

Low CSSLog value check

Nate Olson

2017-02-22

Objective

Check on low csslog normalized count values.

Approach

Generate histograms of raw and log transformed count values

Loading data

Normalized count data from 2017-02-09-Normalization-df-Dev.Rmd

```
count_df <- readRDS("../data/normalize_count_df.rds")
```

Histograms

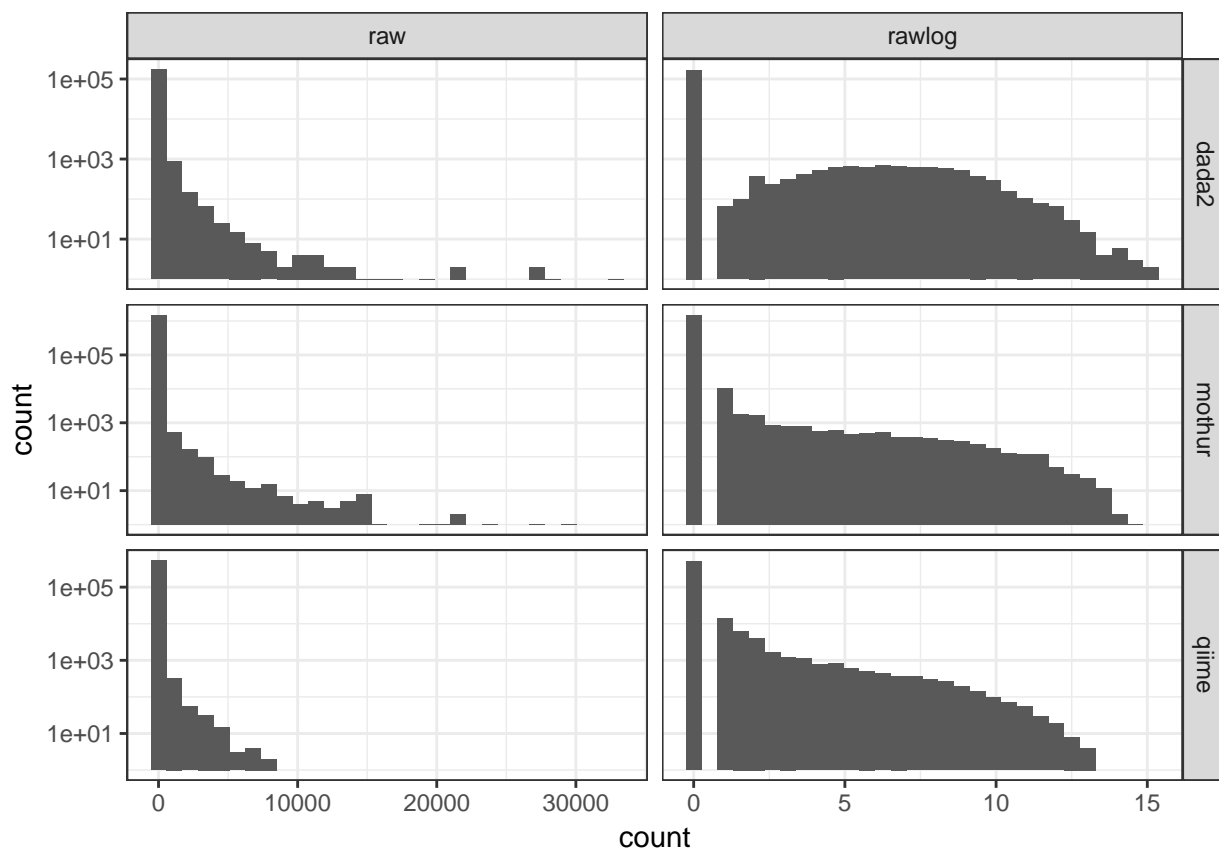
```
filt_counts <- count_df %>% filter(norm_method %in% c("raw", "rawlog"))
```

```
filt_counts %>% sample_frac(size = 0.25) %>%  
  ggplot() +  
  geom_histogram(aes(x = count)) +  
  facet_grid(pipe~norm_method, scales = "free") +  
  scale_y_log10() + theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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

