

Ödevde verilen algoritmaların implementasyon'u gerçekleştirirken ilk önce basit ve daha kısa olan kendi başıma çözebildiğim DNA dizileri(sequence) tercih ettim böylelikle algoritmanın doğru çalıştığından emin olup, daha sonra sizin tarafınızda tanımlanmış olan Dna dizileri(sequence) algoritmamı uyguladım.

Power is nothing without control!



Aşağıdaki çıktılarda ilgi socre matrixlarni ve bu score matrixlerinin sonuçlarını görebilirsiniz. Buna ek olarak gerçekleştirmiş olduğum basit örnekleri görebilirsiniz.

Buna ek olarak algoritmalar hakkında yaptığım çıkarımlara ve yorumlarımı da aşağıda belirmiş bulunmaktayım.

Needleman-Wunsch algorithm

İmplement ettiğim algoritma aşağıdaki gibidir.

Initialization.

$$\begin{aligned} F(0, 0) &= 0 \\ F(0, j) &= -j \times d \\ F(i, 0) &= -i \times d \end{aligned}$$

Main Iteration. Filling-in partial alignments

For each $i = 1, \dots, M$

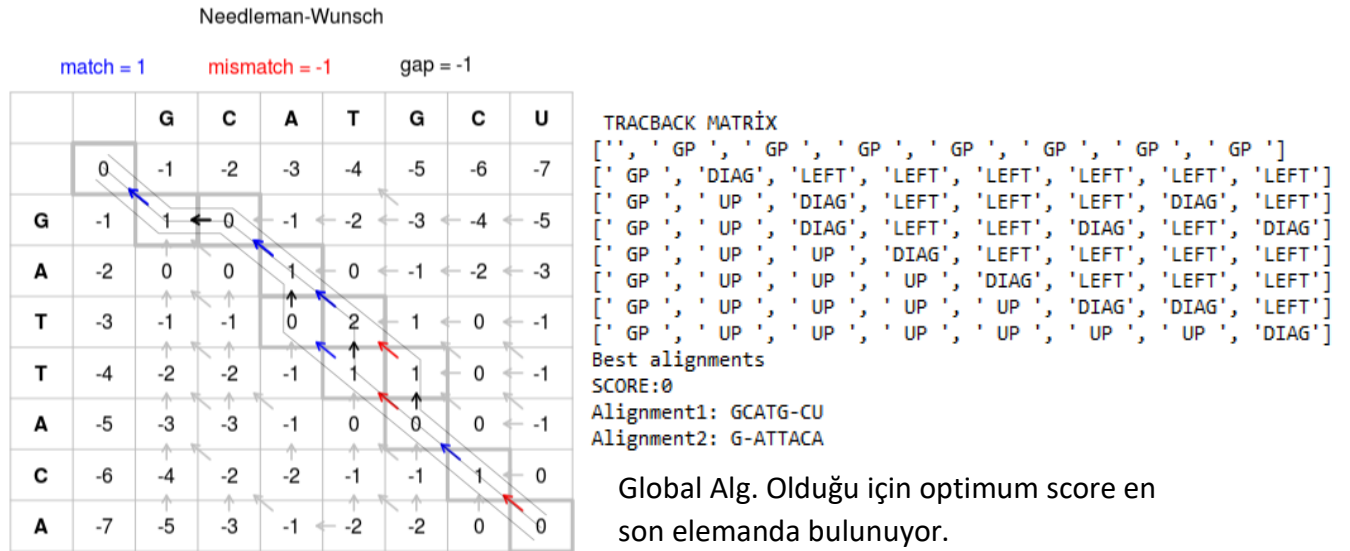
For each $j = 1, \dots, N$

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) & [\text{case 1}] \\ F(i-1, j) - d & [\text{case 2}] \\ F(i, j-1) - d & [\text{case 3}] \end{cases}$$

