CAB201 Programming Principles - Semester 2, 2019 Report for Assignment: Project – Genomic Sequence Retrieval - Part I

Student name and number:

James Norman n10256989

Build and Run Instructions

Please provide <u>clear</u> step-by-step instructions here on how to build your program in Visual Studio and run your program in the command line, given your submitted zip folder. For each step, you should include a screenshot. You may expand the box if needed

[Enter build & run instructions. Include where to find .sln file (with screenshot), how to build your program (with screenshots), how to find the .exe file (with screenshots), and finally a screenshot of your .exe program being run from the command line]

Step 1. Unzip the CAB201_Assignment folder

Step 2. Click into the folder run the .sln program

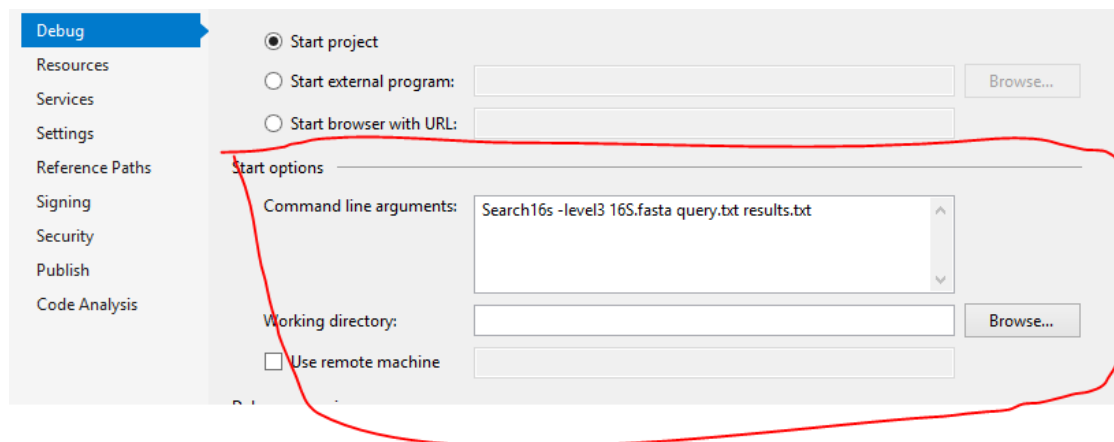|			
📁 CAB201_ASSIGNMENT	4/09/2019 1:51 PM	File folder	
📄 CAB201_ASSIGNMENT.sln	16/08/2019 4:48 PM	Visual Studio Solu...	2 KB

Step 3. (optional) The fasta file should already be in there but just in case it's not or you want to replace it with a specific file go to the folder above the .sln then into the folder labelled bin and then the folder labelled debug and then paste it in there.

Step 4. Enter what level and input arguments you would like to do this click on the properties window in the sln
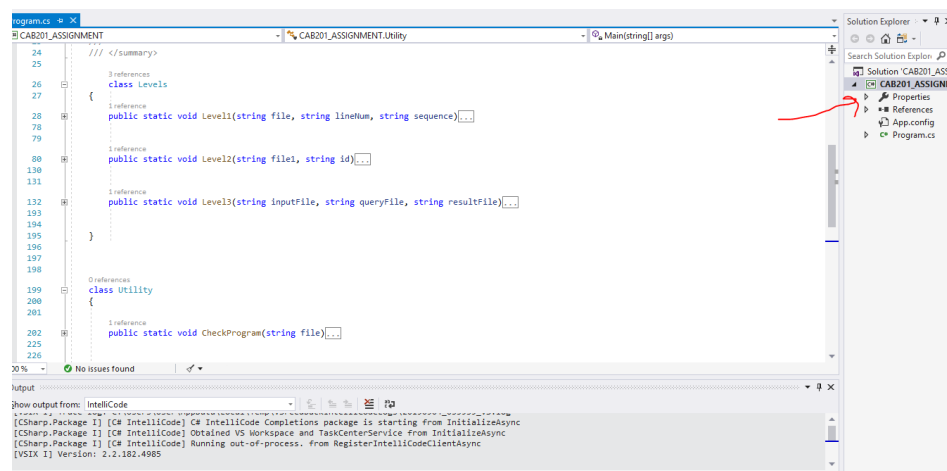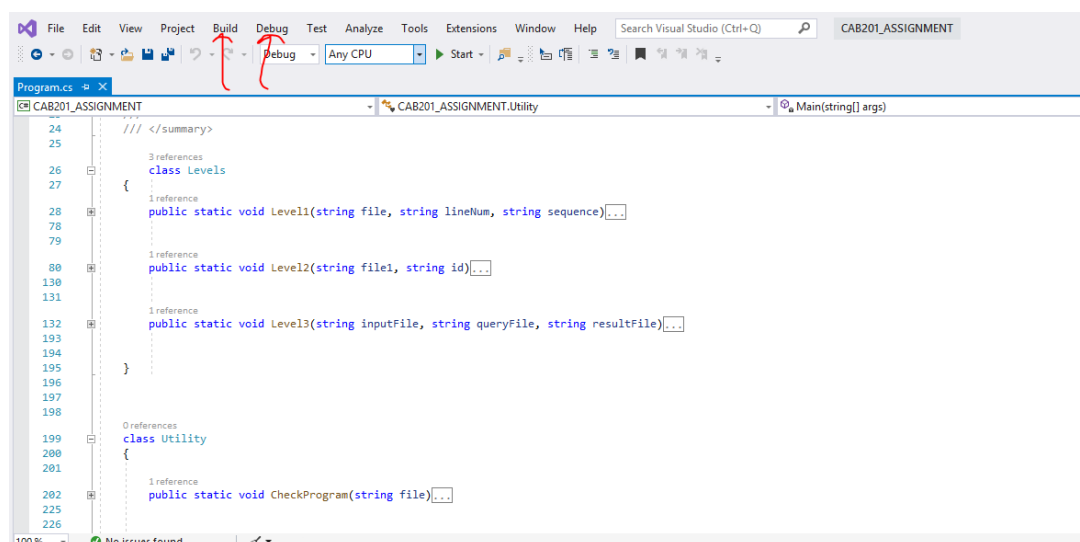

(.sln Filename is now Search16S)

Then click on the debug tab and enter your input arguments in the box for example

Debug

Resources

Services

Settings

Reference Paths

Signing

Security

Publish

Code Analysis

◉ Start project

○ Start external program: Browse...

