Mari Cobb Final Project Summary

MOTIVATION:
My motivation for this project stemmed from my thesis. I was working on a Drosophila actin-microtubule cross linker called Short Stop this year and came across various homologs to Short Stop in other organisms. I wanted to see how similar Drosophila Short Stop is to human ACF1, mouse MACF7, and mouse BPAG1. I put in two mouse proteins because I thought maybe it could serve as a control of sorts because I thought two proteins from the same animal might be inherently more similar in pair-wise comparisons than two proteins from two different animals. This was not the case as the two mouse proteins ended up being not similar at all (data not shown).

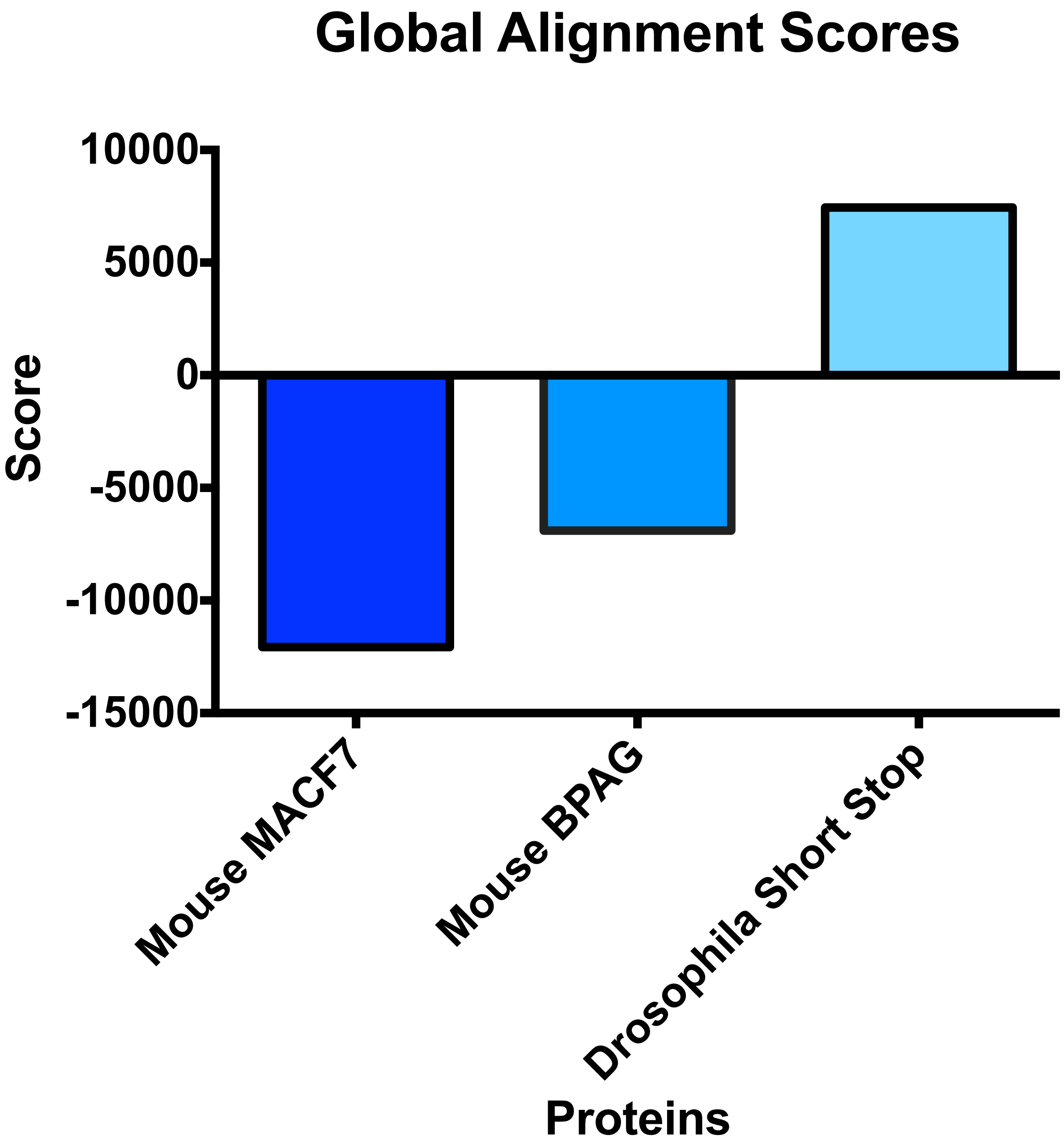
DATA FORMATTING:
I obtained FASTA files from the NCBI's Genbank and manually deleted out all of the new lines on Sublime2 so the entire sequence would be on one line. I initially tried to delete the new lines using python using the string.strip('\n') command, but that didn't work, so I just did it manually.

