count_df %>% glimpse()
```

```
## Observations: 9,519,660
## Variables: 11
## $ biosample_id  <chr> "E01JH0004", "E01JH0004", "E01JH0004", "E01JH000...
## $ titration     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ t_fctr        <fctr> 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20,...
## $ id            <chr> "1-A1", "1-A1", "1-A1", "1-A1", "1-A1", "1-A1", ...
## $ pcr_16S_plate <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ pos           <chr> "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", ...
## $ pcr_half      <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"...
## $ pcr_rep       <chr> "1:1", "1:1", "1:1", "1:1", "1:1", "1:1", "1:1",...
## $ pipe          <chr> "dada2", "dada2", "dada2", "dada2", "dada2", "da...
## $ feature_id    <chr> "SV1", "SV2", "SV3", "SV4", "SV5", "SV6", "SV7",...
## $ count         <dbl> 101.765711, 968.767800, 73.381432, 271.976457, 4...
```

Extracting observed count values for unmixed pre and post samples

```
## Pre and post full features are not necessarily pre or post specific
post_count <- count_df %>%
      filter(t_fctr == 0) %>%
      select(pipe, biosample_id, feature_id, pcr_rep, count) %>%
      dplyr::rename(post_count = count)


pre_count <- count_df %>%
```

```r
    filter(t_fctr == 20) %>%
    select(pipe, biosample_id, feature_id, pcr_rep, count) %>%
    dplyr::rename(pre_count = count)

pre_post_count <-  full_join(pre_count, post_count) %>%
    mutate(pre_count = if_else(is.na(pre_count), 0, pre_count),
           post_count = if_else(is.na(post_count), 0, post_count))
```

Use observed count values and titration factor to calculate expected values for titrations.

```r
titration_list <- data_frame(titration = c(1:5,10,15)) %>%
    mutate(post_prop = 2^-titration) %>% list() %>% rep(nrow(pre_post_count))

titration_pred <- pre_post_count %>% ungroup() %>%
    add_column(titration = titration_list) %>% unnest() %>%
    mutate(exp_count = post_count * post_prop + pre_count * (1-post_prop))
```

```r
count_exp_df <- count_df %>%
    ungroup() %>%
    filter(titration %in% c(1:5,10,15)) %>%
    dplyr::rename(obs_count = count) %>%
    right_join(titration_pred) %>%
    mutate(residual = exp_count - obs_count)
```

```r
count_exp_df %>% glimpse()
```

```
## Observations: 7,404,180
## Variables: 16
## $ biosample_id  <chr> "E01JH0004", "E01JH0004", "E01JH0004", "E01JH000...
## $ titration     <dbl> 1, 2, 3, 4, 5, 10, 15, 1, 2, 3, 4, 5, 10, 15, 1,...
## $ t_fctr        <fctr> 1, 2, 3, 4, 5, 10, 15, 1, 2, 3, 4, 5, 10, 15, 1...
## $ id            <chr> "1-B1", "1-C1", "1-D1", "1-E1", "1-F1", "1-G1", ...
## $ pcr_16S_plate <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ pos           <chr> "B1", "C1", "D1", "E1", "F1", "G1", "H1", "B1", ...
## $ pcr_half      <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"...
## $ pcr_rep       <chr> "1:1", "1:1", "1:1", "1:1", "1:1", "1:1", "1:1",...
## $ pipe          <chr> "dada2", "dada2", "dada2", "dada2", "dada2", "da...
## $ feature_id    <chr> "SV1", "SV1", "SV1", "SV1", "SV1", "SV1", "SV1",...
## $ obs_count     <dbl> 1654.483638, 780.001283, 619.851624, 392.836731,...
## $ pre_count     <dbl> 101.76571, 101.76571, 101.76571, 101.76571, 101....
## $ post_count    <dbl> 1031.00775, 1031.00775, 1031.00775, 1031.00775, ...
## $ post_prop     <dbl> 5.000000e-01, 2.500000e-01, 1.250000e-01, 6.2500...
## $ exp_count     <dbl> 566.38673, 334.07622, 217.92097, 159.84334, 130....
## $ residual      <dbl> -1088.09691, -445.92506, -401.93066, -232.99339,...
```

```r
count_exp_df %>% saveRDS("../data/expected_count_values_df.rds")
```

Session information

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

```
##  setting  value
##  version  R version 3.3.3 (2017-03-06)
```

```
## system   x86_64, darwin15.6.0
## ui       unknown
## language (EN)
## collate   en_US.UTF-8
## tz        America/New_York
## date      2017-04-04
```

```
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	1.10.3	2017-03-28	Bioconductor
GenomicAlignments	1.10.1	2017-03-28	Bioconductor
GenomicRanges	1.26.4	2017-03-28	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	2.8.2	2017-03-28	Bioconductor
knitr	1.15.1	2016-11-22	CRAN (R 3.3.2)
lattice	0.20-34	2016-09-06	CRAN (R 3.3.3)
limma	3.30.13	2017-03-28	Bioconductor
locfit	1.5-9.1	2013-04-20	CRAN (R 3.3.1)
Matrix	1.2-8	2017-01-20	CRAN (R 3.3.3)
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	1.1.0	2017-03-22	CRAN (R 3.3.2)
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	0.12.2	2017-03-28	Bioconductor
sads	0.3.1	2016-05-13	CRAN (R 3.3.2)
savR	1.12.0	2016-11-07	Bioconductor
ShortRead	1.32.1	2017-03-28	Bioconductor
stringr	1.2.0	2017-02-18	CRAN (R 3.3.2)
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)

package	version	date	source
XVector	0.14.1	2017-03-28	Bioconductor