

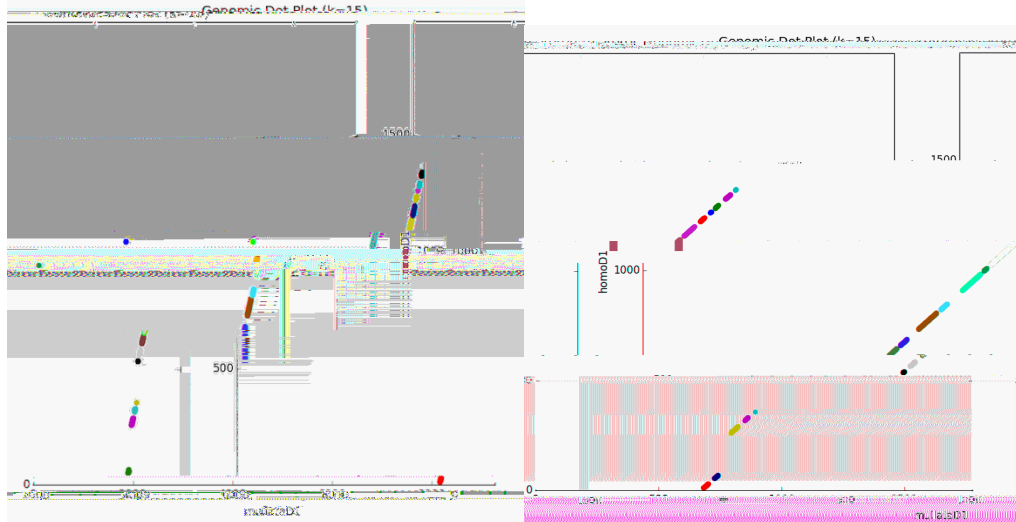
Independent Project Write up

Madeline Robin

- (a) Describe any data formatting you did (where/how you got the data, what manual trimming you needed to do, etc.)

For this project as test data I used the toys and strings that Anna made for Homework 8. The data for mouse, human, monkey and zebra fish genes were all found on NCBI and are all mRNA. The data for the rhesus monkey was predicted mRNA, and was manually trimmed to better align with the human sequence. The rhesus sequence was originally much longer than the human one so it was trimmed to places where there was alignment, which mostly entailed removing the beginning part of the rhesus sequences.

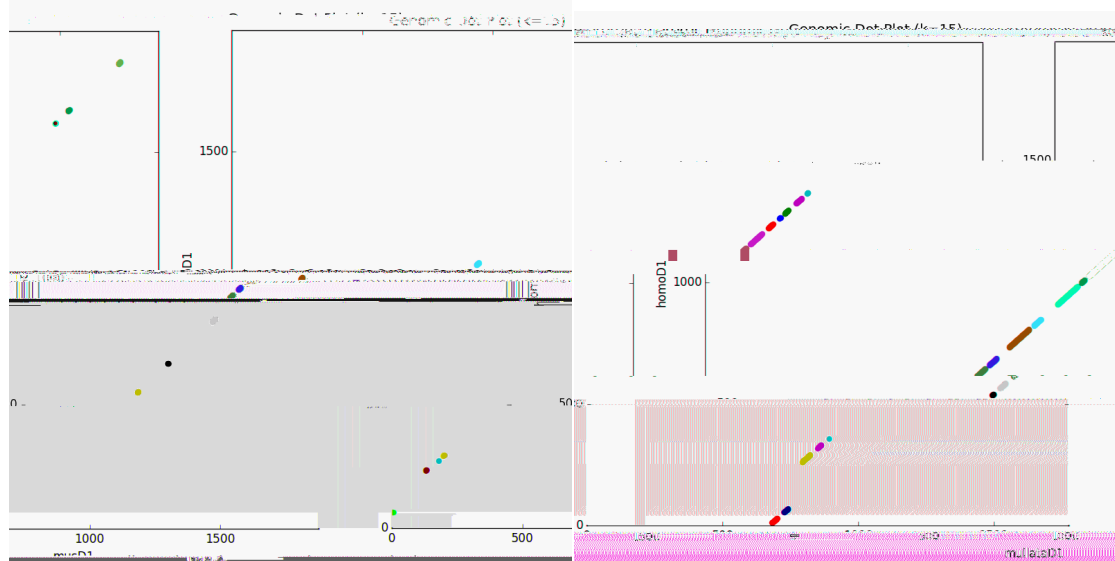
Figure1. Pre-Trimming homo vs rhesus Post-Trimming homo vs. rhesus



- (b) Describe, at a high level, the steps of your programs, similar to Question 4 in the Project Update.

The high level steps in this project involved writing a code that was able to recognize chunks of alignments and identify them as synteny blocks for both the forward and reverse strands of a sequence of nucleotides.

Figure 2. Side by side comparison of Humans vs Mouse and Humans vs Rhesus monkeys



- (c) Describe your results and some discussion/interpretation/conclusions. See details in the first section of the instructions.

From the sequences I chose to compare, you can see that HOX genes are in fact highly conserved across groups. It was expected that a monkey and human genes would have more alignment than that human and a mouse gene and that is seen in comparing the hoxD1 genome

It is important to be able to compare genomes and find conserved sequences across the species, because highly conserved sequences tend to be special sequences. Hox genes play a vital roll in development for all animals. Better understanding of these regions of the genome could shed light on important how evolutionary steps affect development. Although I used this code to look at highly conserved genes, it could be edited and implemented to find look at sequence alignments on the genomic scale.