

COSC 1437, Professor Clark

Assignment 9 Part A: 12 points and Part B: 6 points

Write a Program: COVID-19 Codon Bias Analysis

Your goal is to do analysis on the COVID-19 virus (*technical name: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)*).

To understand the terminology within these requirements, read the Lab9Definitions.pdf file. It is one of the files in the Lab9NeededFiles.zip in the Assignments section of the Chapter 9 Module.

Program Requirements:

You will write a program that will do codon bias analysis for the COVID-19 genome. The DNA sequence is located in the covid19Sequence.txt file and was collected in Texas in January 2024.

You will write a report to a file containing the information below for each amino acid in the genome:

- The full name of the amino acid along with its one letter code
- All codons that can code for that amino acid
- The percentage for which each codon is found for that amino acid within the genome
- The overall percentage for which that specific amino acid is found within the genome

Design Requirements and Suggestions:

You are provided with the following files to begin (located in Lab9NeededFiles.zip):

- aminoAcidTable.csv – a file of all amino acids with their names, abbreviations, and all codons that can create that amino acid
- covid19Sequence.csv – a file for the COVID-19 genome, with all codons already delimited
- Lab9Definitions.pdf – a file containing definitions for the terms used in this assignment

You're going to need 2 classes but may choose to create as many as you wish. Minimally, you'll need one for Amino Acid and a Main class. If you go this design route, your Main class will be large. This is fine, as long as there are many well documented methods in it.

Design this before you start trying to code it! With multiple team members, communicating who is completing which parts is crucial, and determining how those parts will interact with each other ahead of time will make the merging of your code pieces much smoother. I've

provided a (hopefully) helpful file called ProjectDesign.txt to help with your team project management for the components.

You should consider holding the following information for your Amino Acid class design:

- Full name
- Single letter abbreviation
- All codons capable of creating this amino acid

You'll be handling a lot and will likely have several arrays/ArrayLists/variables/methods/etc. being managed in your Main class (or additional classes you choose to create). Be sure to document everything well (Javadoc comments, comments within code as needed).

Let me be clear, this project will not be particularly easy. It will take your group working together to figure out how you want to tackle this.

Submission Requirements:

Upload a .zip file for your group on Canvas containing the following items:

- A folder containing all your code for the project
- Codon bias report

Grading Rubric for Part A: (Note, all group members receive the same score for this portion)

1. Code Correctness (meets specification)	5 points
2. Design (well-constructed solution) – this is important!	5 points
3. Report Correctness and Formatting (Readability)	2 points

Total for Part A: **12 points**

Part B: Individual Project Effort submission:

Look on Canvas for Lab 9, Part B. It has three questions asking about your group dynamics. Part B scored individually and is worth **6 points**. *Do not complete Part B until Part A is submitted.*

Total for Part B: **6 points**

Example Output:

I will not provide example output for the COVID-19 genome you are analyzing. However, I have provided a measles.zip file on Canvas (in our Chapter 9 module) with the requested generated report on a Measles genome. You may look at that report to see the format etc. of what you need to provide.

A note: you may wish to write your project first against the Measles genome (I have provided that file within measles.zip) If you can get that to work and get my output, then it will only require very minor updates to your code to get it to work for COVID-19.