○ Start browser with URL:

Start options

Command line arguments: Search16s -level3 16S.fasta query.txt results.txt

Working directory: Browse...
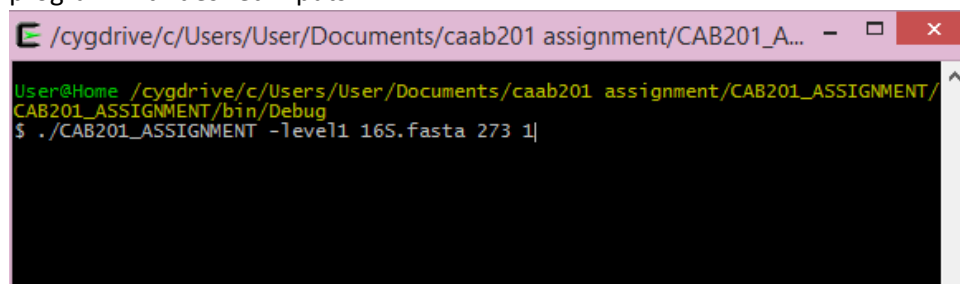
☐ Use remote machine

Step 5. Once you've entered your input arguments go back to the program.cs tab and click into build and then build solution, alternatively you could go into the debug tab and run it from there using either the run without debugging or with debugging.

File   Edit   View   Project   Build   Debug   Test   Analyze   Tools   Extensions   Window   Help   Search Visual Studio (Ctrl+Q)   CAB201_ASSIGNMENT

Debug   Any CPU   ▶ Start ▾

Program.cs + X

CAB201_ASSIGNMENT   CAB201_ASSIGNMENT.Utility   Main(string[] args)

```
24        /// </summary>
25
          3 references
26        class Levels
27        {
          1 reference
28        public static void Level1(string file, string lineNum, string sequence)...
78
79
          1 reference
80        public static void Level2(string file1, string id)...
130
131
          1 reference
132       public static void Level3(string inputFile, string queryFile, string resultFile)...
193
194
195       }
196
197
198
          0 references
199       class Utility
200       {
201
          1 reference
202       public static void CheckProgram(string file)...
225
226
```
100 %   No issues found

Step 6. A command prompt will then popup with your desired output.

Step 7. (optional) alternatively you can use the command line of a bash shell to run this program, to do that repeat steps 1-5 and click build solution, after that open up a shell program eg: Cygwin, command prompt, change directories to the debug folder and run the program with desired inputs.

/cygdrive/c/Users/User/Documents/caab201 assignment/CAB201_A...   ─   □   ✕

```
User@Home /cygdrive/c/Users/User/Documents/caab201 assignment/CAB201_ASSIGNMENT/
CAB201_ASSIGNMENT/bin/Debug
$ ./CAB201_ASSIGNMENT -level1 16S.fasta 273 1
```

Statement of Completeness

This statement of completeness will need to *accurately* state the functionality which has been implemented. There will be a penalty of 3.5 marks (loss of 3.5 marks) for a non-completed or submitted statement of completeness, and a penalty of 1 mark for each inaccurate statement to a maximum of 3 marks.

In the following section, you are required to mark which functionality you have implemented. <u>In the column on the right please mark 'Y' where you have completed this functionality, and 'N' where you have not</u>. Please fill in any additional text boxes requested, and please note any limitations or bugs in the box at the end of each section. You may expand the table if you need more room for comments.

Basic Functionality		
Basic itinerary output	The program displays the data from the file	Y
	The program displays the appropriate line	Y
	The correct amount of information is displayed, e.g. only the relevant entries	Y
	The correct level, provided as a command line flag - *levelN*, is executed	Y
	The program **does not** store the whole file in memory, instead it accesses the file on disk	Y
Error handling	A clear error message is displayed when an incorrect number of arguments is provided	Y
	A clear error message is provided when an incorrect flag is provided (e.g. not -level1, etc.)	Y
	A clear error message is provided when the input file doesn't exist, or is incorrectly formatted	Y
Comments	*I tried my hardest but could not get the .exe to be Search16s so it'll be CAB201_ASSIGNMENT.exe* 🙁	

Searching Algorithm – Part I
Please <u>underline, circle or highlight</u> the levels that were completed.

| Algorithm	Level: Level 1, level 2, level 3
Comments	All 3 levels have been completed correctly with no bugs found in the testing done.

Screenshots of Functionality

In the following section, you are required to provide screenshots that provide evidence of your program working with provided input. You must complete this section.

1) The `16S.fasta file` has been provided with this template. Download them and place them in the same folder as your .exe file. You may have extra files, e.g. a query file, in this folder, and your .exe may be named differently. This is fine.

