**oriC\_finder build instruction**

**Language:**

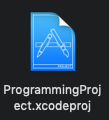
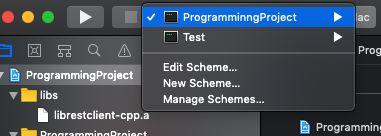
C++

**Running Environment:**

Mac OS

**How to build the program:**

The program can be built in two ways:

1. CMakeList.txt is provided in the project directory, so Makefile can be generated by using CMake under any platforms. In order to save your time on checking, I have generated the Makefile, and it is under “bin” folder of the project directory. Change the directory to that folder and execute command “make all”. Both the test program and the oriC finder program will be built. The finder program is called “ProgrammingProject” and the test program is called “Test”. Run the finder program by using “./ProgrammingProject”. Run the test by using “./Test”.
2. Xcode can be used for opening the project and built the code. The project file has been provided. If you click on that, Xcode will be opened automatically and all files will be loaded for you. Click the run button on the top left, the program will be started. Either the test program and the finder program can be selected from here  to run.
3. A program has been built for you, and it is located in the pre-built folder.

**Usage:**

./ProgrammingProject <accession#>

Ex: ./ProgrammingProject NC\_000964.3

If the program is run in this way, all output will be printed to the console

./ProgrammingProject <accession#> > <outputfilename>

Ex: ./ProgrammingProject NC\_000964.3 > result.txt

If the program is run in this way, the output will be navigated to the file you provide.

**Sample Output:**

downloading the gene file...

sequence.fasta is downloaded

>Computing the distribution of the sequence...

A%=28.182733%

C%=21.806687%

G%=21.707721%

T%=28.302858%

GC%=43.514408%

>locatinng oriC\_1 of B. subtilis subsp. subtilis str. 168, NC\_000964.3...

start:0 end:188

start:-4215419 end:-1 (reverse complementary)

>locatinng oriC\_2 of B. subtilis subsp. subtilis str. 168, NC\_000964.3...

start:4213639 end:4214265

start:-1342 end:-1968 (reverse complementary)

>locatinng dnaA gene of B subtilis subsp. subtilis str. 168...

start:4214265 end:0

start:-1 end:-1342 (reverse complementary)

Relative locations: OriC1 OriC2 DnaA (Circular)

>DnaA boxes oriC\_1 of B. subtilis subsp. subtilis str. 168, NC\_000964.3

caggaccggggatcaatcggggaaagtgtgaataacttttcggaagtcatacacagtCTG

TCCACATGTGGATAGgctgtgtttcctgtctttttcacaacTTATCCACAaatccacagg

ccctactattacttctactattttttataaatatatatattaatacattatccgttagga

ggataaaa

>DnaA boxes oriC\_2 of B. subtilis subsp. subtilis str. 168, NC\_000964.3

ttatgacacctccctcgaggaatagctgttaaagacagtcttacttattatatttgcgtt

acctattcattgtcaacttcactagtgcttttatttcttgcaaccataataggataccat

accttttcaactttcgaaaccttattttttagattccttaattttacggaaaaaagacaa

attcaaacaatttgcccctaaaatcacgcaTGTGGATATctttttcggctttttttaGTA

TCCACAgaggTTATCGACAacattttcacattaccaaccccTGTGGACAAggttttttca

acaggttgtccgcttTGTGGATAAgattgtgacaaccattgcaagctctcgtttattttg

gtattatatttgtgttttaactcttgattactaatcctacctttcctctTTATCCACAaa

gTGTGGATAAgttgtggattgatttcacacagcttgtgtagaaggTTGTCCACAagttgt

gaaatttgtcgaaaagctatttatctactatattatatgttttcaacatttaatgtgtac

gaatggtaagcgccatttgctctttttttgtgttctataacagagaaagacgccattttc

taagaaaaggagggacgtgccggaag

>dnaA gene of B subtilis subsp. subtilis str. 168

ATGGAAAATATATTAGACCTGTGGAACCAAGCCCTTGCTCAAATCGAAAAAAAGTTGAGC

AAACCGAGTTTTGAGACTTGGATGAAGTCAACCAAAGCCCACTCACTGCAAGGCGATACA

TTAACAATCACGGCTCCCAATGAATTTGCCAGAGACTGGCTGGAGTCCAGATACTTGCAT

CTGATTGCAGATACTATATATGAATTAACCGGGGAAGAATTGAGCATTAAGTTTGTCATT

CCTCAAAATCAAGATGTTGAGGACTTTATGCCGAAACCGCAAGTCAAAAAAGCGGTCAAA

GAAGATACATCTGATTTTCCTCAAAATATGCTCAATCCAAAATATACTTTTGATACTTTT

GTCATCGGATCTGGAAACCGATTTGCACATGCTGCTTCCCTCGCAGTAGCGGAAGCGCCC

GCGAAAGCTTACAACCCTTTATTTATCTATGGGGGCGTCGGCTTAGGGAAAACACACTTA

ATGCATGCGATCGGCCATTATGTAATAGATCATAATCCTTCTGCCAAAGTGGTTTATCTG

TCTTCTGAGAAATTTACAAACGAATTCATCAACTCTATCCGAGATAATAAAGCCGTCGAC

TTCCGCAATCGCTATCGAAATGTTGATGTGCTTTTGATAGATGATATTCAATTTTTAGCG

GGGAAAGAACAAACCCAGGAAGAATTTTTCCATACATTTAACACATTACACGAAGAAAGC

AAACAAATCGTCATTTCAAGTGACCGGCCGCCAAAGGAAATTCCGACACTTGAAGACAGA

TTGCGCTCACGTTTTGAATGGGGACTTATTACAGATATCACACCGCCTGATCTAGAAACG

AGAATTGCAATTTTAAGAAAAAAGGCCAAAGCAGAGGGCCTCGATATTCCGAACGAGGTT

ATGCTTTACATCGCGAATCAAATCGACAGCAATATTCGGGAACTCGAAGGAGCATTAATC

AGAGTTGTCGCTTATTCATCTTTAATTAATAAAGATATTAATGCTGATCTGGCCGCTGAG

GCGTTGAAAGATATTATTCCTTCCTCAAAACCGAAAGTCATTACGATAAAAGAAATTCAG

AGGGTAGTAGGCCAGCAATTTAATATTAAACTCGAGGATTTCAAAGCAAAAAAACGGACA

AAGTCAGTAGCTTTTCCGCGTCAAATCGCCATGTACTTATCAAGGGAAATGACTGATTCC

TCTCTTCCTAAAATCGGTGAAGAGTTTGGAGGACGTGATCATACGACCGTTATTCATGCG

CATGAAAAAATTTCAAAACTGCTGGCAGATGATGAACAGCTTCAGCAGCATGTAAAAGAA

ATTAAAGAACAGCTTAAATAG

Execution time:9539ms

**Test result:**

All operations on the Sequence class has been tested.



**How the application is designed:**

The program uses restclient-cpp to send the HTTP request to the GenBank Nucleotide DB. After the response is received from the internet, sequence data will be parsed, and all genes will be stored as the Sequence object in the memory.

The Sequence object involves two arrays for storing genes. One is for the original order, and another is for the reversed complementary order. When the program locates the gene section, the index of the gene section will start from 0 in the original order and from -1 in the reversed complementary order. Also, if the location is (0, 13), it means the section is between 0 and 12 inclusively. If the location is (-1, -13), it means the section is between 0 and 11 in the reversed complementary order. The start index and the end index must be with the same sign.

When the program searches for DNA boxes, an array of locations will be returned as the result. By using these locations, certain sections will be highlighted in the output.