PopPsiSeq Dev1

Charlie Soeder

11/14/2018

16 Nov 2018

experimental data:

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

Table 1: Sequenced Experimental Samples

name	paired	experimental	source
SRR5860570	TRUE	control	NCBI
SRR303333	FALSE	selection	EarlyJones2011
10A	TRUE	selection	EarlyJones2013

Population-wide sample count by species:

Table 2: Number of Sequenced Samples per Species

species	sample_count
drosophila sechellia	3
drosophila simulans	3

load & discuss FASTP summary

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_character(),
##   X4 = col_double()
## )
```
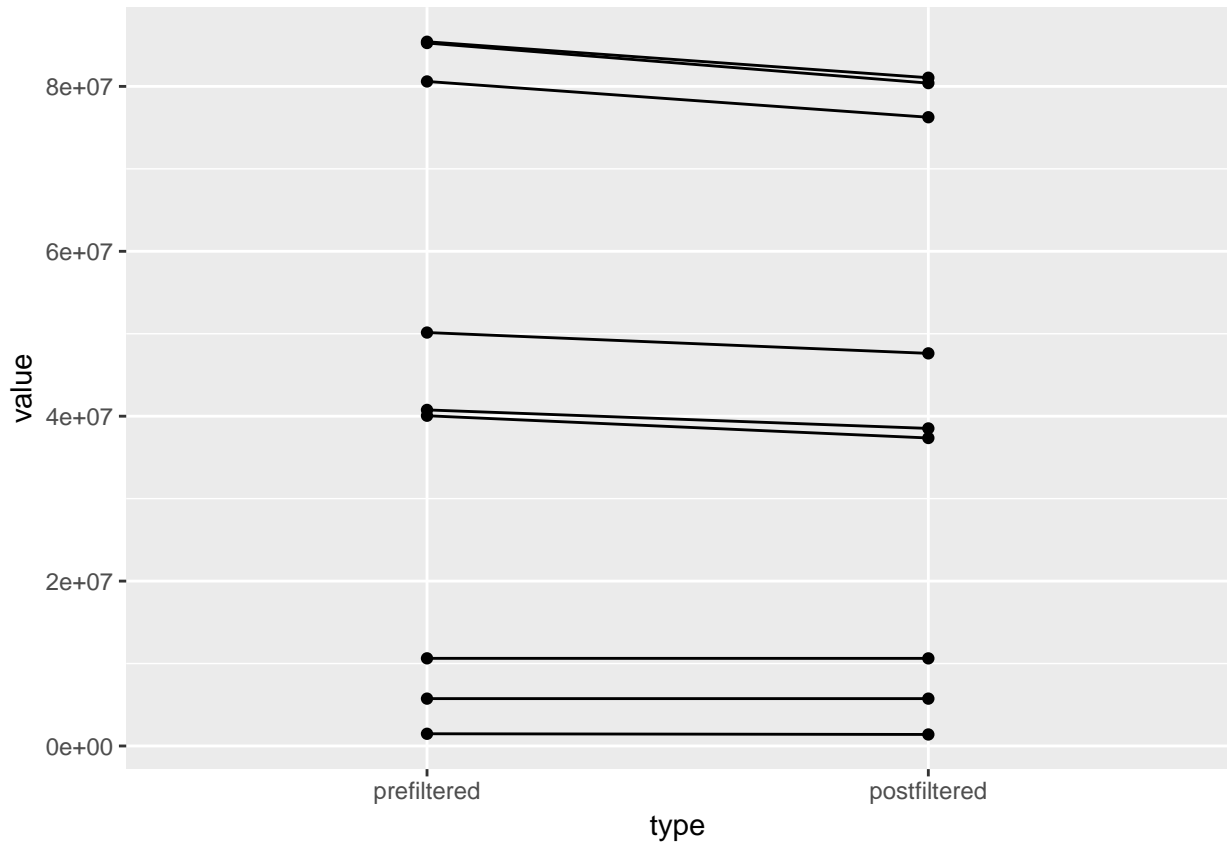
prefilt:

