

Julia Fairbank
Progress Diary
CS321 Bioinformatics Final Project

Hour 1 *with group

Set out to: Review how to generate read pairs and paired de Bruijn graphs
Accomplished: Reviewed how to generate read pairs and paired de Bruijn graph in the book chapters. Generated pseudocode for the necessary functions (paired composition, paired dbgraph, eulerian cycle/path, reconstruction...)

Hour 2 *with group

Set out to: Compile relevant code from previous Rosalind assignments
Accomplished: Pulled from Eulerian Path Rosalind, established code-sharing platform, researched more on read-pairs

Hour 3 *with group

Set out to: Generate Read Pairs from full data sequence
Accomplished: Pulled from Generate the k-mer Composition of a String Rosalind, established code-sharing platform, researched more on read-pairs

Hour 4 *with group

Set out to: Make paired DeBruijn graph
Accomplished: Successfully generated a paired DeBruijn graph using text example ("TAATGCCATGGGATGTT", $k=3, d=1$) worked for this scenario, but not others

Hour 5

Set out to: Create shell for poster
Accomplished: Looked at examples of previous posters as well as other scientific posters to create our own, then chose relevant sections headers

Hour 6

Set out to: Do background research on genome
Accomplished: Wrote background section explaining why we were researching this genome in particular and what makes it interesting (its size)

Hour 7

Set out to: Get feedback from Prof. Linderman about current status of poster
Accomplished: brainstormed different ways to show results of project and got feedback to trim down text in the background portion of poster

Hour 8 *with group

Set out to: Create Eulerian Cycle
Accomplished: added edge from the end node to the start node, found the eulerian cycle *again difficulty dealing with tuples vs. lists

Hour 9 *with group

Set out to: Create Eulerian path and reconstruct genome

Accomplished: Split the eulerian cycle between start and end node to create an eulerian path, generating two reconstructed portions of the string (one using first kmer and one using second).

Merged the two together (disregarding overlap (only seen once))

Hour 10 *with group & Prof. Linderman

Set out to: Meet with Prof. Linderman and figure out the next step for furthering our project

Accomplished: brainstormed ideas of data analysis: sweep k and sweep d and compare N50 or randomly generate portions of the genome and compare N50

Hour 11

Set out to: run our program once and gather data

Accomplished: tried to run our data once, but splitting our entire genome into reads of 100 was unreasonable. Randomly generated 10 reads of length 1000 to analyze. Wrote to txt file for our data as k and d swept

Hour 12

Set out to: Write Algorithm/Methodology & Reconstruction section

Accomplished: created pseudocode for our current algorithm

Hour 14

Set out to: Put together reference section

Accomplished: compiled resources used and construct casual bibliography section

Hour 15

Set out to: Input results & graphs into poster

Accomplished: Aligned graphs on poster and created sections for analysis of graphs

Hour 16 *with group

Set out to: Adapt methodology section

Accomplished: wrote a concise methodology that explained how we modified our class programs to fit our new goal. Wrote figure legends that explained what conclusions could be made from each graph.

Hour 17 *with group

Set out to: Make conclusions based on findings and what we were showing on our poster

Accomplished: Created 4 bulletpoints of takeaways

Hour 18/19/20 *with group

Set out to: compare poster with expectations and requirements, making sure everything was present, correct, and visually appealing

Accomplished: Realigned, recolored and rearranged different sections to make sure poster contained all required information and was easy to track and understand.