

LAB 9 PROJECT PLAN HELP (IF NEEDED)

NOTES

- A. Be sure to test as you go along! Just try printing what you have to the screen or something to make sure that your section actually has the data you are expecting. Do not wait to try to run it until all of the pieces are there!
 - B. I'd strongly suggest writing this program for measles first, since you have all of my files and example output. Then when that works, just change the filenames to be the covid ones.
-

Here are tasks that need to be done that have dependencies, and tasks that can be done concurrently.

CONCURRENT 1

The amino acid ArrayList has to be built first from the aminoAcidTable.csv file. No one can get any real work done without that.

- A. You'll need the AminoAcid class built. That class needs to contain the name of the acid, the one letter code it goes by, and a list of the codons that make it up. Store those few codons however you want but remember that different amino acids have different numbers of codons. Therefore, you can't just have attributes codon1, codon2, etc. in your class.
 - B. In Main, write a method to read in the aminoAcidTable.csv file and split it. Every line in the file is one amino acid, so make an instance of an AminoAcid class, fill it with the data, and then add it to your ArrayList of amino acids.
 - C. Break up the covid19Sequence.csv file into its codons. You'll need the ability to loop through the whole genome later.
-

With those necessities out of the way (items everyone will need), you can start working on the bigger functionality.

CONCURRENT 2

- A. Someone begin the "print all amino acids" report.
 - Open a codon bias output file to write to, loop through every amino acid in your amino acid ArrayList, and print the name, letter, and its codons to the file.
 - ***(Note that at the moment, this will not display the codon percentages for those written codons. That's fine. We'll come back to that later.) ***
 - B. Someone write the code to count all of the codons for a specific amino acid. Choose an amino acid (just pick one) and grab all of its codons from the ArrayList from 1.B. Then, loop through the entire codon array from 1.C. Keep track of all of the counts for each codon in that amino acid. (***This will take some thought***)
 - For example, if the amino acid is 'N', then the codons for it are AAT and AAC. Start at the beginning of the codon array. Every time you find an AAT, add one to your count for that, and every time you find an AAC, add one to your count for that.
-

LAST PARTS

3. When 2.B is finished, you have the counts for each codon in the genome. Figure out the percentage of each codon that makes the amino acid you're working on.
 - I'll keep using the AAT and AAC example for amino acid 'N'. Figure out the percentage of AATs vs AACs.
 - i.e. if there are 100 AATs and 150 AACs in the arraylist from 1.C, then there are 250 total that code for N (some AAT and some AAC). The percentage of AATs is 40% and AACs are 60%. Write those percentages to your file.
 4. Determine the percentage for an amino acid used in the entire genome. Write that information to the file.
 - For example, if there are 250 codons that code for N, and there are 5000 codons in the whole genome, then N would account for 5% of the genome.
 5. Now that 3 and 4 are working, you can put all of that in a loop where you're checking every amino acid in your ArrayList from 1.B instead of just one specific one.
-

I hope this helps! It's not every detail of course, but it might help you break out the work to be done.