Caroline Cutter
CSCI 321 Bioinformatics
May 16th, 2022

Final Project Diary

Hour 1 *with group

- **Goal:** pick a project, write a proposal for the project
- **Accomplished:** Decided to use challenge problem #3 in the book to reconstruct the Carsonella Ruddii genome using paired De Bruijn graph, wrote our proposal to submit to Professor Linderman

Hour 2

- **Goal:** researching Paired De Bruijn graphs to familiarize myself with the algorithms and changes necessary
- **Accomplished:** read papers and watched some online videos to understand paired De Bruijn graphs a little bit better

Hour 3 *with group

- **Goal:** review and talk about how to generate read paired De Bruijn graph
- **Accomplished:** reviewed our De Bruijn graph Rosalind algorithms and talked about pseudocode for the necessary function and changes to these (paired composition, paired graph, eulerian cycle/path, reconstruction…)

Hour 4 *with group

- **Goal:** Start modifying previous code from Rosalind assignments to make the new De Bruijn graph algorithm
- **Accomplished:** started looking at the Eulerian Path (didn't get as far into the code as we would like because had trouble coding with 3 people), created a code sharing platform

Hour 5 *with group

- **Goal:** start working on generating read pairs from a full data sequence
- **Accomplished:** Was able to re-use the k-mer composition of a string in Rosalind but spent most of the time re-coding it to work with our files

Hour 6 *with group

- **Goal:** continue making the paired DeBruijn graph algorithm

- **Accomplished:** was able to successfully generate the paired De Bruijn graph using a smaller example, but it only worked sometimes

Hour 7

- **Goal:** try to figure out why the code worked with some data and not with others
- **Accomplished:** didn't make a lot of progress, did find a few bugs that helped improve the accuracy but didn't quite work yet

Hour 8 *with group

- **Goal:** create the Eulerian Cycle for the paired De Bruijn graph
- **Accomplished:** added edge from the end node to start node, found trouble with the coding of this because the program wouldn't accept the tuples versus the lists

Hour 9 *with group

- **Goal:** Start working on the poster by working on the methodology section
- **Accomplished:** wrote about our methodology so far to explain what changes we made to the De Bruijn graph algorithm, talked as a group about how we wanted to analyze the data

Hour 10

- **Goal:** Come up with hypotheses and conclusions of why our data is what it is
- **Accomplished:** did research on possible reasons why the data looks the way it does, wrote out some questions that people might ask during presentations and came up with some answers

Hour 11 *with group

- **Goal:** start to finalize our poster by writing all our findings and conclusions on them
- **Accomplished:** Made conclusions based on the data and findings we had, started to finalize the poster

Hour 12 *with group

- **Goal:** Meet with Professor Linderman to ask how we could make our poster and project better
- **Accomplished:** Met with professor Linderman, talked about how we could better analyze the k-mer length and come up with conclusions, started talking as a group of how we could do this

Hour 13

- **Goal:** create a program to analyze the different repeated k-mers in a sequence
- **Accomplished:** pulled together the genome, started writing the code to do this, researched how to write to a csv file

Hour 14

- **Goal:** Keep working on analyzing the different repeated k-mers in a sequence
- **Accomplished:** Finished the program to do this, wrote to CSV file, came up with some thoughts and hypotheses of why the data looked like this

Hour 15

- **Goal:** Use R to create graphs of the repeated k-mer data
- **Accomplished:** analyzed the data using R, made the graphs, talked as a group about how we could incorporate these findings but decided to not include the graphs, just talk about the data in the conclusions (we still put these in our code folder)

Hour 16 * with group

- **Goal:** Finalize the project by compiling all of our work
- **Accomplished:** finalized the poster, created a readMe for all the files, finished progress diaries, checked the project rubric to make sure we didn't forget anything, and submitted our final project :)