T.C. İSTANBUL TİCARET ÜNİVERSİTESİ MÜHENDİSLİK VE TASARIM FAKÜLTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BIL458 BİYOİNFORMATİK

DNA SEKANS HİZALAMA ALGORİTMALARI UYGULAMASI

Alperen KÖYLÜ

Bilgisayar Mühendisliği Programında Hazırlanan

ÖDEV RAPORU

Danışman: Arzu Kakışım

Number of used sequence for comparision is 27. Each sequence is compared by first sequence in order to find the best global and local alignment with respect to their score.

Global Alignment

Needleman-Wunsch algorithm is implemented for finding global alignments score. In the figure below, there are details of the highest scoring pairings found at the end of the Needleman-Wunsch alignment, which operates 104 times using 4 different cost matrices and 26 different sequence pairs.

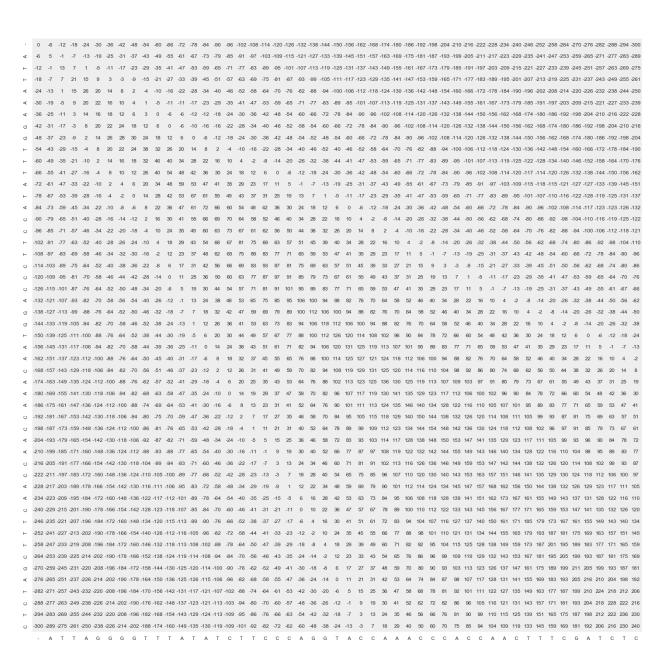
```
MAXIMUM SCORE WITH PROVIDING COST MATRIX
DIRECTION MATRIX:
[ׁ'V' 'D' 'H' 'H' 'H' ... 'H' 'H' 'H' 'H' 'חֹ
 וְיׁט' 'v' 'ס' 'א' 'א' ... 'א' 'א' 'א' 'א' 'א' '
 וֹ'v' 'v' 'v' 'ס' 'h' ... 'א' 'א' 'א' 'א' 'א'
 ['V' 'V' 'V' 'V' 'D' ... 'H' 'H' 'H' 'H' 'H']
 ['V' 'V' 'V' 'V' 'V' ... 'D' 'H' 'H' 'H' 'H']
 ['v' 'v' 'v' 'v' 'v' ... 'v' 'ס' 'H' 'H' 'H']
 רְׁ' 'א' 'v' 'v' 'v' ... 'v' 'v' 'ם' 'H' 'H']
 רְׁיִע' 'v' 'v' 'v' 'v' ... 'v' 'v' 'v' 'b' 'H'וָ
וֹיִט' 'v' 'v' 'v' 'v' ... 'v' 'v' 'v' 'v' 'ס' וֹן
SCORING MATRIX:
[[ 0. -6. -12. -18. -24. ... -276. -282. -288. -294. -300.]
        5. -1. -7. -13. ... -265. -271. -277. -283. -289.]
 [ -12. -1. 13. 7. 1. ... -251. -257. -263. -269. -275.]
 [ -18. -7. 7. 21. 15. ... -237. -243. -249. -255. -261.]
 [ -24. -13. 1. 15. 26. ... -226. -232. -238. -244. -250.]
 [-276. -265. -251. -237. -226. ... 216. 210. 204. 198. 192.]
 [-282. -271. -257. -243. -232. ... 210. 224. 218. 212. 206.]
 [-288. -277. -263. -249. -238. ... 204. 218. 228. 222. 216.]
 [-294. -283. -269. -255. -244. ... 198. 212. 222. 236. 230.]
 [-300. -289. -275. -261. -250. ... 192. 206. 216. 230. 240.]]
ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC
```

```
COST MATRIX:

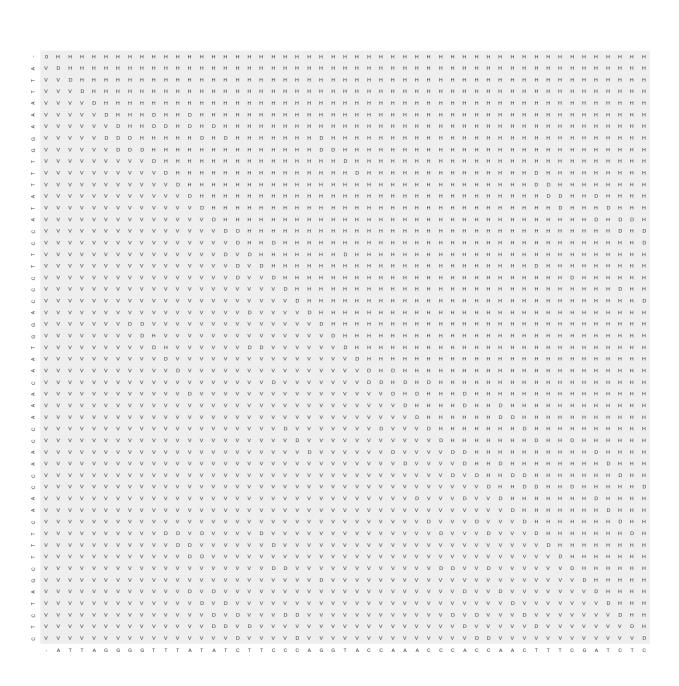
- A T C G
- -6 -6 -6 -6 -6
A -6 +5 -1 -4 -4
T -6 -1 +8 -3 -4
C -6 -4 -3 +4 -4
G -6 -4 -4 -4 +6
```