```
## Warning: Removed 48 rows containing missing values (geom_bar).
```



CSS and CSSLog Transformation

CSS and CSS log transformation normalized count values calculated with $p = 0.75$ and $sl = 1$ (no additional scaling).

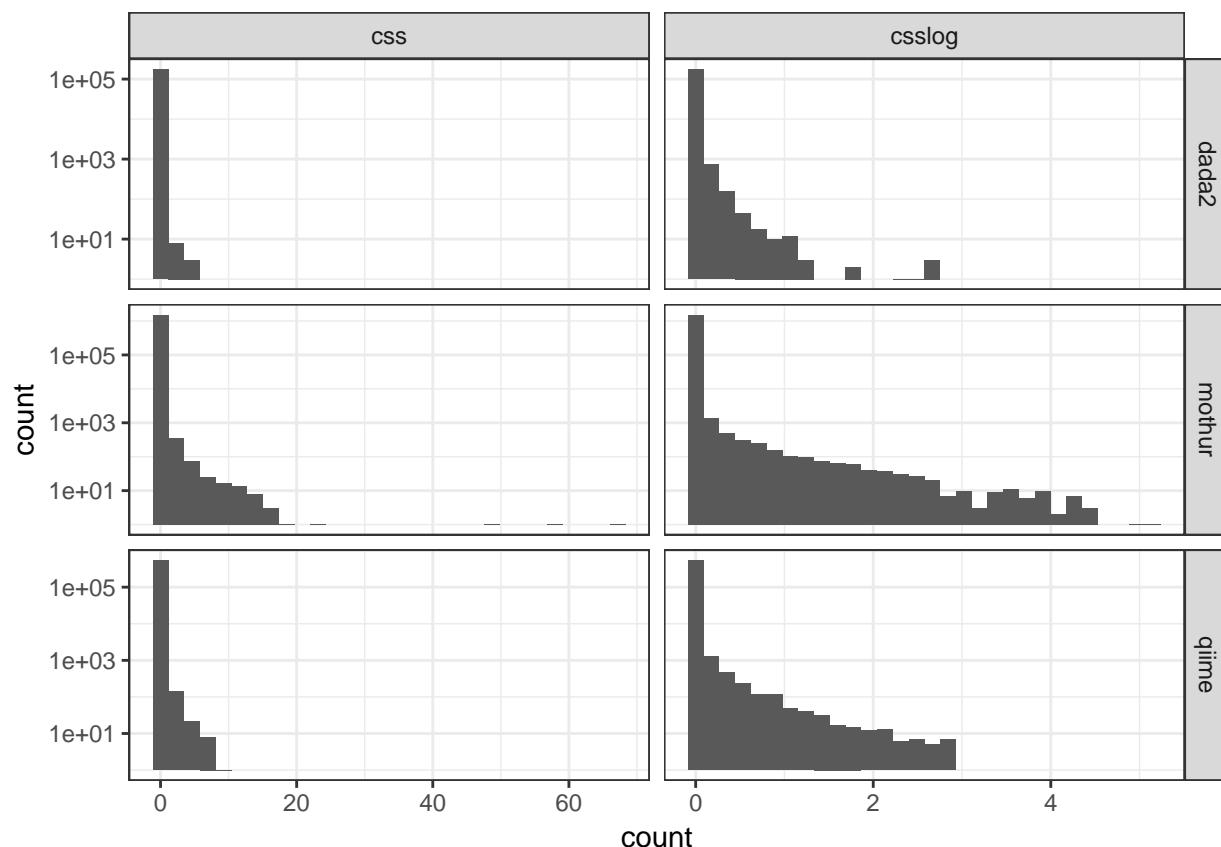
```
filt_counts <- count_df %>% filter(norm_method %in% c("css", "csslog"))
```

```
filt_counts %>% sample_frac(size = 0.25) %>%
  ggplot() +
  geom_histogram(aes(x = count)) +
  facet_grid(pipe~norm_method, scales = "free") +
  scale_y_log10() + theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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

```
## Warning: Removed 102 rows containing missing values (geom_bar).
```



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 2017-02-22
```

```
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)

package	version	date	source
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