2) Open the command prompt and go to the above folder. In the command line, type the name of the .exe file and copy and paste following arguments:

   `Search16s -level1 16s.fasta 273 10`

Place screenshot of entered command line arguments here:
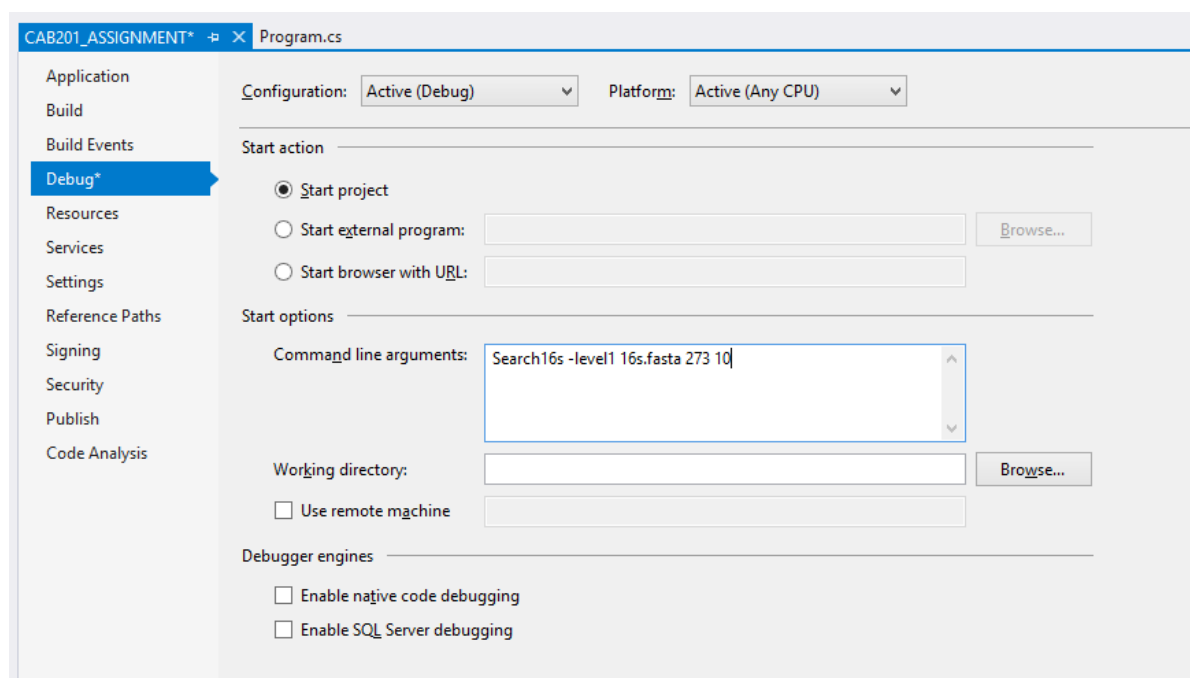
3) Hit enter to run your program.

