

2. Übung zur Vorlesung “Einführung in die Bioinformatik I, 2. Teil”

Sommersemester 2020

Prof. Peter Dittrich, Marcus Ludwig

Ausgabe: 12. Mai 2020,
Abgabe: 19. Mai 2020 per Email

Aufgabe 1 (3 Punkte): Gegeben die Ähnlichkeitsfunktion S mit $S(a, a) = 1$, $S(a, b) = -1$ für $a \neq b$ und $S(a, -) = S(-, b) = -1$. Berechnen Sie den Sum-of-Pairs-Score für eine Spalte mit k Einträgen, in der sich nur 'A's befinden und für eine Spalte mit k Einträgen, in der sich $k - 1$ 'A's und ein 'T' befinden. Vergleichen Sie die beiden Werte asymptotisch.

Aufgabe 2 (4 Punkte): Bestimmen Sie die besten Consensus-Sequenzen des multiplen Alignments

```
G T A A C A T C C A
A T G - C C - - G A
A T A - C C G G C T
A T G C C G - G A T
A T G C T C - G A T
```

1. für obige Ähnlichkeitsfunktion S mit $S(-, -) = 0$.
2. für die Ähnlichkeitsfunktion S' mit
 - $S'(A, A) = 4$, $S'(C, C) = S'(G, G) = S'(T, T) = 2$,
 - $S'(C, G) = S'(G, C) = 1$, $S'(a, b) = -1$ für alle anderen $a \neq b$, und
 - $S'(a, -) = S'(-, b) = -1$, $S'(-, -) = 0$.

Aufgabe 3 (8 Punkte): Berechnen Sie die Matrix D für das lokale multiple Alignment der drei Sequenzen $u = \text{ACAG}$, $v = \text{GGA}$ und $w = \text{CAGG}$ für den Sum-of-Pairs-Score zur Ähnlichkeitsfunktion S'' mit $S''(a, a) = 3$, $S''(a, b) = -1$ für $a \neq b$, $S''(a, -) = S''(-, b) = -1$ und $S''(-, -) = 0$. Schreiben Sie die zweidimensionalen Matrizen $D[i, j, 0]$, $D[i, j, 1]$, \dots , $D[i, j, 4]$ einzeln auf. Was sind die optimalen Alignments, was ihre Ähnlichkeit?

Aufgabe 4 (5 Punkte): Erstellen Sie den (vollständig beschrifteten) Edit-Graphen für die Sequenzen $u = \text{GTTC}$ und $v = \text{GGAT}$

1. mit Einheitskosten
2. mit der Kostenfunktion

$$\delta(a, b) = \begin{cases} 0 & a = b \\ 1 & a, b \in \{\text{A}, \text{G}\} \text{ und } a \neq b \\ & a, b \in \{\text{C}, \text{T}\} \text{ und } a \neq b \\ 2 & \textit{sonst} \end{cases}$$