Lin Lin Mei

**Documentation – Project #14 – Finding Prophages**

-One paragraph description of the project.

This project is about finding prophages within a genome. Given two files, one have the genome sequence (with 1 line that has information about the sequence), and the other have the start and stop positions of the sub-sequences and descriptions of the genes. The purpose of the project is to find all prophage sequences using the file (that have the positions and gene descriptions), then extract those sub-sequences (out of the genome sequence) and write them to a file.

-One paragraph description of the data structures used and ***why*** you selected this data structure(s) for the implementation.

The data structure I used is a vector. Considering the project, I will have to read in files and use their data, and I usually used vectors to do that, so I thought vectors would be simple and useful for this project too. Out of all the other data structures (arrays, linked-lists, stacks, queues, etc.), I understand vectors more than the others. Vectors are like dynamic arrays, so I don’t have to worry about the size, and it is flexible enough to take in any size that can store data from files. Though it takes in a lot of memory for storage, I think I’m able to access the elements in a vector better than other data structures like linked-lists.

-A step-by-step instruction of how to use the code. This should include a test case for running your code.

Once you download the code (prophage\_project.cpp) and the files (genetics file.txt and sequence.txt), put the files into the same project as the code (for Code::Blocks, I just created a project and added all the files and code into it). After that, you can hit the build and run button. All the files are inputted into the code, so the code should build and run smoothly. Once it finish running, you should see a “phage sequences.txt” file inside the project folder.

If you want to run smaller versions of the files, they’re included. Everything is in main (or the main function), there are no other functions, so if you want to run the smaller versions, just enter “seq short.txt” in genome.open() and “genes short.txt” in genes.open(). But, I doubt they will show any prophage sequences in the written file (“phage sequences.txt”).

-List the IDE(s) and OS(s) used during development and testing.

The IDE used during development and testing is Code::Blocks and the OS used was Windows 8.1