Place screenshot(s) of the full output to console of your program. You may expand the box as necessary, and use as many screenshots as needed:

```
>NR_025900.1 Thermus aquaticus strain YT-1 16S ribosomal RNA, partial sequence
GCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCAAGTCGTGCGGGCCGTGGGGTATCTCACGGTCAGCGGCGGACG
GGTGAGTAACGCGTGGGTGACCTACCCGGAAGAGGGGGACAAACATGGGGGAAACCCAGGCTAATCCCCCATGTGGACACAT
ACCAAGGCGACGACGGGTAGCCGGTCTGAGAGGACGGCCGGCCACAGGGGCACTGAGACACGGGCCCCACTCCTACGGGA
GGCAGCAGTTAGGAATCTTCCGCAATGGGCGCAAGCCTGACGGAGCGACGCCGCGCTTGGAGGAGGAAGCCCTTCGGGGTGT
AAACTCCTGAACCCGGGACGAAACCCCCGATGAGGGGACTGACGGTACCGGGGTAATAGCGCCGGCCAACTCCGTGCCAG
CAGCCGCGGTAATACGGAGGGCGCGAGCGTTACCCGGATTTACTGGGCGTAAAGGGCGTGTAGGCGGCTTGGGGCGTCCC
ATGTGAAAGGCCACGGCTCAACCGTGGAGGAGCGTGGGATACGCTCAGGCTAGACGGTGGGAGAGGGTGGTGGAATTCCC
GGAGTAGCGGTGAAATGCGCAGATACCGGGAGGAACGCCCGATGGCGAAGGCAGCCACCTGGTCCACTCGTGACGCTGAGG
CGCGAAAGCGTGGGGAGCAAACCGGATTAGATACCCGGGTAGTCCACGCCCTAAACGATGCGCGCTAGGTCTCTGGGTTA
TCTGGGGGCCGAAGCTAACGCGTTAAGCGCGCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGG
GGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATGCTAGGGAAC
CTGGGTGAAAGCCTGGGGTGCCCCGCGAGGGGACCCTAGCACAGGTGCTGCATGGCCGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCGTTAGTTGCCAGCGGGTGAAGCCGGGCACTCTAACGGGACTGC
CTGCGAAAGCAGGAGGAAGGCGGGGACGACGTCTGGTCATCATGGCCCTTACGGCCTGGGCGACACACGTGCTACAATGC
CCACTACAGAGCGGAGGCGACCTGGCAACAGGGAGCGAATCGCAAAAAGGTGGGCGTAGTTCGGATTGGGGTCTGCAACCC
GACCCCATGAAGCCGGAATCGCTAGTAATCGCGGATCAGCCATGCCGCGGTGAATACGTTCCCGGGGCCTTGTACACACCG
CCCGTCACGCCATGGGAGCGGGTTCTACCCGAAGTCGCCGGGAGCCTTAGGGCAGGCGCCGAGGGTAGGGCCCGTGACTG
GGGCGAAGTCGTAACAAGGTAGCTGTACCG)
>NR_118761.1 Actinobacillus ureae strain Henrikson 3520/59 16S ribosomal RNA, pa
rtial sequence
ATTGAAGAGTTTGATCNNGGCTCAGATTGAACGCTGGCGGCAGGCTTAACNNNTGCNAGTCGAACGGTAACAGGGATTAG
CTTGCNNATTTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGGAATCTGGCTCATGGAGGGGGATAACTACGGGAAA
CTGTAGCTAATACCGCGTAATATCTTAGGATTAAAGTGTGGGACCGCAAGGCCACATGCCATGAGATGAGCCCAAGTGGG
ATTAGGTAGTTGGTGGGGTAAAAGGCCTACCAAGCCGACGATCTCTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAAC
TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCNCNATGGGGGGGCAACCCTGATGCAGCCATGCCG
CGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTAGCGAGGAAGGTATCAAATTTAATAGATTTGGTAATTGA
CGTTAACTACAGAAGAAGCACCNGCNNACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATA
ACTGGGCGTAAAGGGCACGCAGGCGGTTGATTAAGTGAGATGTGAAAGCCCCNGGCTTAACCTGGGAATTGCATTTCATA
CTGGTCAACTAGAGTACTTTAGGGAGGGNTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGA
AGGCGAAGGCNGCCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTA
GTCCACGCTGTAAACGCTGTCGATTTGGGGATTGGNCTTTGAGCTTGGTGCCCGTAGCTAACGTGATAAATCGACCGCCT
GGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAT
GNAACGCGAAGAACCTTACCTACTCTTGACATCCATGGAATCTTGTAGAGATACGAGAGTGCCTTCGGGAACCATGAGAC
AGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGT
TGCCAGCGATTTGGTCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCAT
GGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACNGAGGGAAGCAATATGGCGACATGGAGCAAATCTCA
CAAAGTACGTCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCAAATCAGAATG
TTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATAGCTT
AACCGCAAGGGGGGGCGTTTACCACGGTATGATTCATGA)
>NR_118745.1 Actinobacillus seminis strain ATCC 15768 16S ribosomal RNA, partial
 sequence
AATTGAAGAGTTTGATCANGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAACGGTAACGGGGGAGGG
AGCTTGCTTTCTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTNGGGAATCTGGCTTATGGAGGGGGATAACGACGGG
AAACTGTCGCTAATACCGCGTAGTATCGGGGAGATGAAAGGGTGGGACTGAAGGGCCACCTGCCATAAGATGAGCCCAAGT
GGGATTAGGTAGTTGGTGGGGTAAAGGCCTACCAAGCCTGCGATCTCTAGCTGGTCTGAGAGGATGGCCAGCCACACTGG
AACTGAGACACGGTCCAGACTNNTACGGGAGGCAGCAGTGGGGAATATTGCGCNATGGGGGGAACCCTGACGCAGCCATG
CCGCGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTNGTGAGGAAGGTGATGAAGTGAAGAGTTTCATCAAT
TGACGTTAATCACAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACNGGGGGTGCGAGCGTTAATCGGA
ATAACTGGGCGTAAAGGGCACGCAGGCGGACTTTTAAGTGAGGTGTGAAATCCCCGGGCTTAACCTGGGAATGGCATTTC
AGACTGGGGGTCTAGAGTATTTTAGGGAGGGNTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATAC
CGAAGGCGAAGGCAGCCCCTTGGGGAGTACTGACGCTCAAGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTG
GTAGTCCACGCTGTAAACGATGTCGATTTGGGGATTGGCTNTAAGCTTGGTGCCCGAAGCTAACGTGATAAATCGACCG
CCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATT
CGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCATAGGAGAGTGNNGAGATGTGCTTGNGCCTTCGGGAACTATG
AGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCT
TTGTTGCCATCGATTAGGTCGGGAACTCAAAGGAGACTGCCGGTGATAAACCGGAGGAAGGTGGGGATGACGTCAAGTCA
TCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGTGCATACAGAGAGCGCGAGGCGAGCGAGGCCGGNGGTGGAGTGAAT
CTCAGAAAGTGCATCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGGATCAG
AATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATA
```

```
AATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGGTTGTACCAGAAGTAGATA
GCTTAACCGNAAGGGGGGGCGTTTACCACGGTATGATTCA>
>NR_118759.1 Actinobacillus capsulatus strain CCUG 12396 16S ribosomal RNA, part
ial sequence
ATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAAAGTA
CTTGTACTTTTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGGAATCTGGCTTATGGAGGGGGATAACTACTGGAAA
CGGTAGCTAATACCGCGTAATATCGGAAGATTAAAGGGTGGGACCATTTGGCCACCTGCCATAAGATGAGCCCAAGTGGG
ATTAGGTAGTTGGTGGGGTAAGGGCCTACCAAGCCTNCGATCTCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAC
TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGCCG
CGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTGATGAGGAAGGCGGTTAGTTTAATAGACTAATTGATTGA
CGTTAGTCACAGAAGAAGCACCGGCTNACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGNGAGCGTTAATCGGAATA
ACTGGGCGTAAAGGGCACGCAGGCGGATTTTTAAGTGAGATGTGAAAGCCCCGAGCTTNACTTGGGAATTGCATTTCAGA
CTGGAAGTCTNGAGTATTTTAGGGAGGGGTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGA
AGGCGAAGGCAGCCCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTNGATACCCTGGTA
GTCCACGCTGTAAACGCTGTCGATTTGGGGATTGGGCTTTNAGCTTGGTGCCCGTAGCTAACGTGATAAATCGACCGCCT
GGGGAGTACGGCCGCAAGGTTAAAACTCAAATNAATTGACGGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTNATTCGA
TGNAACGCGAAGAACCTTACCTACTCTTGACATCCTCAGAATCCGATAGAGATATTGGAGTGCCTTTGGGAACTGAGAGA
CAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTNAGTCCCGCAACGAGCGCAACCCTTATCCTTTG
TTGCCAGCGATTAGGTCGGGAACTCAAAGGAGACTGCCAGTGATNAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCA
TGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGTGCATACAGAGGGCGACAAACCTGCGAGGGGGAGTGAATCTC
AGAAAGTGCATCTAAGTCCGGATTGGAGTCTGCNACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCAGAAT
GTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGGTTGTACCAGNAGTAGATAGCT
TAACCGAGAGGGGGGGGCGTTTACCACGGTATGATTCATGACTGGGG)
>NR_118758.1 Actinobacillus lignieresii strain NCTC 4189 16S ribosomal RNA, part
ial sequence
ATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAACGGTAACGGGAAGGGA
GCTTGCTTTCTTTGCCCGACGAGTGGCGGACGGGTGAGTAATGCTTGGGAATCTGGCTTATGGAGGGGGATAACTACGGGA
AACTGTAGCTAATACCGCGTAATATCTTAGGATTAAAGGGTGGGACTTTCGGGCCACCTGCCATAAGATGAGCCCAAGTG
GGATTAGGTAGTTGGTTAGGTAAAGGCTGACCAAGCCGACGATCTCTAGCTNGTCTGAGAGGATGACCAGCCACACTGGA
ACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGC
CGCGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTAGCGAGGAAGGTATCAAATTTAATAGATTTGGTAATT
GACGTTAACTACAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAA
TAACTGGGCGTAAAGGGCACGCAGGCGGTTGATTAAGTGAGATGTGAAAGCCCCGGGCTTNACCTGGGAATTGCATTTCA
TACTGGTCAACTAGAGTACTTTAGGGAGGGGTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACC
GAAGGCGAAGGCAGCCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGG
TAGTCCACGCTGTAAACGCTGTCGATTTGGGGATTGGACTTTNAGTCTGGTGCCCGAAGCTAACGTGATAAATCGACCGC
CTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTC
GATGCAACGCGAAGAACCTTACCTACTCTTGACATCCATGGAATCTTGTAGAGATACGAGAGTGCCTTCGGGAACCATGA
GACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTNAGTCCCGCAACGAGCGCAACCCTTATCCTT
TGTTGCCAGCGATTNGGTCGGGAACTCAAAGGAGACTGCCGGTGATNAACCGGAGGAAGGTGGGGATGACGTCAAGTCAT
CATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAGAGGGAAGCAAGATGGCGACATGGAGCAAATC
TCACAAAGTACGTCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCAGA
ATGTTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATAG
CTTAACCGCAAGGGGGGGCGTTTACCACGGTATGATTCATGACTGGGG)
>NR_118755.1 Actinobacillus capsulatus strain NCTC 11408 16S ribosomal RNA, part
ial sequence
ATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAAAGTA
CTTGTACTTTTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGGAATCTGGCTTATGGAGGGGGATAACTACTGGAAA
CGGTAGCTAATACCGCGTAATATCGGAAGATTAAAGGGTGGGACCATTTGGCCACCTGCCATAAGATGAGCCCAAGTGGG
ATTAGGTAGTTGGTGGGGTAAGGGCCTACCAAGCCTNCGATCTCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAC
TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGCCG
CGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTGATGAGGAAGGCGGTTAGTTTAATAGACTAATTGATTGA
CGTTAGTCACAGAAGAAGCACCGGCTNACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATA
ACTGGGCGTAAAGGGCACGCAGGCGGATTTTTAAGTGAGATGTGAAAGCCCCGAGCTTNACTTGGGAATTGCATTTCAGA
CTGGAAGTCTNGAGTATTTTAGGGAGGGNTAGAATTCCACGTGTAGCGGTNAAATGCGTNGAGATGTNGAGGAATACCGA
AGGCGAAGGCAGCCCCCTTGGGAATGTACTGACGCTCATGTGCNAAAGCGTGGGGAGCAAACAGGATTNGATACCCTGGTA
GTCCACGCTGTAAACGCTGTCGATTTGGGGATTGGGCTTTNAGCTTGGTGCCCGTAGCTAACGTGATAAATCGACCGCCT
GGGGAGTACGGCCGCAAGGTTAAAACTCAAATNAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGA
TGNAACGCGAAGAACCTTACCTACTCTTGACATCCTCAGAATCCGATAGAGATATTGGAGTGCCTTTGGGAACTGAGAGA
CAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGNTGGGTTNNGTCCCGCAACGAGCGCAACCCTTATCCTTTG
TTGCCAGCGATTAGGTCGGGAACTCAAAGGAGACTGCCAGTGATNAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCA
```