```
## Warning: Column `name` joining factors with different levels, coercing to
## character vector
```
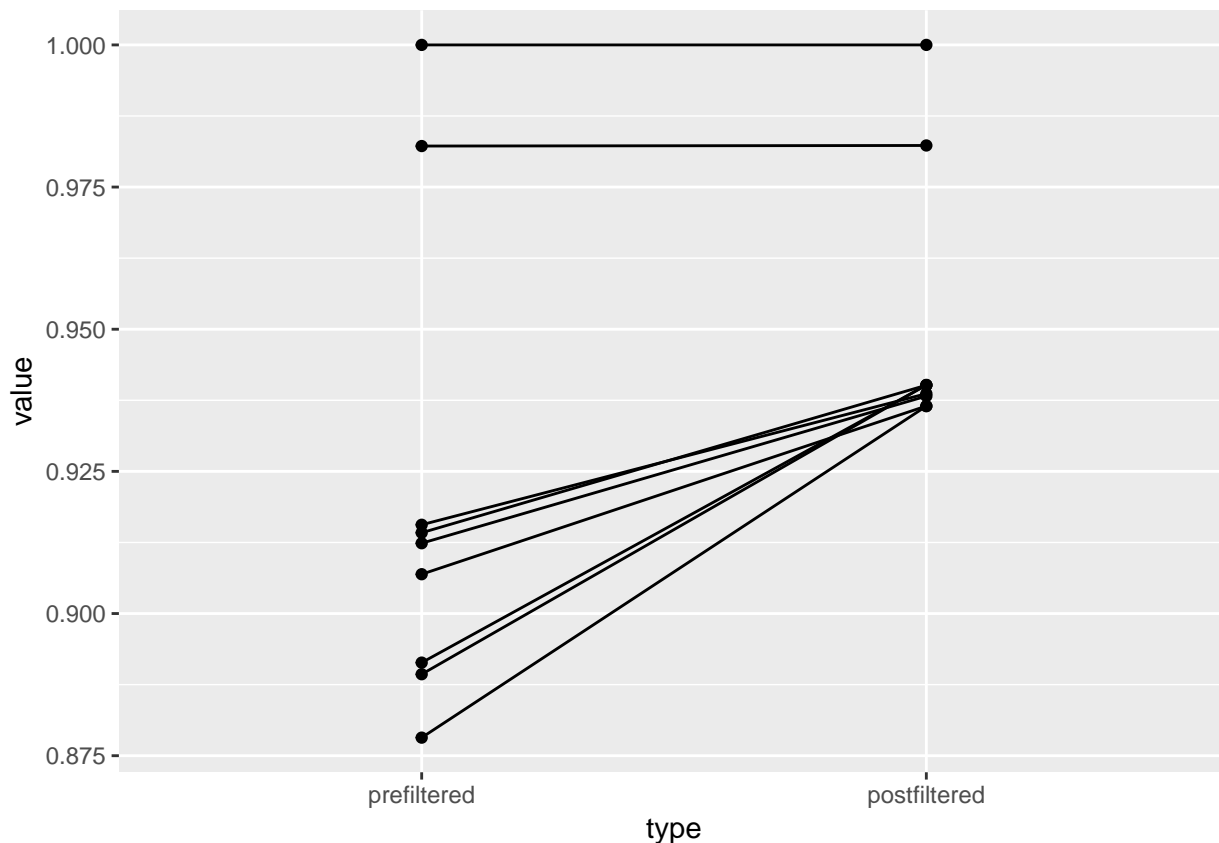
```
## # A tibble: 1 x 3
##   minimum  average  maximum
##     <dbl>    <dbl>    <dbl>
## 1 1481482 44455123. 85417202
```

```
## Warning: Column `name` joining factors with different levels, coercing to
## character vector
```

type	minimum	average	maximum
prefiltered	1.481482e+06	4.445512e+07	85417202
postfiltered	1.397152e+06	4.210807e+07	81052256

type	minimum	average	maximum
percent retention	9.326349e+01	9.564348e+01	100