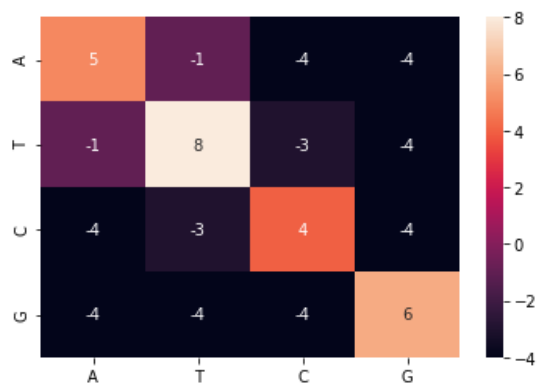
$$\text{Ptr}(i, j) = \begin{cases} \text{DIAG}, & \text{if } [\text{case 1}] \\ \text{LEFT}, & \text{if } [\text{case 2}] \\ \text{UP}, & \text{if } [\text{case 3}] \end{cases}$$

3. **Termination.** $F(M, N)$ is the optimal score, and from $\text{Ptr}(M, N)$ can trace back optimal alignment

Basit örnek:



Tanımlanmış olan Score matrisi ve Dna Sequence için elde edilen sonuçlar :



Score:240

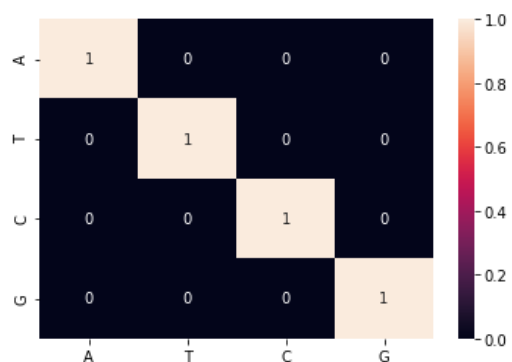
Alignment1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

Alignment2:ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC

ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

|||| |

ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC



Score:45

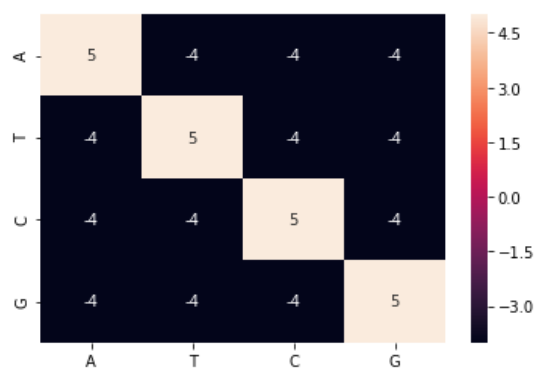
Alignment1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

Alignment2:ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC

ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

|||| |

ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC



Score:205

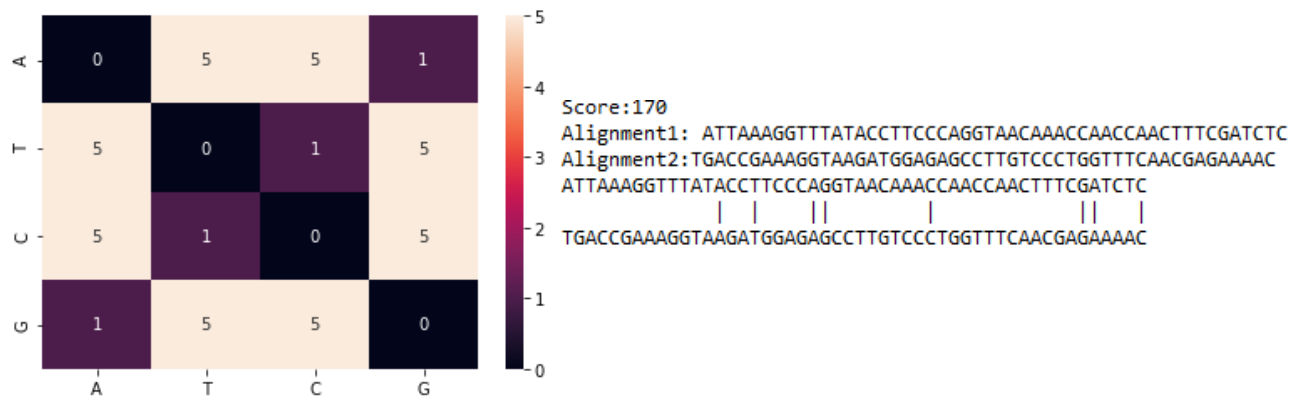
Alignment1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