TTGCCAGCGATTAGGTCGGGAACTCAAAGGAGACTGCCAGTGATNAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCA
TNGNCCTTACGAGTAGGGCTACACACGTGCTACAATGGTGCATACAGAGGGCGACAAACCTGCGAGGGGGGAGTGAATCTC
AGAAAGTGCATCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCAGAAT
GTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATAGCT
TAACCGAGAGGGGGGCGTTTACCACGGTATGATTCA>
>NR_118760.1 Actinobacillus equuli strain NCTC 8529 16S ribosomal RNA, partial s
equence
ATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAACGGTAGCACGAAGGAG
CTTGCTCCTTGGGTGACGAGTGGCGGACGGGTGAGTAATGCTTGGGAATCTGGCTTATGGAGGGGGATAACTACGGGAAA
CTGTAGCTAATACCGCGTAATATCTTCGGATTAAAGTGTGGGACCGCAAGGCCACATGCCATGAGATGAGCCCAAGTGGG
ATTAGGTAGTTGGTGGGGTAAAGGCCTACCAAGCCGACGATCTCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAC
TGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGGAACCCTGATGCAGCCATGCCG
CGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTAGCGAGGAAGGTATCAAATTTAATAGATTTGGTAATTGA
CGTTAACTACAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATA
ACTGGGCGTAAAGGGCACGCAGGCGGTTGATTAAGTGAGATGTGAAAGCCCCGGGCTNNACCTGGGAATTGCATTTCATA
CTGGTCAACTAGAGTACTTTAGGGAGGGGTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGA
AGGCGAAGGCAGCCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTA
GTCCACGCTGTAAACGCTGTCGATTTGGGGATTGGGACTNNNAGTCTGGTGCCCGAAGCTAACGTGATNAATCGACCGCCT
GGGGAGTACGGCCGCAAGGTTAAAAACTCAAATNAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGA
TNNAACGCGAAGAACCTTACCTACTCTTGACATCCATGGAATCTTGTAGAGATACGAGAGTGCCTTCGGGAACCATGAGA
CAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTNAGTCCCGCAACGAGCGCAACCCTTATCCTTTG
TTGCCAGCCGGTTCGGCCGGGAACTCAAAGGAGACTGCCAGTGATNAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCA
TGGNCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAGAGGGAAGCAATATGGCGACATGGAGCAAATCTC
ACAAAGTACGTCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCAAATCAGAAT
GTTGCGGTGAATACGTTCCCGGGCNTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATAGCT
TAACCGCAAGGGGGGCGTTTACCACGGTATGATTCAT>
>NR_029171.1 Aggregatibacter actinomycetemcomitans strain 277/49-50(8) 16S ribos
omal RNA, partial sequence
TTGAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGGACGGTAGCAGGTA
AGTACTTGTACTTATGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGGNATCTGTCTTATGGAGGGGGATAACGACGG
GAAACTGTCGCTAATACCGCGTAGAATCGGGAGATGAAAGTGTGGGACCTTCGGGCCACATGCCATAGAATGAGCCCAAG
TGTGATTAGGTAGTTGGTGGGGTAAAGGCCTACCAAGCCGACGATCGCTAGCTGGTCTGAGAGGATGGCCAGCCACACCG
GGACTGAGACACGGCCCNGNCTCCTACGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGCGAACCCTGACGCAGCCAT
GCCGCGTGAATGAAGAAGGCCTTCGGGTTGTAAAGTTCTTTCGGTGATGAGGAAGGTTGGTGTGTTAATAGCATGCCAAA
TTGACGTTAAATACAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGGGGGTGCGAGCGTTAATCGG
AATAACTGGGCGTAAAGGGCACGTAGGCGGACCTTTAAGTGAGGTGTGAAATCCCCGGGCTTAACCTGGGNATTGCATTT
CATACTGGGGGTCTGGAGTACTTTAGGGAGGGNTAGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATA
CCGAAGGCGAAGGCAGCCCCTTGGGGATGTACTGACGCTGATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT
GGTAGTCCACGCTGTAAACGGTGTCGATTTGGGGATTGGGGTTTAGCCCTGGTGCCCGAAGCTAACGTGATAAATCGACC
GCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAAT
TCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGAAGAACTCAGAGATGGGTTTGTGCCTTAGGGAGCTTT
GAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTNAGTCCCGCAACGAGCGCAACCCTTATCC
TTTGTTGCCAGCGACTTGGTCGGGGAACTCAAAGGAGACTGCCGGTGATAAACCGGAGGAAGGTGGGGATGACGTCAAGTC
ATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGACGGGTATACAGAGGGCTAACCAACCACGCGATGGGGAGTGAA
TCTCAGAAAGTACGTCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCA
GAATGTTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGAT
AGCTTAACCGAGAGGGGGGCGTTTACCACGGTATGATTCANGAC>
>NR_044754.1 Aerococcus urinae 16S ribosomal RNA, partial sequence
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACCGACGAAGTGCTTGCA
CTTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTAAGGAACCTACCGATAAGCGGGGGACAACATCCGGAAACGGGTG
CTAATACCGCATAGGAAAGGTCACCNCATGGTGACCTTTGGAAAGACGGCTTTGCTGTCACTTATCGATGGCCTTGCNGT
GCATTAGCTCGTTGGTGGGGTAACGGCCTACCAAGGCAATGATGCATAGCCGACCTGAGAGGGTAATCGGCCACATTGGG
ACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGC
CGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTAAGAGAAGAACAAATTGGAGAGTAACTGCTCCAGTCTT
GACGGTATCTTACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAT
TTATTGGGCGTAAAGGGGGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAAGTGCATTGGA
AACTGGGNAACTTGAGTACAGAAGAGGAAAGTGGAACTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACACC
AGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCTGAGGCCCGAAAGCGTGGGTAGCAAACAGGATTAGATACCCTGG
TAGTCCACGCCGTAAACGATGAGTGCTANGTGTTGGAGGGTTTCCACCCTTCAGTGCCGGAGTTAACGCATTAAGCACTC
CGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGACCGCACAAGCGGTGGAGCATGTGGTTTAAT
TCGAAGCAACGCGAAGAACCTTACCAAGTCTTGACATCCGTTGACCACTCTAGAGATAGAGCTTTCTTCGGGGACNAAGT

```
C:\WINDOWS\system32\cmd.exe
```

```
AGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCTGAGGCCCGAAAGCGTGGGTAGCAAACAGGATTAGATACCCTGG
TAGTCCACGCCGTAAACGATGAGTGCTANGTGTTGGAGGGTTTCCACCCTTCAGTGCCGGAGTTAACGCATTAAGCACTC
CGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGACCGCACAAGCGGTGGAGCATGTGGTTTAAT
TCGAAGCAACGCGAAGAACCTTACCAAGTCTTGACATCCGTTGACCACTCTAGAGATAGAGCTTTCTTCGGGGGACNAAGT
GACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTGTT
AGTTGCCAGCATTNAGTTGGGCACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAAGGCGGGGATGACGTCAAATCATC
ATGCCCCTTATGACTTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTCAGCGACCTTGTGAAAGCAAGCGAATC
TCTTAAAGCCATTCTCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGCGGATCAGC
ACGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCTGAAGTCGGTGA
GGTAACCTTTGGAGCCAGCCGCCGAAGGTNNGACAGATATT>
>NR_044742.1 Afipia felis ATCC 53690 strain B-91-007352 16S ribosomal RNA, parti
al sequence
TTCCAACTTGAGAGTTTGATCCTGGCTCAGAGCGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGGGCGTAGCA
ATACGTCAGCGGCAGACGGGTGAGTAACGCGTGGGAACATACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAAT
ACCGAATAAGCCTTACGGGGAAAGATTTATCGCCGAAAGATTGGCCCGCGTCTGATTAGCTTGTTGGTGAGGTAACGGCT
CACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCNNAACTCCTACGGG
AGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTNATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTG
TAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAA
TACGAAGGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGTGCGTAGGCGGGTCTTTAAGTCAGAGGTGAAAGCCT
GGAGCTCAACTCCAGAACTGCCTTTGATACTGAGGATCTCGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTG
AAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTGACGCTGAGGCACGAAAGCGTG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATGCCAGCCGTTGGGGAGTTTACTCTTCAGT
GGCGCAGTTAACGCTTTAAGCATTCCGGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCGCA
CAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATGTCCAGGACCGGTCGCAG
AGATGTGACCTTCTCTTCGGAGCCTGGAGCACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAA
GTCCCGCAACGAGCGCAACCCTNGTCCTTAGTTGCTACCATTTAGTTGAGCACTCTAAGGAGACTGCCGGTGATAAGCCG
CGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGCGGGTGACAATG
GGAGCAAAGGCCGCGAGCCCTAGCAAATCTCAAAAAAGCCGTCTCAGTTCGGATTGGACTCTGCAACTCGAGTCCATGAAGT
TGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATG
GGAGTTGGTTTTACCTGAAGACGGTGCGCTAACCCGCAAGGAGGCAANCGGCCACGCTAGGGTCA>
Press any key to continue . . .
```

