Report:

1. System configuration:

CPU: 2.9 GHz Intel Core i5

Clock rate: 2.9GHz

RAM: 8GB 1867 MHz DDR3

Cache size:

L2 Cache (per Core): 256 KB

L3 Cache: 3MB

2. Construction time:

chr12.fas: Segmentation Fault: 11

colorblind\_human\_gene\_fasta: 7 us

colorblind\_mouse\_gene\_fasta: 9 us

Human-BRCA2.cds.fasta: Segmentation Fault: 11

Slyco.fas: Segmentation Fault: 11

3. Justification:

The suffix tree construction run times were faster than I expected, considering the number of data points per file. I did not have tangible expectations but when the first colorblind file for humans ran in less than 13ms, I was surprised. After that file had run, it was no surprise that the mouse colorblind file and BRCA2 files would also run very quickly. The ~98 second runtime for Slyco then fell into expectations as well.

4. Implementation constant:

I do not yet know how to determine this.

5. BWT Index is included in the zip file.

When I run it for Human-BRCA2.cds.fasta, my results do not match the provided sample file.

I suspect that I have an error in the insertion functions.

Testing my code on the s1.fas and s2.fas files, output matches what is expected for DFS and BWT.

6. Exact matching repeat:

This can be determined by finding the lowest internal node and reading all of it’s children nodes to get the start indexes.

During DFS, maintain a pointer to the internal node with the longest string depth. Once DFS is complete, return to that node and read the start indexes of it’s children to find the locations in the string of the exact matches. The string depth will be the length of the match.