STEPS:
-Downloaded FASTA files.
-Deleted new lines from FASTA files so sequences would be on 1 line.
-Opened FASTA files in Python and printed the lengths of each sequence.
-Computed the global alignment score of human ACF1 vs another cross linker by using Anna Ritz's solution to Homework 7.2. I tried to use my own code that I had submitted for that assignment first, but failed to correctly integrate backtracking in it, so I just used the code in the solutions. The global alignment function took 4 arguments: sequence a, sequence b, the indel penalty, and whether or not the sequences were of DNA (instead of amino acids). I chose an indel penalty value of 1, so every time there is an insertion or deletion in the alignment, 1 is added to the alignment score.
-Computed the local alignment score of human ACF1 vs another cross linker. I used the code that I submitted for Lab 7. It took the same four arguments as my global align function and the indel penalty value stayed as 1.
-Computed the alignment. This function took three arguments, the backtrack table of the values, sequence a, and sequence b.
-Copied and pasted the alignment to word and found all of the hyphens in the alignment. I then subtracted that number from the total number of characters and divided that difference by the total number of characters. This allowed me to get the percentage of matches in the alignment.
-Compiled all my data together to draw a basic phylogenetic tree. This was not too difficult as all of my data points (local alignment score, global alignment score, and percentage) led to the same result.

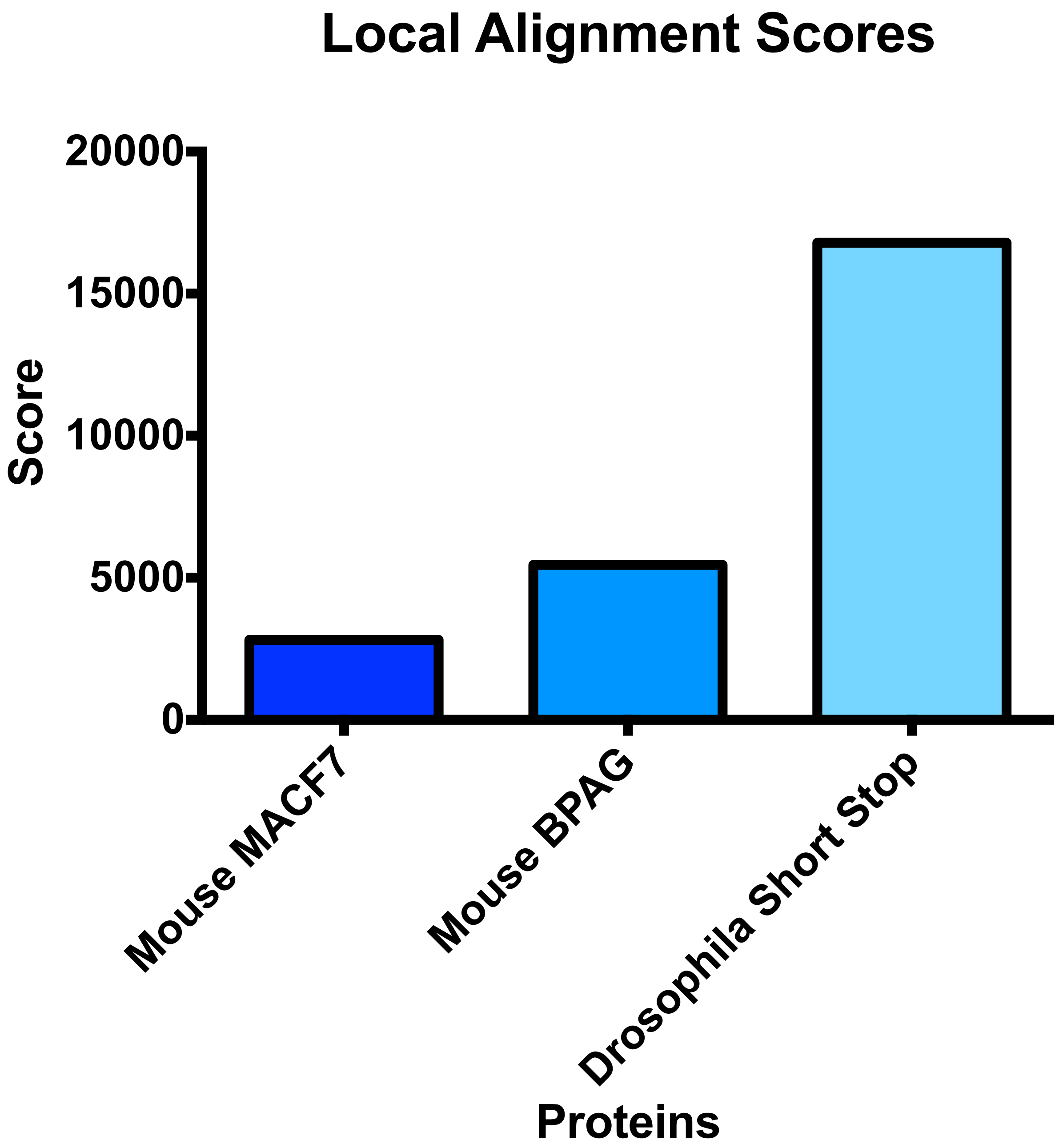
RESULTS AND CONCLUSION:
-I found that Drosophila Short Stop and human ACF7 are the most similar as far as the pairwise local and global alignments show. This pair also had the highest percentage of matches. This could likely be due to the fact that these two proteins were the most similar in length while the two mouse proteins were comparatively really short.