SCORE: 240

Representation of all values of the score matrix for Needleman-Wunsch alignment algorithm;



Representation of all values of the direction matrix for Needleman-Wunsch alignment algorithm; (D: Diagonal inheritence, V: Vertical inheritence, H: Horizontal inheritence)

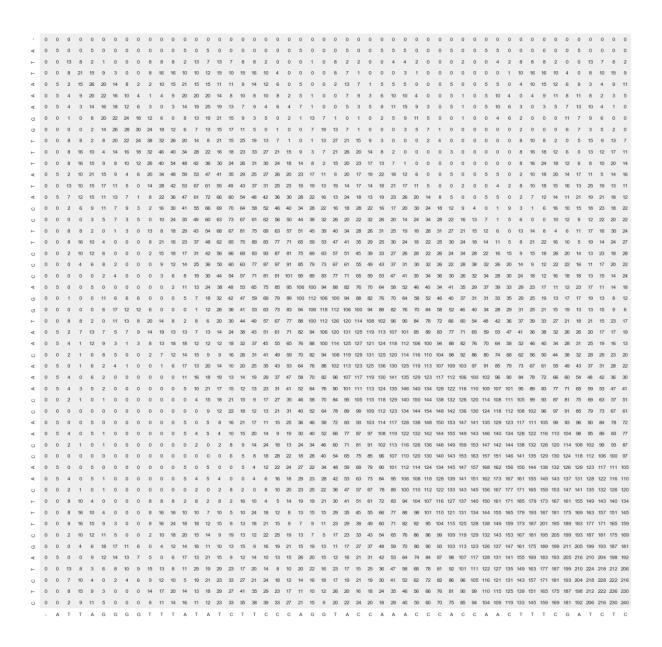


Local Alignment

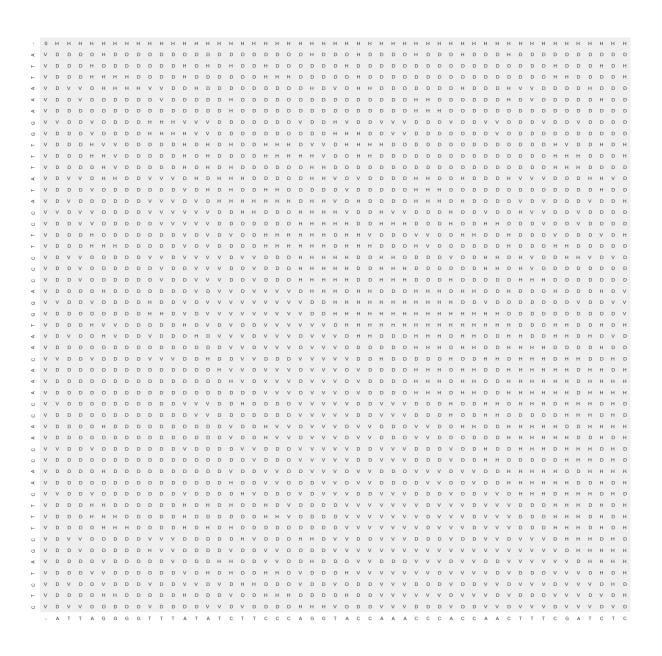
Smitch-Waterman algorithm is implemented for finding local alignments score. In the figure below, there are details of the highest scoring pairings found at the end of the Smitch-Waterman alignment, which operates 104 times using 4 different cost matrices and 26 different sequence pairs.

```
MAXIMUM SCORE WITH PROVIDING COST MATRIX
DIRECTION MATRIX:
'V' 'D' 'D' 'D' 'D' ... 'D' 'D' 'D' 'D' [
 ['V' 'D' 'D' 'D' 'H' ... 'D' 'D' 'H' 'D' 'H']
 ['v' 'ס' 'ס' 'ס' 'א' ... 'א' 'ס' 'ס' 'ס' 'ס' 'ס']
['V' 'D' 'V' 'V' 'D' ... 'D' 'H' 'D' 'D' 'D']
 ['V' 'D' 'D' 'D' 'D' ... 'D' 'H' 'H' 'H' 'H']
 ['V' 'D' 'D' 'D' 'V' ... 'V' 'D' 'H'
                              'D'
 ['V' 'D' 'V' 'D' 'D' ... 'V' 'V' 'D'
                              'н'
 ['V' 'D' 'D' 'D' 'D' ... 'V' 'D' 'V' 'D' 'H']
['V' 'D' 'V' 'V' 'D' ... 'V' 'V' 'D' 'V' 'D']]
SCORING MATRIX:
                                          0.]
              0. 0. ... 0. 0.
                                 0.
[[ 0. 0. 0.
                                     0.
  0. 5.
         0. 0. 5. ... 5. 0. 0. 0. 0.
 [ 0. 0. 13. 8. 2.... 0. 13. 7. 8.
                                         2.]
 0. 0. 8. 21. 15. ... 0. 8. 10. 15.
                                         9.]
0. 5. 2. 15. 26. ... 9. 3. 4.
[ 0. 5. 0. 0. 9. ... 216. 210. 204. 198. 192.]