Self-Assessment:

1) How do I think I went with this assignment?

> **I think I did well on this assignment I managed to complete all 3 parts to full functionality in a decent amount of time. All my error messages were good, my code was well constructed and all up I achieved all of what the task asked from me.**

2) What did I find difficult in this assignment?

> **The assignment as a whole wasn't too bad the hardest part was trying to figure out how to get the sequence to come along with the actual inputted id in parts 2 and 3. Although the actual hardest part was outside the assignment in trying to keep up with weekly ams tasks throughout the duration of the assignment. Also exceptions were weird because I already had all my error stuff done before we learnt them making me go back and do something more complex that did the same thing.**

3) What would I do differently next time?

> **Next time I would use the time in the tutorials more so to focus on the assignment instead of the tutorial & ams exercises, (sorry Ben ily)**

4) Were there any bugs in my assignment, if so what were they?

> No bugs were found in the testing of my assignment.

CAB201 Programming Principles - Semester 2, 2019

Project – Genomic Sequence Retrieval

CRA

You must test your program in one of the CAB201 labs, using Visual Studio and .NET Framework prior to submission. Markers will attempt to compile and run your code in that environment only. It is your responsibility to ensure that your code compiles and runs on PCs in the QUT computer labs of CAB201 as the target platform. Code that does not compile will not receive any marks. Plan for time to test your submission.

Note – you must test it from a USB Disk (disk on key), and you must make sure that you load the assignment from the USB. Make sure that all files were uploaded. It is recommended that you try running your assignment in a CAB201 lab PC **by starting a new session** and following your user manual instructions (including unzipping the assignment directory that you intend to upload) to ensure that it is all in order.

Part I and Part II must be submitted **twice each** by the specified due dates on the assignment document. **A week before each part is due, a progress report must be submitted, and a week later, the final submission.**

Progress Submission: This progress report must contain all the work done up until that date, along with a statement of completeness explaining what has been done and what still needs doing, and self-reflection on progress. This is worth 5 marks from 100 for each part. No marks will be awarded unless the final submission of the respective Part is made. The progress reports will not be assessed in the same way as the final submission – it is intended to checkpoint your work and will be marked for completeness and accuracy of the statement of completeness and self-reflection.

Failure to submit the progress report on time will result in loss of 20 marks (20%) of the score – even if the final submission of the assignment part is made on time!!

Final Submission: Each assignment part will be marked out of 100. You are required to submit a report explaining your submission, which will include a statement of completeness, a short self-reflection and a brief user manual. Instructions for this report will be included in the submission guide. This statement of completeness will need to accurately state the functionality which has been implemented. There will be a penalty of 7 marks for missing statement of completeness, and a penalty of 3 marks for missing self-reflection. For incomplete statements of completeness, a penalty of 1 mark is deducted per missing section (up to -6 marks). For an incomplete self-reflection, a penalty of 1.5 marks is deducted.

Note: some additional marks are available in this assignment for extended functionality in Part-II. While these additional grades cannot allow your score for part-II to go above 100%, they can make up for potential weaknesses in other areas of your assignment. Additional marks are indicated with a '+'.