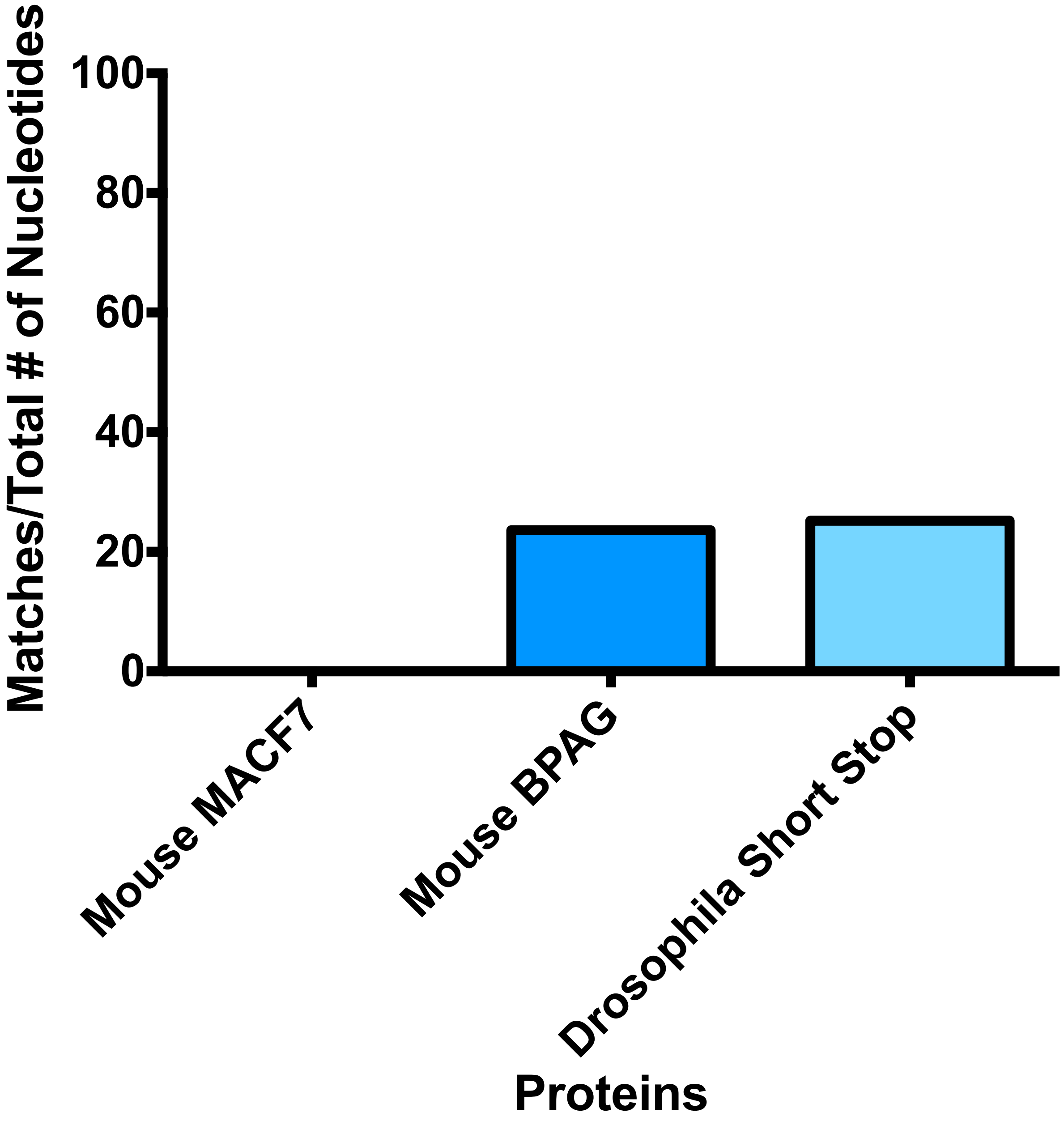
-My phylogenetic tree obtained from this data shows that in order of similarity, it goes mouse MACF7, Mouse BPAG1, Drosophila Short Stop, and Human ACF1. This is a confusing result because I expected to see that Drosophila and Human cross linkers were the least similar out of the four proteins I was comparing between.
-I think in order to apply this to broader situations, I would have to fix my local alignment function so that it would not just compute the length of the shorter of the two proteins as its local alignment score. These functions all took a really long time as well. It took about an hour for the human v drosophila alignment/scores to compute. It would be good to optimize it to run faster in the future.