Alignment2:ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC

ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

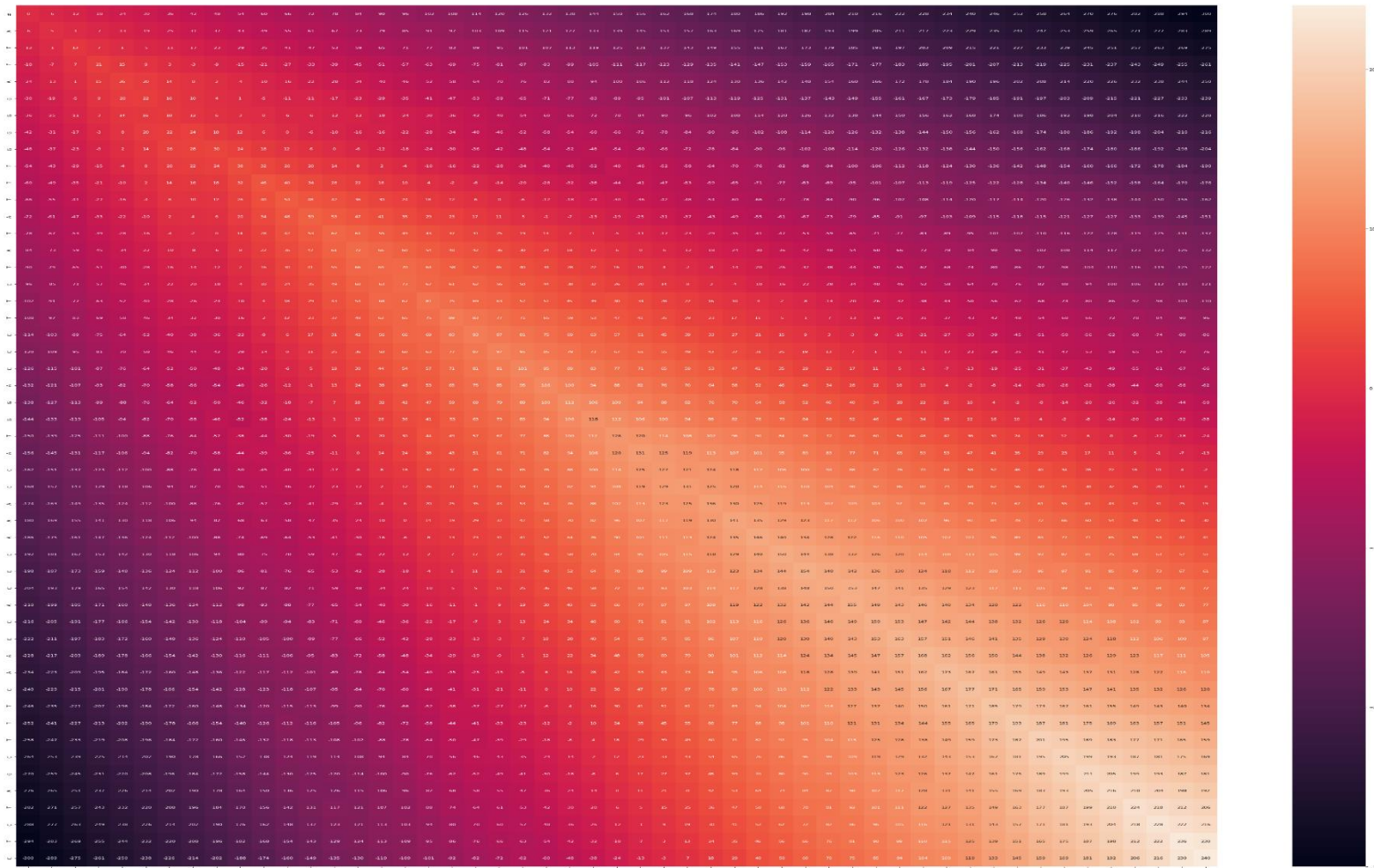
|||| |

ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC



Burada mismatchlerin daha yüksek deęerler sahip olmasından dolayı böyle kötü bir alignment elde ettik. Algoritmamız üzerinde deęiřikliğe gidersek (max yerine min) deęerleri alırsak daha iyi bir alignment elde edebiliriz fakat bu seferde gap(-6) deęerinde deęiřikliğe gitmemiz gerekmektedir.

