T.C.
Marmara University
Faculty of Engineering
CSE4065 Computational Genomics
Programming Assignment 2
Report

Group Members:

Ayberk Ömer Altuntabak - 150119640
Bilal Tan - 150119630

In this assignment, we are going to implement a dynamic pairwise sequence alignment algorithm and test with the given example test sequences. Our goal with this algorithm is to find the optimum alignment on the test sequences.

We will use predetermined score values for this algorithm.
    Default values for scores are as follows:
        • Match score: +2
        • Mismatch score: -1
        • Gap opening penalty: -1, Gap extension penalty: -0.5

We came up with an algorithm based on what we learned in the lesson and the research we did. We did not use the linear scoring model here, as requested by us for the project.

We initially created 3 Matrices. These are M, Ix, Iy matrices. Ix matrix is for representing aligning a character Y with a gap (a vertical move). Ix does the same job for X axis. After we initialize matrices, we calculate values of those matrices according to the given configuration. We use match, mismatch, gap open and extension penalties to determine values of M, Ix, Iy. After we calculate the values of those 3 matrices, we can get the score of those matrices and the maximum of them determines the start matrix. Then after, We go through from those three matrices and determine the return matrix by calculating true/false values like if Ix[i][j] score equals M[i][j-1] - g then we apply gaps if necessary then change the matrix pointer.

Algorithm that we applied explained in those slides
https://www.cs.cmu.edu/~ckingsf/bioinfo-lectures/gaps.pdf

The scores we found with our algorithm are as follows; 153, 103, 196, 648.5 and 94. We tried to show some sequence alignment in outputs but test sequence 3 and 4 was too long, their graphical representation is not very good, but in the 1,2 and 5 numbered test sequences graphical representation fits to the screen.

Outputs:

Test Sequence 1:

```
Score of this sequence alignment 153.0
CGG--GTAGTTAACCCT-ACA-GCATAGAGTCGCGAGATAAAGTGCAGGA-GTCTTTCGCGGCAGATTCGTACCTCAACCA-C-GTGCTACTT---TCTGGCATCACGAATCTGCCGCATAGGTCCGTGAGT-CCATATGA
  || |||||| | ||| |||  |||||||||||||| ||| |  ||| |||    || |||   ||| | |||  ||||||  || ||    |||   || || | ||  | ||| |  | |||  | ||| ||||||||
AGGAAGTAGTT-AGCCTAACAGGCATAGAGTCGCGACAT-ATGTG-AAGATGTCATT--CGG--TATTCAAACCTCATGCATCATTGC--CTTGAGTC--GC-TC------CTGGAGCATA-GTCCCTGAGTGCCATATGA
```

Test Sequence 2:

```
Score of this sequence alignment 103.0
-C-GAGACCGA-C-GAAGAGGTTTGGC-CCCAA-CCAGGTTCC-CTGATCACGTAACTT----ACCGGCCAAAAGGACTG--GCCTT-ACTAAGG---CCT-TTGTCTACT-GC-G-G--GT-C--CGG-G-GGC-CGTT--GGT-T-TC-GGCAGAACACTTC
 | | |||||| | ||| ||    ||| |||   |||  ||   ||| |||  |       |     | |    ||| ||    ||   |   || |  |  ||| | | |  | | |  | ||  |  || | | || ||    || | || ||  | | ||
CCTG-GACCGAGCTTAA-A---TT-GCTAGCAATACAGATGCCGCT--TC------CTTGGGGA-GGGTGTGTAGGA-TGTAG-GTTAACGAATGCAAGTTCCGGGGTA-TCGCAGAGTCGTGCTACGGCGTGGCAC-TTAGGGTCTCTCGGGAAAAAGAGTAG
```

Test Sequence 3:

```
Score of this sequence alignment 196.0
GTGT-GGTTGCTTGCATCACTCCGTGTACATGTGACAACCGAACGAGTTGATCCAGCTTTGTTAAGTCAGCTTCGAATG-CGGTAGCTCTC-AAATATGAT-ATGACTCTTGGG-GTAGAT--GCTGG--GGACCTATTGCGCCC--AAAGCGATAT--TCGGGGCA-C
CGGTTTAGGGTACCCTATCAAGGCGAT-AC-TTCGATGCAGTGTG-ATGCCGG-A-G-GTGTCG-GTCAGCATGTGAGA--GCT
 | | ||  ||  ||  | | || ||| ||| ||| |  | |  | ||| || ||    |||| | ||   |||| | |       ||| || || | |  |    | |  || || |||| ||||  || ||| | | ||    |||
  |   ||||    |   |||| | |||| ||  || ||| | ||| ||| || | |  ||| | || ||||| ||||| |   | |
G-GTAGGTT-CGTGAAGCACTCC-TGGAC-TCTGACCAC--AAC-A--TAATCAAGC---G--CAG--AG-TT-GAATGACCAAAGCGCTCAAGCT-T--TCA-CGAAC---GTCTCATATCCGC-GGTCGTA-CAAACGCGCTCTTAAAAC-A-ATCGTCTATCCATC
CGG----GGATACC--A--ATGG-G-TGACATT--A---ACTGGGCAGGCCGGCACGCG-GACGCGACAG-AT-TGAAATGGAT
```