Section		Score	Value in CAB201 grade
	Part I	/100	/20%
	Part II	/100	/30%
	Total		**/50%**

Penalties		Part I Marks Lost:	Part II Marks Lost:
If the statement of completeness, self-reflection is incomplete or missing, the student will lose marks		**/-10**	**/-10**
Statement of Completeness	Missing	/-7	/-7
	Incomplete (up to -6)	/-6	/-6
	Total	**/-7**	**/-7**
Self-Reflection	Missing	/-3	/-3
	Incomplete	/-1.5	/-1.5
	Total	**/-3**	**/-3**

Missing Progress report	**-0 from 100**	**-0 from 100**

Code Quality	Part I Marks Available:	Part II Marks Available:
To score points in this section, the student must follow the code quality guidelines as specified in the C# Coding Style Guide on Blackboard	**28/30**	/30
Maintained consistent, clear, and meaningful standard in variable and method naming. No magic numbers.	3/3	/3
Well structured – consistent and appropriate white spacing, line length, indentation, and separation into files within the project (i.e. one class per file)	2/2	/2
Well commented – class header comment at beginning of each class, comment before every method, and in-line comments to explain complex or not easily discernible code. In-line comments are not excessive.	4/4	/4
The DRY principle (Don't repeat yourself) is followed where appropriate	3/3	/3
Methods are single purpose and clear	4/4	/4
Classes are well designed, with high cohesion and low coupling	6/8	/8
Classes are separated into reusable modules where appropriate	3/3	/3
Exceptions are thrown and handled appropriately	3/3	/3

Basic Functionality		Part I Marks Available:	Part II Marks Available:
To score marks in this section, your program must be able to be run from the command line with the appropriate arguments.		**15/15**	**/15**
Basic Output	The program displays the data from the file	1/1	/1
	The program displays the appropriate line	1/1	/1
	The correct amount of information is displayed, e.g. only the relevant entries	1/1	/1
	The correct level, provided as a command line flag *-levelN*, is executed	1/1	/1
	The program **does not** store the whole file in memory, instead it accesses the file on disk	5/5	/5
	Total:	**9/9**	**/9**
Error Handling	A clear error message is displayed when an incorrect number of arguments is provided	2/2	/2
	A clear error message is provided when an incorrect flag is provided (e.g. not -level1, etc.)	2/2	/2
	A clear error message is provided when the input file doesn't exist, or is incorrectly formatted	2/2	/2
	Total:	**6/6**	**/6**

Part I		Marks Available:
To score marks in this section, your program must be able to run levels 1-3.		**55/55**
Level 1	The program displays the correct data from the file	4/4
	The program displays the appropriate number of entries	4/4
	A clear error message is provided when a non-existent starting sequence number is requested	1/1
	A clear error message is provided when a problematic number of sequences is provided	1/1
	Total:	**10/10**
Level 2	The program correctly locates and prints the requested sequence	12/12
	A clear error message is provided when the sequence-id does not exist	3/3
	Total:	**15/15**
Level 3	The program correctly accepts and interprets a list of instructions as specified by the query file	10/10
	The output is stored in the results file	10/10
	If a sequence is not found, the error message is displayed on the console and not in the results file	5/5
	Total:	**25/25**
Progress Report	Statement of completeness 2.5 Self-reflection 2.5	**5/5**

Part I final score	**98 /100**

Part II		Marks Available:
To score marks in this section, your program must be able to run levels 4-7.		**/55**
Level 4	The program creates a file as specified by the command line arguments	/2.5
	The index file contains a list of all the sequence ids with the appropriate byte-offset	/2.5
	The searching program makes use of the created index file to execute a number of queries	/2
	A clear error message is provided when the index file does not exist	/1
	Clear error messages are provided when the query file cannot be found, or when a bad query is given, like in Level 3	/2
	Total:	**/15**
Level 5	The program correctly locates and prints the requested sequence ids	/10
	A clear error message is provided when the sequence does not exist	/5
	Total:	**/15**
Level 6	The program correctly locates and prints the requested sequence ids	/15
	A clear error message is provided when the keyword does not exist	/5
	Total:	**/20**
Level 7 (Optional)	The program correctly decodes the expression given and identifies any matching sequences	/5
	A clear error message is provided when the sequence does not exist	/5
	Total:	**/+10**
Progress Report	Statement of completeness 2.5 Self-reflection 2.5	**/5**

Note: maximum mark in part II is 55, but level 7 has 10 additional marks that can offset lost marks.

Part II final score	/100

Test planning

Category	Input	Expected result	Actual output	Passed
Wrong amount of inputs	-level1 16S.fasta 1	Please enter more inputs	Please enter more inputs	Y
Wrong files	-level1 315 1	Please enter a valid file name	Please enter a valid file name	Y
Limit breaker	-level1 16S.fasta 400001 2	Please enter a line number in the length of the file	Please enter a line number in the length of the file	Y
Etc	-level1 16S.fasta 40000 2	Please enter an odd line number	Please enter an odd line number	Y
Wrong amount of inputs	-level 3 16S.fasta Query.txt results.txt p	Please enter less inputs	Please enter less inputs	Y
Limit breaker	-level1 16S.fasta 3 20000000	Please enter a number of sequences that doesn't exceed file length	Please enter a number of sequences that doesn't exceed file length	Y
Wrong files	-level 3 16S.fasta Q.txt results.txt	Please enter a valid file name	Please enter a valid file name	Y
Level 1 working output	-level1 16S.fasta 11 1	>NR_118908.1 Amycolatopsis methanolica 239 16S ribosomal RNA, partial sequence GTGAGTGGCGAACG….	>NR_118908.1 Amycolatopsis methanolica 239 16S ribosomal RNA, partial sequence GTGAGTGGCGAACG….	Y
Level 2 found	-level2 16S.fasta >NR_118908.1	>NR_118908.1 Amycolatopsis methanolica 239 16S ribosomal RNA, partial	>NR_118908.1 Amycolatopsis methanolica 239 16S ribosomal RNA, partial	Y
Level 2 not found	-level2 16S.fasta >NR_9999999	Sequence >NR_9999999 Was not found	Sequence >NR_9999999 Was not found	Y
Level 3 not found	-level 3 16S.fasta Query.txt results.txt	Sequence >NR_9999999 Was not found	Sequence >NR_9999999 Was not found	Y
Level 3 found	-level 3 16S.fasta Query.txt results.txt	Please press enter to continue	Please press enter to continue	Y