Needleman Scora Matrix Sample



Smith-Waterman Algorithm

İmplement ettiğim algoritma aşağıdaki gibidir.

Smith & Waterman Algorithm

- The recurrence:

$$s_{i,j} = \max \begin{cases} 0 \\ s_{i-1,j-1} + \delta(v_p, w_j) \\ s_{i-1,j} + \delta(v_p, -) \\ s_{i,j-1} + \delta(-, w_j) \end{cases}$$

There is only this change from the original recurrence of a Global Alignment

initialization: first row and first column initialized with 0's

- traceback:

—find maximum value of $F(i, j)$; can be anywhere in matrix
—stop when we get to a cell with value 0

Basit örnek:

| | | T | G | T | T | A | C | G | G |
|---|---|---|---|---|---|----|----|----|---|
| G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 3 | 3 |
| T | 0 | 3 | 1 | 6 | 4 | 2 | 0 | 1 | 4 |
| T | 0 | 3 | 1 | 4 | 9 | 7 | 5 | 3 | 2 |
| G | 0 | 1 | 6 | 4 | 7 | 6 | 4 | 8 | 6 |
| A | 0 | 0 | 4 | 3 | 5 | 10 | 8 | 6 | 5 |
| C | 0 | 0 | 2 | 1 | 3 | 8 | 13 | 11 | 9 |
| T | 0 | 3 | 1 | 5 | 4 | 6 | 11 | 10 | 8 |
| A | 0 | 1 | 0 | 3 | 2 | 7 | 9 | 8 | 7 |

Traceback path (blue arrows) from (C, 13) to (G, 3):

Sequence 1: TGTTACGG
Aligned sequence: GTT-AC

[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 0, 3, 3, 1, 0, 0, 3, 1]
[0, 3, 3, 1, 1, 6, 4, 2, 1, 0]
[0, 1, 1, 6, 4, 4, 3, 1, 5, 3]
[0, 0, 0, 4, 9, 7, 5, 3, 4, 2]
[0, 0, 0, 2, 7, 6, 10, 8, 6, 7]
[0, 0, 0, 0, 5, 4, 8, 13, 11, 9]
[0, 3, 3, 1, 3, 8, 6, 11, 10, 8]
[0, 3, 6, 4, 2, 6, 5, 9, 8, 7]

SCORE:13

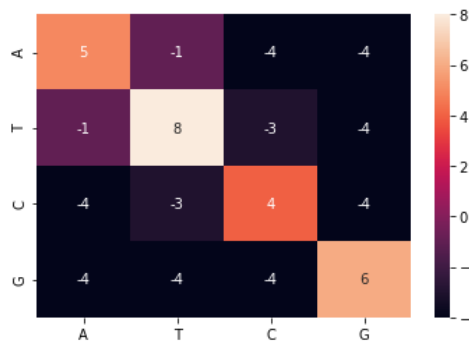
Sequence 1: TGTTACGG

Aligned sequence: GTT-AC

Sequence 2: GGTTGACTA

https://en.wikipedia.org/wiki/Smith%E2%80%93Waterman_algorithm

Tanımlanmış olan Score matrisi ve Dna Sequence için elde edilen sonuçlar :