  0. 0. 13. 8. 3. ... 210. 224. 218. 212. 206.
 [ 0. 0. 7. 10. 4. ... 204. 218. 228. 222. 216.]
  0. 0. 8. 15. 9. ... 198. 212. 222. 236. 230.]
[ 0. 0. 2. 9. 11. ... 192. 206. 216. 230. 240.]]
ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC
SCORE: 240
COST MATRIX:
 - A T C G
- -6 -6 -6 -6
A -6 +5 -1 -4 -4
T -6 -1 +8 -3 -4
C -6 -4 -3 +4 -4
G -6 -4 -4 -4 +6
```

Representation of all values of the score matrix for Smitch-Waterman alignment algorithm;



Representation of all values of the direction matrix for Needleman-Wunsch alignment algorithm; (D: Diagonal inheritence, V: Vertical inheritence, H: Horizontal inheritence)



Personal Conclusion Comment:

Since the given dna sequences are very close to each other, that is, there is not much protein difference between them and the gap penalty value (-6) we want to use is smaller than the mismatch penalty in the given cost matrices, the algorithm that is applied prefers to do mismatch instead of gap. In local alignment, this method often gives more accurate results, but in global alignment, accurate results cannot be obtained.

Especially if we look at global alignment and local alignment maximum score matches, almost the same results are obtained due to the errors mentioned above.

In my opinion, either to give the gap penalty value close to 0, or to encourage it to replace the mismatch instead of mismatch by making the gap penalty value less than the gap penalty value, especially for the global alignment to work correctly.

Another arrangement which is not very important but needs to be made is to make the similarities (or lengths) of the given dna sequences as different as possible.