

# End Joining Signatures - dev

*Charlie Soeder*

*2/15/2019*

**25 February 2019**

Rebuilding, starting with summary stats for the materials/methods section.

Reference genomes

Table 1: Size and Consolidation of Reference Genomes

Reference Genome:	dm6
number_bases	144 M
number_contigs	1.87 k

Sequenced reads

Table 2: Number of Sequenced Samples by Treatment

experimental	sample_count
control	30

Table 3: Sequenced Experimental Samples

name	paired	experimental	source
Mcm5-A7	TRUE	control	dannyMiller
mcm5-28	TRUE	control	dannyMiller
mcm5-27	TRUE	control	dannyMiller
mcm5-26	TRUE	control	dannyMiller
mcm5-25	TRUE	control	dannyMiller
mcm5-24	TRUE	control	dannyMiller
mcm5-23	TRUE	control	dannyMiller
mcm5-22	TRUE	control	dannyMiller
mcm5-21	TRUE	control	dannyMiller
mcm5-20	TRUE	control	dannyMiller
mcm5-19	TRUE	control	dannyMiller
mcm5-18	TRUE	control	dannyMiller
mcm5-17	TRUE	control	dannyMiller
mcm5-16	TRUE	control	dannyMiller
mcm5-15	TRUE	control	dannyMiller
mcm5-14	TRUE	control	dannyMiller
mcm5-13	TRUE	control	dannyMiller
mcm5-12	TRUE	control	dannyMiller
mcm5-11	TRUE	control	dannyMiller
mcm5-10	TRUE	control	dannyMiller
mcm5-09	TRUE	control	dannyMiller
mcm5-08	TRUE	control	dannyMiller

name	paired	experimental	source
mcm5-07	TRUE	control	dannyMiller
mcm5-06	TRUE	control	dannyMiller
mcm5-05	TRUE	control	dannyMiller
mcm5-04	TRUE	control	dannyMiller
mcm5-03	TRUE	control	dannyMiller
mcm5-02	TRUE	control	dannyMiller
mcm5-01	TRUE	control	dannyMiller
DfMcm5	TRUE	control	dannyMiller

Total Starting Reads: 3.88G Post-QC Reads: 3.64G.

Table 4: Read Count and Percent Retention

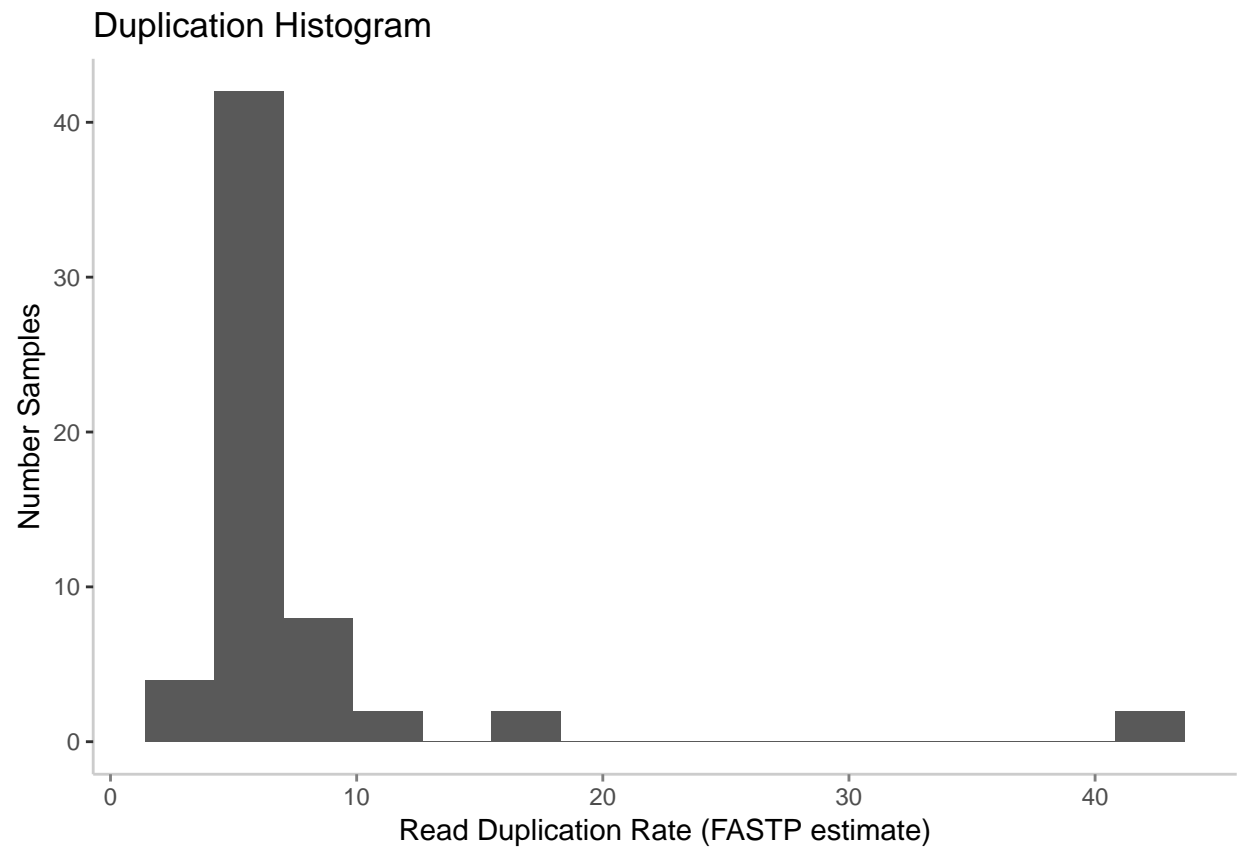
type	minimum	average	maximum
prefiltered	42.3 M	64.7 M	75.7 M
postfiltered	39.4 M	60.6 M	71.1 M
percent retention	92.4	93.7	95.5

This framework is general-purpose enough that it might be a good template.....

Dupes:

Table 5: Percentage Duplication

minimum	average	median	maximum
4.1	7.6	5.8	43.5



**27 February 2019**

Bioinformatics tips on INDEL calling & normalization with DSB background: [https://genome.sph.umich.edu/w/images/b/b4/Variant\\_Calling\\_and\\_Filtering\\_for\\_INDELS.pdf](https://genome.sph.umich.edu/w/images/b/b4/Variant_Calling_and_Filtering_for_INDELS.pdf)