Score:240

Alignment1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCACTTTCGATCTC

Alignment27: ATTAAGGTTTATATCTTCCCAGGTACCAAAACCAACCACTTTCGATCTC

||||| ||||||| ||||||||| ||||||| ||||||| |||||||

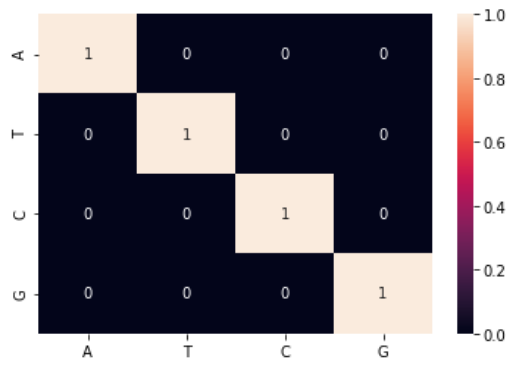
ATTAAGGTTTATATCTTCCCAGGTACCAAAACCAACCACTTTCGATCTC

Bu iki sequence en iyi eşleştirmeyi gerçekleştirdi ve

global alignment gibi çalıştı fakat uygulamayı

çalıştırdığınızda çıktılarında algoritmanın local

alignmentsı başarı ile gerçekleştirdiğinizi göreceksiniz.



Score:45

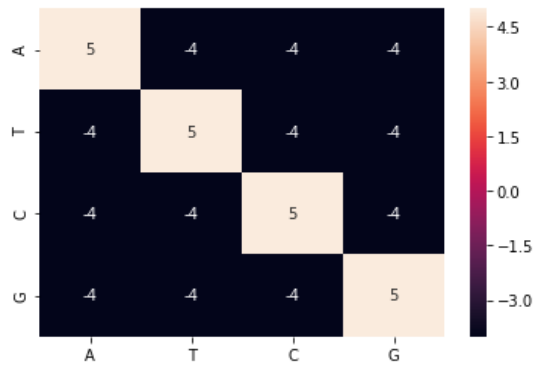
Sequence1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTC

Sequence2: ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC

ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTC

|||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC



Score:205

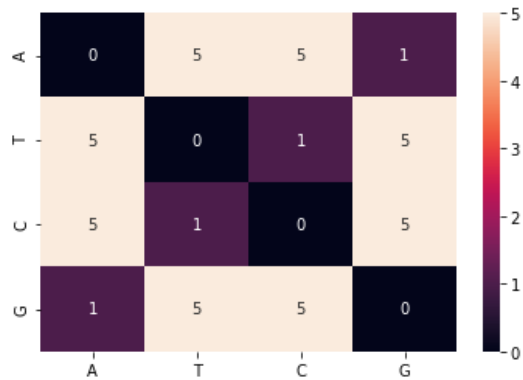
Sequence1: ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTC

Sequence2: ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC

ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTC

|||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

ATTAGGGGTTTATATCTTCCCAGGTACCAAACCCACCAACTTTCGATCTC



Score:176

Sequence1: ATTAAGG-TTTATACC-TTCCCAGGTAACAA-AC-CAACCAACTTTCGA

Sequence2: TACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTC

ATTAAGG-TTTATACC-TTCCCAGGTAACAA-AC-CAACCAACTTTCGA

TACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTC

Yine daha önce açıkladığım gibi bu scora matrix için algoritmada değişikliklere gidilmelidir.