```
## Warning: Column `name` joining factors with different levels, coercing to
## character vector
```

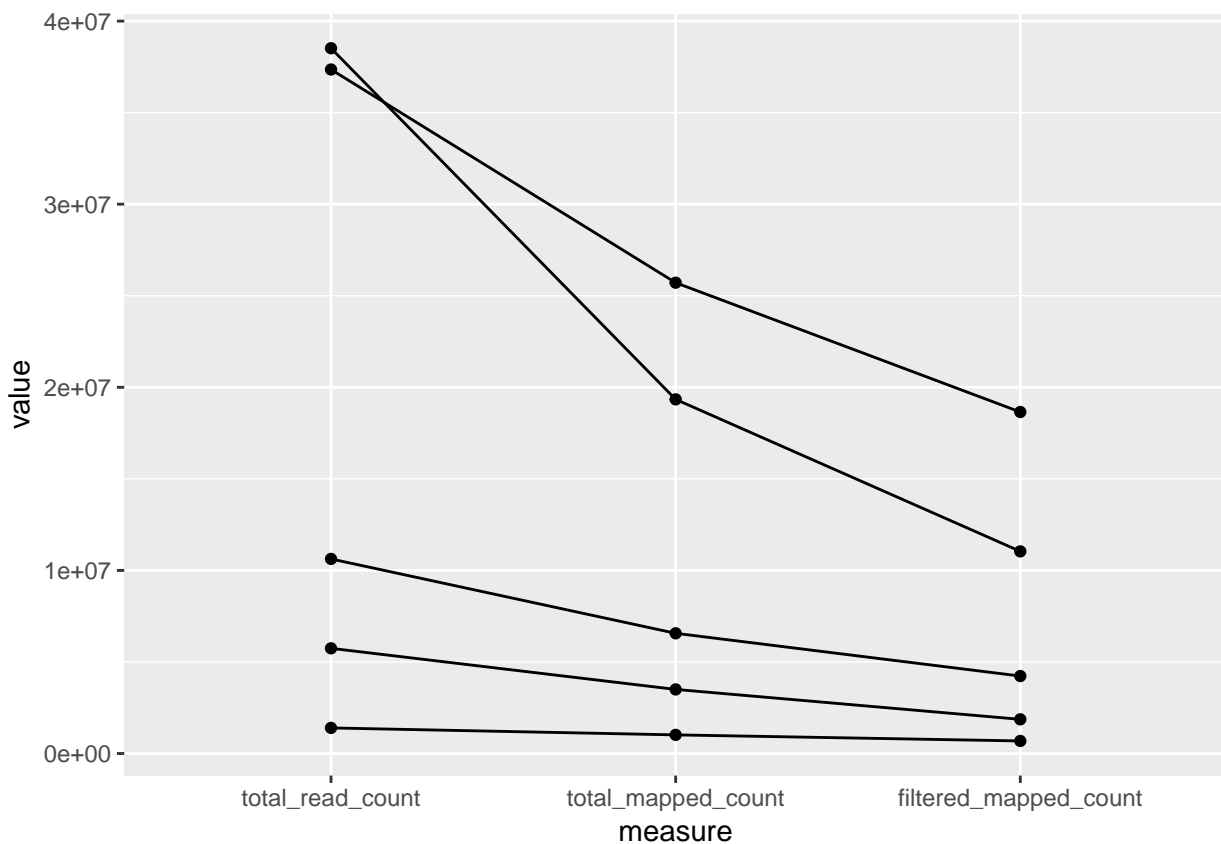## 19 Nov 2018

load and discuss bam summary

depth of coverage is effed???

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_double()
## )
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_double()
## )

## # A tibble: 10 x 4
##    sample   measure               value aligner
##    <fct>    <fct>                 <dbl> <fct>
##  1 10A      total_read_count    5743832 bwa
##  2 10A      total_mapped_count  3499415 bwa
##  3 SECH1    total_read_count   38516580 bwa
```

```
##  4 SECH1     total_mapped_count 19340203 bwa
##  5 SECH3     total_read_count   37356234 bwa
##  6 SECH3     total_mapped_count 25711919 bwa
##  7 SRR303333 total_read_count   10625978 bwa
##  8 SRR303333 total_mapped_count  6563682 bwa
##  9 SRR5860570 total_read_count    1397152 bwa
## 10 SRR5860570 total_mapped_count  1016591 bwa

## # A tibble: 5 x 4
##   sample     measure                 value aligner
##   <fct>      <chr>                   <dbl> <fct>
## 1 10A        filtered_mapped_count  1865642 bwaUniq
## 2 SECH1      filtered_mapped_count 11038965 bwaUniq
## 3 SECH3      filtered_mapped_count 18649404 bwaUniq
## 4 SRR303333  filtered_mapped_count  4229353 bwaUniq
## 5 SRR5860570 filtered_mapped_count   685616 bwaUniq
```



```
## # A tibble: 3 x 5
##   measure               minimum  average   median  maximum
##   <chr>                   <dbl>    <dbl>    <dbl>    <dbl>
## 1 filtered_mapped_count  685616  7293796  4229353 18649404
## 2 total_mapped_count    1016591 11226362  6563682 25711919
## 3 total_read_count      1397152 18727955. 10625978 38516580
```

