CS 6670 Advanced Bioinformatics Assignment 4

Gopal Menon

July 20, 2016

1. **Output format**

   The program generates two random nucleotide strings and then finds the alignment score or edit distance using dynamic programming. Then it uses the traceback algorithm to print out the string alignment.

   The program generates the following output

   a) The two strings are printed out

   b) The alignment string or edit distance is printed out

   c) The alignment is then printed out in reverse order of the string. The following convention is used:

      i. If a character in one string is substituted by another, then it is shown in the form $T \times A$, where T in the first string is replaced by an A.

      ii. If a character in one string matches with a character in the other string, then this is shown using the form $C = C$.

      iii. If a character needs to be inserted into the first string to match the second string, then it is shown using the form $\_ \quad C$.

2. **Test Cases and Corresponding Output**

   Here are some instances of randomly generated strings and the output produced by the program.

First string: TCTAAAGCACG
Second string: GGAATTGTGAC

Alignment score: 9

```
G X C
C   _
A = A
_   G
C X T
G = G
_   T
A X T
A = A
A = A
T   _
C X G
T X G
```

First string: TGCCCAATCCCACTTT
Second string: ATTTAGTTGTGGAGTGAGATTCTTG

Alignment score: 17

```
_   G
T = T
T = T
_   C
T = T
C X T
A = A
C X G
C X A
C X G
T = T
_   G
A = A
_   G
_   G
A X T
C X G
C X T
C X T
G = G
_   A
_   T
_   T
T = T
_   A
```

First string: GCTTTCGTACTTTC
Second string: ACGGGCCGCGGATCGTAGCC

Alignment score: 13

```
_   C
C = C
_   G
T X A
T = T
T X G
C = C
_   T
A = A
_   G
_   G
T X C
G = G
_   C
C = C
T X G
T X G
T X G
C = C
G X A
```