Exercise sheet 5: Local Sequence Alignment

Exercise 1 - Smith-Waterman

Consider the following sequences $S1$ and $S2$, a similarity scoring via $s(x, y)$ and a gap cost function $g(k)$.

$$S1 = TCCGA \tag{1}$$
$$S2 = TACGCAGA \tag{2}$$
$$s(x, y) = \begin{cases} +1 & if \ x = y \\ 0 & else \end{cases} \tag{3}$$
$$g(k) = -k \tag{4}$$

Question 1A   Compute the local alignment matrix $S_{i,j}$ for the given sequence.

Matrix

Hint: Formulae

$$H_{0,0} = 0 \tag{5}$$
$$H_{i,0} = 0 \tag{6}$$
$$H_{0,j} = 0 \tag{7}$$
$$H_{i,j} = max \begin{cases} H_{i-1,j-1} + s(a_i, b_j) \\ H_{i-1,j} + s(a_i, -) \\ H_{i,j-1} + s(-, b_j) \\ 0 \end{cases} \tag{8}$$

Solution

Question 1B   Give all optimal local alignments and the according score.

Solution   score: 3

```
TCCG     TCCGA    TCCG-A    CCGA
|:||     |:||:    |:|| |    |:||
TACG     TACGC    TACGCA    CAGA
```

Exercise 2 - Affine gap costs

Question 2A Which of these statements are correct?

☐ Distance and similarity scores are equally useful for local alignment scoring
☐ Similarity scores are not suited for local alignment scoring
☐ Distance scores are not suited for local alignment scoring

Solution

☐ Distance and similarity scores are equally useful for local alignment scoring
☐ Similarity scores are not suited for local alignment scoring
☒ Distance scores are not suited for local alignment scoring

Question 2B You want to extend the Smith-Waterman algorithm for local alignment to more general gap scoring functions $g(k)$ (where k denotes the gap length).

The following recursions were created analogously to the Waterman-Smith-Beyer algorithm. Which of these (if any) represents a variant of the Smith-Waterman algorithm that allows for an arbitrary gap scoring function?

$$D_{0,0} = 0, \ D_{i,0} = 0, \ D_{0,j} = 0$$

$$D_{i,j} \stackrel{(i)}{=} max \begin{cases} Di-1, j-1 + s(a_i, b_j) \\ max_{1 \le k \le j} D_{i,j-k} + g(1) \\ max_{1 \le k \le i} D_{i-k,j} + g(1) \\ 0 \end{cases} \qquad D_{i,j} \stackrel{(ii)}{=} max \begin{cases} Di-1, j-1 + s(a_i, b_j) \\ max_{1 \le k \le j} D_{i,j-k} + g(k) \\ max_{1 \le k \le i} D_{i-k,j} + g(k) \\ 0 \end{cases} \tag{9}$$

$$\tag{10}$$

$$D_{i,j} \stackrel{(iii)}{=} min \begin{cases} Di-1, j-1 + s(a_i, b_j) \\ max_{1 \le k \le j} D_{i,j-k} + g(k) \\ max_{1 \le k \le i} D_{i-k,j} + g(k) \\ 0 \end{cases} \qquad D_{i,j} \stackrel{(iv)}{=} min \begin{cases} Di-1, j-1 + s(a_i, b_j) \\ max_{1 \le k \le j} D_{i,j-k} + g(1) \\ max_{1 \le k \le i} D_{i-k,j} + g(1) \\ 0 \end{cases} \tag{11}$$

$$\tag{12}$$

$$D_{i,j} \stackrel{(v)}{=} max \begin{cases} Di-1, j-1 + s(a_i, b_j) \\ max_{1 \le k \le j} D_{i,j-k} + g(1) \\ max_{1 \le k \le i} D_{i-k,j} + g(1) \\ -1 \end{cases} \tag{13}$$

Solution Only Formula *ii* allows for an arbitrary gap scoring.

Note

Distance scores are not suited for local alignment scoring, thus all minimization approaches are not suitable. Further $g(1)$ does not represent an affine scoring.

Question 2C The following recursions were created analogously to the Gotoh algorithm. Which of these (if any) represents a variant of the Smith-Waterman algorithm that allows for an affine gap scoring function?

$$D_{0,0} = 0, \ D_{i,0} = 0, \ D_{0,j} = 0$$

$$D_{i,j} \stackrel{(i)}{=} max \begin{cases} D_{i-1,j-1} + s(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \\ 0 \end{cases} \qquad P_{i,j} = max \begin{cases} D_{i-1,j} + g(1) \\ P_{i-1,j} + \beta \end{cases}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad Q_{i,j} = max \begin{cases} D_{i,j-1} + g(1) \\ Q_{i,j-1} + \beta \end{cases}$$
(14)

(15)

$$D_{i,j} \stackrel{(ii)}{=} max \begin{cases} D_{i-1,j-1} + s(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \\ 0 \end{cases} \qquad P_{i,j} = max \begin{cases} D_{i-1,j} + g(k) \\ P_{i-1,j} + \beta \end{cases}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad Q_{i,j} = max \begin{cases} D_{i,j-1} + g(k) \\ Q_{i,j-1} + \beta \end{cases}$$
(16)

(17)

$$D_{i,j} \stackrel{(iii)}{=} min \begin{cases} D_{i-1,j-1} + s(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \\ 0 \end{cases} \qquad P_{i,j} = min \begin{cases} D_{i-1,j} + g(k) \\ P_{i-1,j} + \beta \end{cases}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad Q_{i,j} = min \begin{cases} D_{i,j-1} + g(k) \\ Q_{i,j-1} + \beta \end{cases}$$
(18)

(19)

$$D_{i,j} \stackrel{(iv)}{=} min \begin{cases} D_{i-1,j-1} + s(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \\ 0 \end{cases} \qquad P_{i,j} = min \begin{cases} D_{i-1,j} + g(1) \\ P_{i-1,j} + \beta \end{cases}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad Q_{i,j} = min \begin{cases} D_{i,j-1} + g(1) \\ Q_{i,j-1} + \beta \end{cases}$$
(20)

Solution Only Formula *i* allows for an affine gap scoring.

Note

Distance scores are not suited for local alignment scoring, thus all minimization approaches are not suitable. Further $g(1)$ does not represent an affine scoring.

Exercise 3 - Programming assignment

Programming assignments are available via Github Classroom and contain automatic tests.

We recommend doing these assignments since they will help you to further understand this topic.

Access the Github Classroom link: Programming Assignment: Sheet 05.