CIS 554

Project 3: Pairwise Alignment

Wenyuan Yu

1. Introduction.

In this project, Java programming language was used for the pairwise alignment. The following files are included in the zip file:

(1) *PairwiseAlignment.java*

This Java program imports Fasta files from local working directory. Two default files (trna.fasta and test.fasta) are included which are copied from the align folder of the project 3.

(2) *PairwiseAlignmentURLFasta.java*

This Java program imports Fasta files directly from URL which would be more convenient to retrieve the file. The URLs for the two files are shown as below.

trna.fasta http://ix.cs.uoregon.edu/~wyu/trna.fasta

test.fasta http://ix.cs.uoregon.edu/~wyu/test.fasta

(3) *trna.fasta*: testing file copied directly from the align folder without any modification

(4) *test.fasta*: testing file copied directly from the align folder without any modification

(5) *short writeup* (this file)

2. Program instruction.

Basically, the two Java programs (*PairwiseAlignment.java, PairwiseAlignmentURLFasta.java*) are almost the same, except the way to import Fasta file.

The program is easy to test, and the result will be produced after the program is run. If *PairwiseAlignment.java* is tested, the testing Fasta file (*trna.fasta* or *test.fasta*) needs to put in the right directory in case the system cannot find the file specified. There will not be any problem if running *PairwiseAlignmentURLFasta.java*, since testing files are directly imported from URL.

The scores for the *gap*, *mismatch* or even *match* can be changed at the start of the program. If you want to test other sequence files, simply change it at line 28. In addition, each step of alignment can be shown in the program by simply uncomment line 148 and 149 in the both Java program.

The output of the program includes three parts:

(1) The original two sequence files.

(2) The *n* * *m* matrix: showing the cost of traceback. For instance, the output of matix for the two sequences of *test.fasta* are shown below.

```
      A  T  T  C  A
   0  2  4  6  8 10
A  2  0  2  4  6  8
T  4  2  0  2  4  6
G  6  4  2  1  2  4
T  8  6  4  3  2  3
A 10  8  6  5  4  2
```

(3) The output sequences of after pairwire alignment. For instance, the sequences of original and after alignment are as follows.

```
original seq1 is
GGTGAGGTGGCCGAGAGGCTGAAGGCGCTCCCCTGCTAAGGGAGTATGCGGTCAAAAGCTGCATCCGGGGTTCGAATCCCCGCCTCACCGCCA
original seq2 is
GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCGAAAACCGGAGTAGGGGCAACTCTACCGGGGGTTCAAATCCCCCTCTCTCCGCCA

The two pairs after Pairwise Alignment are:
GGTGAGGTGGCCGAGAGGCTGAAGGCGCTCCCCT-GCTAAG-GGAGTATGCGGTCAAAAGCTGCATCCGGGGTTCGAATCCCCGCCTCACCGCCA
GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCGAAAACCGGAGTAGG-GG-CAACT-CTAC--CGGGGGTTCAAATCCCCCTCTCTCCGCCA
```