Read Counts by Processing Step: Raw, Mapped, Filtered
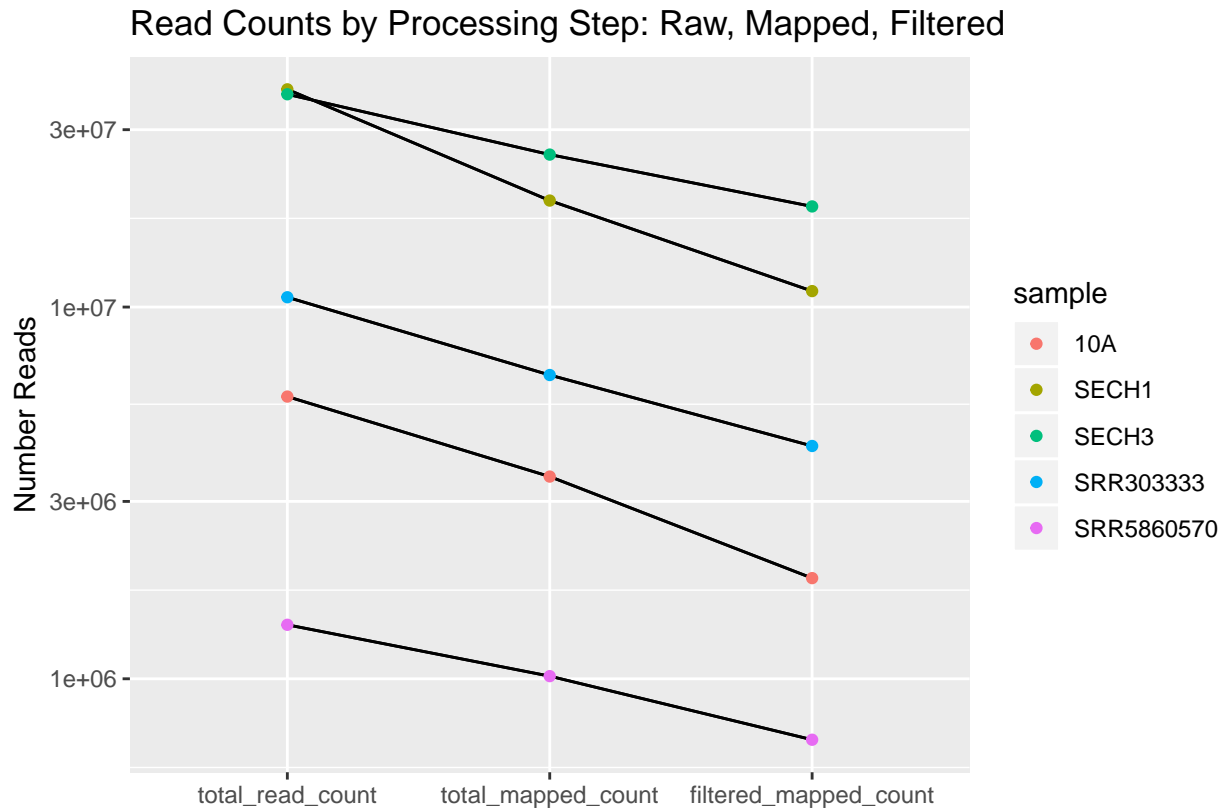
20 Nov 2018

Depth of coverage:

