

14. Übung

Einführung in die Bioinformatik I, 2. Teil
Sommersemester 2021

Aufgabe 1 (3 Punkte): Gegeben seien die Sequenzen $u = \text{GATCAGGT}$ und $v = \text{GACC}$. Bestimmen Sie die Matrix D und alle optimalen **globalen** Alignments mit Einheitskosten.

	ϵ	G	A	T	C	A	G	G	T
ϵ	0	1	2	3	4	5	6	7	8
G	1	0	1	2	3	4	5	6	7
A	2	1	0	1	2	3	4	5	6
C	3	2	1	1	1	2	3	4	5
C	4	3	2	2	1	2	3	4	5

GATCAGGT
GACC-----

GATCAGGT
GA-CC----

GATCAGGT
GA-C-C---

GATCAGGT
GA-C--C-

GATCAGGT
GA-C---C

Aufgabe 2 (6 Punkte): Gegeben seien die beiden Sequenzen aus Aufgabe 1.

Wir definieren die Ähnlichkeitsfunktion S mit $S(a, a) = 1, S(a, b) = -1$ für $a \neq b$ und $S(a, -) = S(-, b) = -1$. Bestimmen Sie die Matrix H für ein **lokales** Alignment von u, v . Wie lauten die optimalen Alignments (Traceback in die Matrix einzeichnen), wie ihre Ähnlichkeit?

	ϵ	G	A	T	C	A	G	G	T
ϵ	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	1	1	0
A	0	0	2	1	0	1	0	0	0
C	0	0	1	1	2	1	0	0	0
C	0	0	0	0	2	1	0	0	0

GATC
GACC

GATC
GA-C

GA
GA

Aufgabe 3 (6 Punkte): Gegeben seien wiederum die Sequenzen aus Aufgabe 1. Wir definieren die Ähnlichkeitsfunktion $S' : \{A, C, G, T\} \times \{A, C, G, T\} \rightarrow \mathbb{R}$ mit

- $S'(A, A) = 2, S'(C, C) = S'(G, G) = S'(T, T) = 1,$
- $S'(C, G) = S'(G, C) = 1, S'(a, b) = -1$ für alle anderen $a \neq b,$ und
- $S'(a, -) = S'(-, b) = -1.$

Bestimmen Sie die Matrix H' für ein **lokales** Alignment von $u, v.$ Was sind diesmal die optimalen Alignments (Traceback in die Matrix einzeichnen) und deren Ähnlichkeit?

	ϵ	G	A	T	C	A	G	G	T
ϵ	0	0	0	0	0	0	0	0	0
G	0	1	0	0	1	0	1	1	0
A	0	0	3	2	1	3	2	1	0
C	0	1	2	2	3	2	4	3	2
C	0	1	1	1	3	2	3	5	4

CAGG
GACC

Aufgabe 4 (5 Punkte): Gegeben sei die Sequenz $u = \text{CCAT}$ sowie das Profil:

Pos.	1	2	3	4
A	0,2	0	0,7	0
C	0,6	0,4	0,2	0
G	0	0,5	0	0
T	0,1	0	0,1	1
-	0,1	0,1	0	0

Bestimmen Sie die Ähnlichkeit des Profil-Sequenz-Alignments für die Ähnlichkeitsfunktion S' aus Aufgabe 3, wenn keine zusätzlichen Gaps in u eingefügt werden.

Aufgabe 4 (5 Punkte): Gegeben sei die Sequenz $u = \text{CCAT}$ sowie das Profil:

Pos.	1	2	3	4
A	0,2	0	0,7	0
C	0,6	0,4	0,2	0
G	0	0,5	0	0
T	0,1	0	0,1	1
-	0,1	0,1	0	0

Bestimmen Sie die Ähnlichkeit des Profil-Sequenz-Alignments für die Ähnlichkeitsfunktion S' aus Aufgabe 3, wenn keine zusätzlichen Gaps in u eingefügt werden.

$$C: 0,2 * S'(A, C) + 0,6 * S'(C, C) + 0,1 * S'(T, C) + 0,1 * S'(-, C) = -0,2 + 0,6 - 0,1 - 0,1 = 0,2$$

$$C: 0,4 * S'(C, C) + 0,5 * S'(G, C) + 0,1 * S'(-, C) = 0,4 + 0,5 - 0,1 = 0,8$$

$$A: 0,7 * S'(A, A) + 0,2 * S'(C, A) + 0,1 * S'(T, A) = 1,4 - 0,2 - 0,1 = 1,1$$

$$T: 1,0 * S'(T, T) = 1,0$$

$$\text{Gesamt: } 0,2 + 0,8 + 1,1 + 1,0 = 3,1$$