2/26/2020
Kaitlyn Jacobs, Paolo Marra-Biggs, Nick Glaser

1. We chose the fungus Saccharomyces cerevisiae, brewer's yeast, because I was enjoying a cold one when picking. It's a highly occurring eukaryote not only in baking and brewing, but also in the functioning of the gut. We found a collection of mutations to align to the genome that was sequenced on an Illumina HiSeq X Ten.

2.

https://www.ncbi.nlm.nih.gov/sra/SRX6900124[accn] --> SRA Database

https://www.ncbi.nlm.nih.gov/genome/15?genome_assembly_id=22535 --> Genome

3. We chose to use BWA and Bowtie2 as our aligners, and the index/mem (BWA) parameter space. BWA gave us a 93% match rate while Bowtie matched at 94%. For BWA the modified parameters are able to change the minimum seed length (mem -k), band width (mem -w), penalties for gaps or mismatches (mem -B, -O), as well as many other facets of these parameters. Bowtie shares many similar parameters, both for building the index and the matching algorithm. For our best performance, we created a small index and ran the aligner with the --sensitive preset option. We found these preset options (4 total: very-fast, fast, sensitive, very-sensitive) a helpful addition to the aligner as it allowed for easy modification of the algorithm without having to tune each individual parameter.

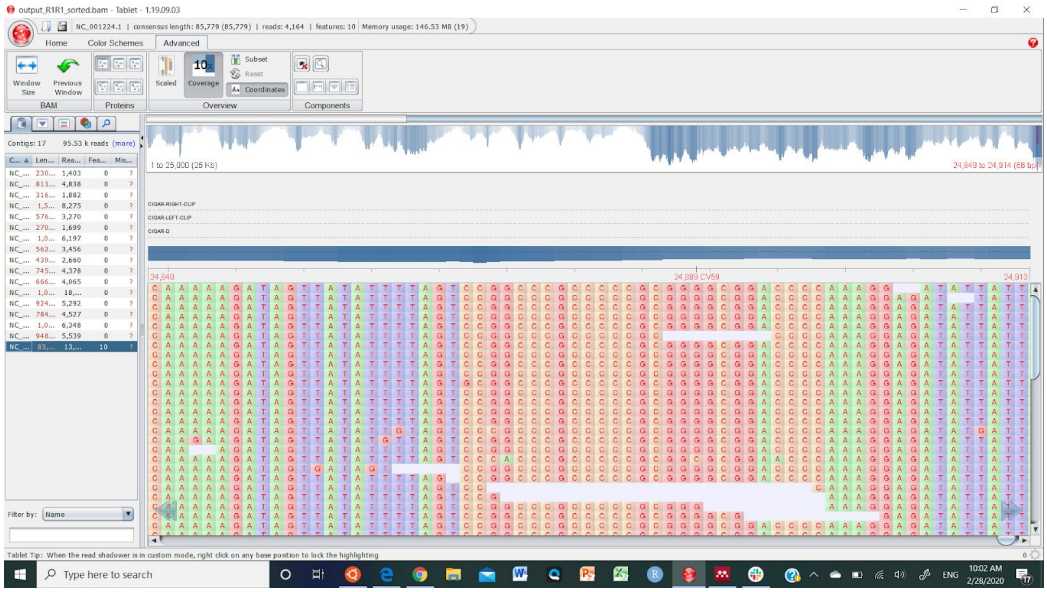
4. <See Github link>

Window number	Position in the genome	Number of reads starting in the window	%GC content
NC_001133.9-1	0-100	31	53
NC_001133.9-2	101-200	36	51
NC_001133.9-3	201-300	27	42
NC_001133.9-4	301-400	34	44

Full table link included on Github

6. Using Bowtie 2 results, R is -0.015106311727713484 A similar R-value achieved with BWA, of -0.015106311727713484.

7. From what we can see, there is no evidence of bias towards GC rich areas. Using the viewer Tablet, here is one instance of a GC rich area with high coverage (average depth of 71, 99.083% bases covered).



In contrast, below is a representation of a place in the genome with the same coverage statistics that's low in GC counts.