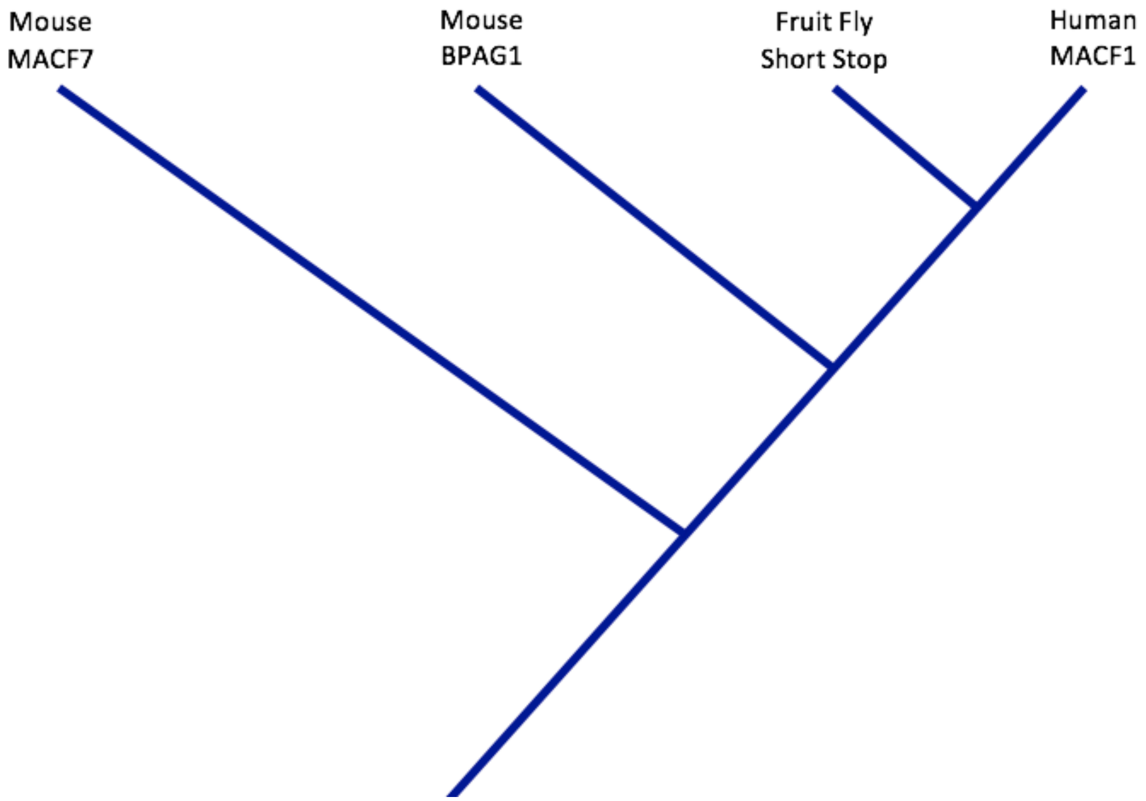
FIGURES:



Global Alignment Scores
10000
5000
0
-5000
-10000
-15000
Score
Mouse MACF7
Mouse BPAG
Drosophila Short Stop
Proteins


Local Alignment Scores
Score
20000
15000
10000
5000
0
Mouse MACF7
Mouse BPAG
Drosophila Short Stop
Proteins


Matches/Total # of Nucleotides
100
80
60
40
20
0
Mouse MACF7
Mouse BPAG
Drosophila Short Stop
Proteins


Mouse MACF7
Mouse BPAG1
Fruit Fly Short Stop
Human MACF1