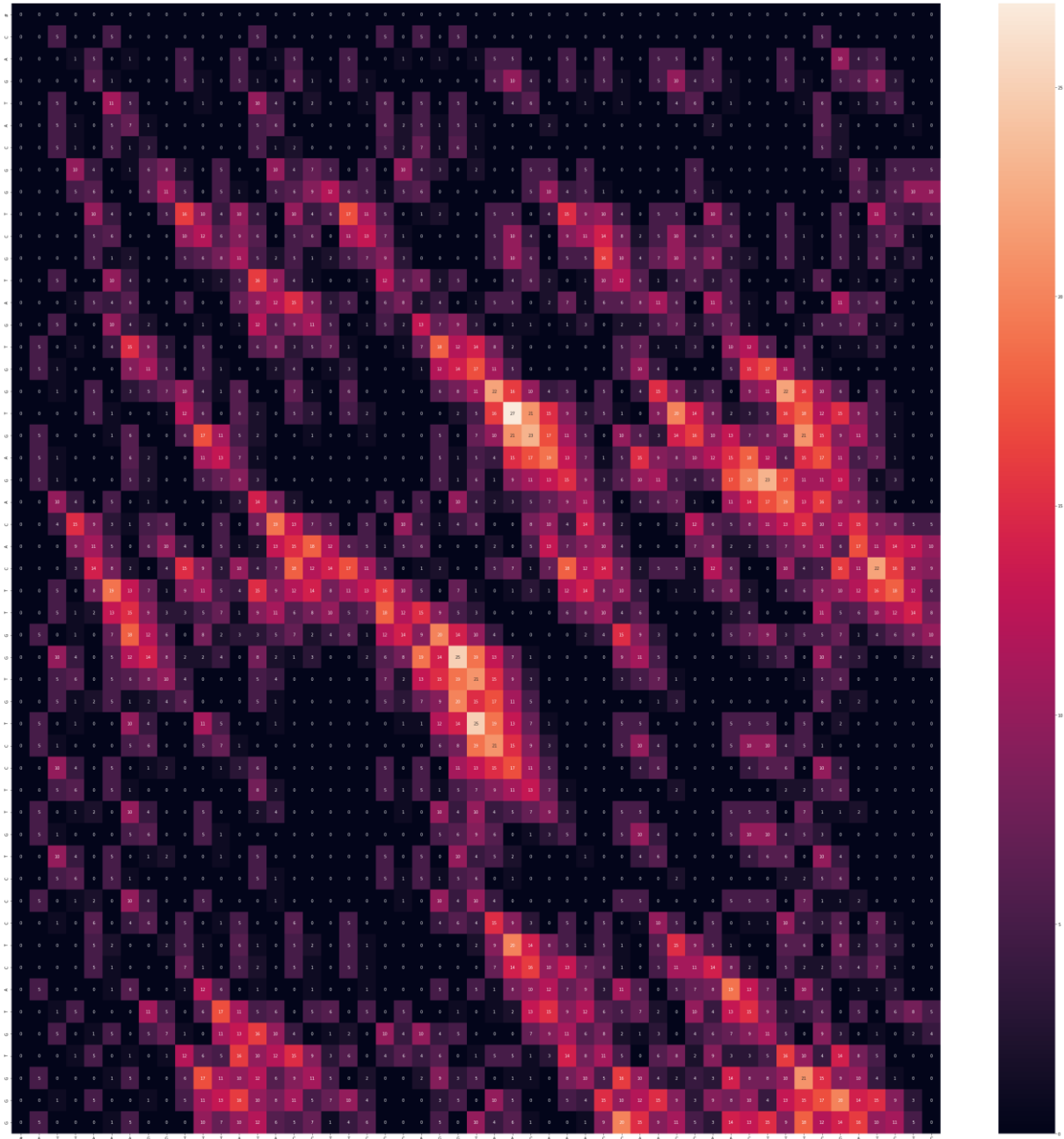
Local alignment'ın gerçekleştiğini gösterebilmek için en iyi örneklerden biri ,
Burada görüldüğü üzere diğer örnekler gibi optimum score matrix'in sonunda
bulunmuyor ve back tracking yapılırken de bu fark görülüyor.

