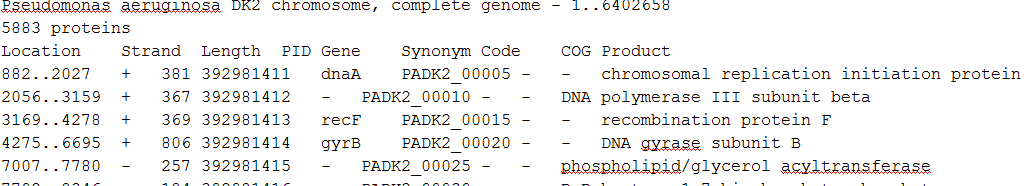
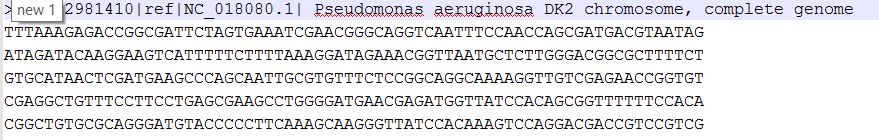
Final Project

The purpose of the project is to take in two files, a gene file (.fna) and a phage file (.ptt) and extract the sections specified in the phage file from the gene file. The gene file is comprised of A’s, T’s, G’s, and C’s which represent the four DNA nucleotides. The phage file is comprised of the names, products and location of certain genes within the gene file. The goal of the project is to extract the genes from the gene file and write the sequence to an .out file using the phage file as a guide. Below are snippets of what the .ptt and the .fna files look like.



The .ptt file sample



The .fna file sample

Vectors were used in this project to store the data from the phage file which included the start, the stop, the length (calculated), and other information not yet in use. The program indexed through the Vector by moving to the needed position and extracting required number of characters. A string data type was used in the formation of a pseudo data structure knowns as the “buffer.” The buffer contained a number of gene lines from the gene file. The buffer allowed the program to move back and forth in the buffer without rewinding the file this is particularly useful with a phage file which overlaps.

To allow for code reuse the classes were broken into their own .cpp and .hpp files with main in a third .cpp file. The project consists of the follow files:

* ReadGeneFile.hpp, ReadGeneFile.cpp
* ReadPhages.hpp, ReadPhages.cpp
* Main.cpp

The main prompts the user for a gene file (.fna) and a phage file (.ptt) both of which are also provided, although any .fna or .ptt file could be used. The main also prompts the user for the name of the .out file. Next all the user has to do is let the program run until the output is produced, which can take some time in comparison to other projects done for this course. The user can then look at the .out file at their convenience.

CodeBlocks 16.10 was used on Windows 10 on a Lenovo u410 laptop pc.