Table 4: Average Depth of Coverage for Raw and Filtered Alignments

step	minimum	average	median	maximum
pre-filtration depth	0.0071379	0.0788253	0.0460866	0.1805350
post-filtration depth	0.0048140	0.0512130	0.0296962	0.1309459
depth retention	0.5331297	0.6296025	0.6443568	0.7253214

```
## Warning: Column `sample`/`name` joining factors with different levels,
## coercing to character vector

## # A tibble: 2 x 6
##   species            step                  minimum average median maximum
##   <fct>              <chr>                   <dbl>   <dbl>  <dbl>   <dbl>
## 1 drosophila sechellia pre-filtration depth 0.136    0.158  0.158   0.181
## 2 <NA>               pre-filtration depth 0.00714  0.0259 0.0246  0.0461

## Warning: Column `sample`/`name` joining factors with different levels,
## coercing to character vector
```
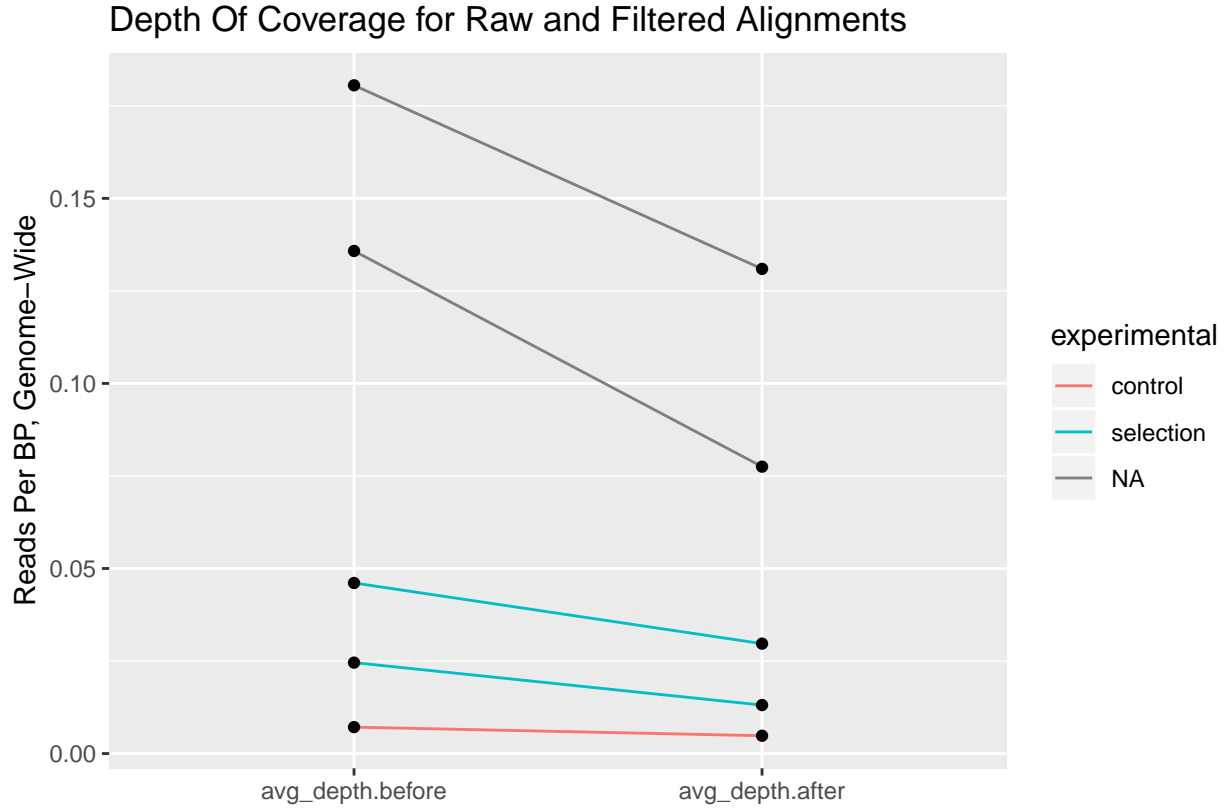
Depth Of Coverage for Raw and Filtered Alignments

Table 5: Breadth of Coverage Statistics for Raw and Filtered Alignments

step	minimum	average	median	maximum
pre-filtration breadth	24.1	61.6	64.4	81.9
post-filtration breadth	19.7	52.6	57.2	71.2
breadth retention	79.4	84.7	86.4	88.8

Breadth Of Coverage for Raw and Filtered Alignments