SCORE:27

Sequence 1: ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTTCGATCTC

Aligned sequence: TAAAGGTT-ATAC-CTT

Sequence 2: CAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCTCATGTGGG

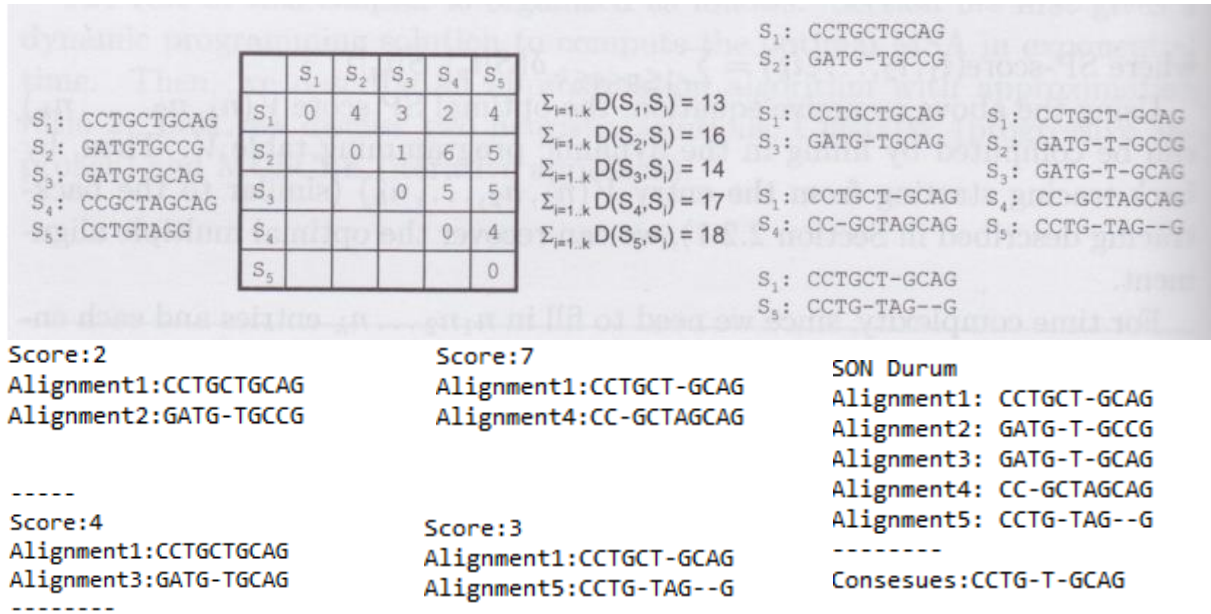


Star Aligment Algorithm

Star Aligment algoritmasının arka tarafında dizilere global alignment uygulandığı için daha önceden gerçekleştirmiş olduğumu Needleman W. Algoritmasını obje olarak yaratarak kullanmayı tercih ettim.

Basit örnek:

Gerçekleştirilen Uygulama: Star Aligment/sample(StarAligment).py



<https://www.site.uottawa.ca/~lucia/courses/5126-11/lecturenotes/12-13MultipleAlignment.pdf>

Tanımlanmış olan Dna Sequence için elde edilen sonuçlar :

! Bu yöntemde herhangi bir score matrisi tanımlanmadığı için kendim tarafında belirlenen score matrisini kullandım.

#gap=-1 #match=1 #mismatch=-1 (Kendi değerlerinizi parametre olarak vererek test edebilirsiniz)

Score:2
Alignment1:ATTAAAGGT-T-TATACCTTCCCAGGTA--ACAAACCAACCAA-CTTTCGA-TC-TC
Alignment2:-TTGTAGATCTGT-TA--TTAA-AGGTTTAA-AAATCTGTGTGGCTGTC-ACTCCTC

Score:0

Alignment1:ATTAAAGGT-T-TATACCTTCCCAGGTA--ACAAACC--AACCAA-CT-T-TCGA--TC--TC

Score:-7

Alignment1:AT-TAAAGGT-T-TA-TAC-CTTC-C-CAGGTA--ACA--AACC--AACCAA-CT-T-TCGA--TC--TC

Alignment4:GGCTGCATGC-T-TAGTGCACT-CACGCAG-TAT-A-ATTAAT---AACTAA--T-TA-C----TG--T-

SON Durum

Sequence.1 : AT-TAAAGGT-T-TA-TAC-CTTC-C-CAGGTA--ACA--AACC--AACCAA-CT-T-TCGA--TC--TC #Center

Sequence.2: -T-TGTAGATCTGT--TA---TTAA---AGGTTTTA-A--AATCTGTGTGG--CTGT--C-AC-TCC-TC

Score:-11

Sequence.3: -T-TGTAGATCTGT--T-C--T---CT--A---A--ACG--AACTTTAA--AATCTGTGTGGCTGTCACTC

Score:-8

Sequence.4: GGCTGCATGC-T-TAGTGCACT-CACGCAG-TAT-A-ATTAAT---AACTAA--T-TA--C----TG--T-

Score:-8

Consensus: -T-TGTAGGT-T-TA-TAC-CTTC-C-CAGGTAT-A-A--AATC--AACCAA-CT-T--CGA--TCC-TC