Test Sequence 4:

```
Score of this sequence alignment 648.5
CGGGGA-AAGA-CGGA---ATGCATCG-ACCATCGGACAAT-GCCTCAC-TGGAAGCGCTG--CTGAT-TTTTGCGCAACGAGCCTCGGA-CCTCCCG-C-TCAAACTT--ACGAA--A-ATGACTCCACC-----AG-CACTG-AACCAACTGGCCTCCAG-TGA-AG
AT---AGTAT-CTACAAATCGTT-TGCCGGGAGAAACATT-TGTA-TGGTA-GTCAGC-GTTCTGCACGTCACGTAATCGT--TCACTAGTT-GTGG-GGTA---TACCAGGTC---GTAATGAG--A-T-GT---TAGTA---TGAATCGTTTATAAC-GCTTGTCAT
AGGAG-TTC---C-GAATAATCGTCA-CTAGGTC-AAT-G---GCCCCCTAC-TTGTA--ATAACTAC-GTCTGATTGAGAAA----CAGATCGTTAGTCATC-GGTT-AATAGTCCCGGAAGATAACGGGTT--CTTGCGTTTTTGCGAATACTACTATCTC-ATGGC
GATGG--AG----C-GT---T-T---GGTCCCAT-CC---AGCCGCGCGAGTATCACTTGTTCGCCTGC-AC--CT-GTCCGACCTTTCGA-----TGGGGAGC--TCCTTTCATTGG--TTGTAG-GTAC-TAAGGGTGAGCAATGTCACGTGACCCAA-GGA-GCCG
TGTGTAATTTCCA---CTTGCTCAGAAAAG--CCTC-GACA----AT--CTG-GGA---CCGACACCTGAGTGATG-GC-TTACA-GACCAGGGGAG-GG-GGGCAGGTTCCCTGGCCACAGAATGGCACGCCCTGAGGAGGCACCGGCCCAA-CGTCG-CTGG
  | ||||| |||  | | ||  || ||  |||||  |   |  |||  | ||  || ||||  || ||| ||    || || || |||     ||||||||| || || ||| | || ||| | ||||     || ||| | ||  | || ||| |||  |  ||
   ||  || |  ||| || ||   || | |||||  ||| | || |  ||| || | | || |||  | |||  | | ||| | || |   || |  |  || | ||| || || | |||  |  || |||  || |||  | |  || || ||  |  || ||  |
   |||  ||  |    | ||  | || ||| || | || |||   |   ||  | |||   || ||| ||  | | ||| ||  ||| |   |  |||| | || |  ||| ||  |  |  ||| | | | |||  ||  || | ||  | || | || | ||  ||
     || ||||  ||  | | ||| |  | | |  |  ||  | ||  | ||  || |  | || |  |  |||  |||  |  || | |||  |  || |  |  || ||| |||  | ||| || | ||| | |  |||| || | |  || ||  | |  ||  ||
      |||  ||  |  ||||  |   | |  |  |||   || || |   |   |   || | |  ||  | ||  | | ||  || ||| |  | |  ||||  ||  | ||| |||| | ||  || ||| | ||  | |  || ||| |  | ||   | || ||
-AGTGAGAAGACCGGATATATGCAACGAACAATCGCAAAATAG-CTC-CATGTACGC-CTGTCCCG-TAATACCGCATGCGAGCCTCGAAGCC-CCCGACGTCAAACCTCAAC-AAGTATAGGGTTCCACCGTATGAGCCACAGAATTCAGTT--CCTCC-GCTGATAG
-TCCCGGT-TCCTCCAA-----TCAGC----AG----ATTCTG-ACT--TAGGTCA-CTGGAAGAAACCTGGCGT-ATTGTGAT-GAAATTTCGTGGTGG-ACGCTACC--GTCTAGGTA--GAGCCACTAGTGCATTGTACACTG--TC-TTT-TTCCGGC-TATAAT
CGGAGTTTCAGGCTGACT--T-GCCACCCAGGTCAAATAGTATGTC----ACGGTGTATGATCTCT-CGGGCTGAT--AGCAAGGGGCGG--CGTCGGGCA-CTCCTTGAATTGACC------AT------TTGGCGTGCCTTGTTGC-----CTCCT-TC-CTCAAGC
-AGGGTTAGGCTTCAGTAAGTGTGGGGGTGGC-TGCCGAAAGACGGGC---T-CGGGGTGTT-GCC--CAACGTCTGGT-CGA-CTTCCCATAATCTGGGCA-CAAGACGT--A-TGGACCCGT-GCG--CTTAATGGT--GC-AT-TTCACGCGAGAAACGGAGGCAG
-GTG------CCATCCGCGGC---GAAAAGATCATCAGACATGACATAAC-GAGGAACCCG-ATCC-GGAGTGAATACGTCACATG-CCAGGGG-GCGGTGTGCA-GTT--CTCGAC-CGGAAACGCCCACCC--ACGACGTACC-GCCAAAGCGTCGCCTGG
```

Test Sequence 5:

```
Score of this sequence alignment 94.0
AGGC--CGAAAACGTCGCGAAT-T-GACCCTGGCGACGCCGCCGAACGGGACCTCCGTTAGT-G----TGGGAGGTCATCAATC---TCGTTCGCTAGCGGCTGACACCAATCACTATAAGTCTGTCATGAC
   |   ||  ||| || | |||  | ||||||||  || || || ||||| ||  ||  || |     | ||||  | ||||   | ||||  | |||| | ||| |  ||| ||  |||||  ||     || |
---CTTC---AA-GT--CAAATATAGATCCTGGC--CG-CTCC--ACGGG--CT---TAAGTCGTTCTCCGAAGGT-A-CGATCTGGTTGGATGCTTCCGTCT-A-AACAA-GAAGATAA-----TC--G--
```