

12. Übung zur Vorlesung “Einführung in die Bioinformatik I, 1. Teil”

Wintersemester 2019/2020

Prof. Peter Dittrich, Emanuel Barth, Marcus Ludwig

Ausgabe: 22. Januar 2020,
Abgabe: 29. Januar 2020 zu Beginn der Übung

Aufgabe 1 (6 Punkte): Berechnen Sie mit Hilfe einer DP-Matrix alle optimalen globalen Alignments (mit Einheitskosten) von GTGCACA und TCACTTA. Wie viele sind es?

Aufgabe 2 (6 Punkte): Die vier organischen Basen Adenin, Cytosin, Guanin und Thymin der DNA lassen sich in sogenannte *Purine* (Adenin und Guanin) und *Pyrimidine* (Cytosin und Thymin) unterteilen. Bei Mutationen im genetischen Code ist es wahrscheinlicher, dass eine Purinbase durch eine andere oder eine Pyrimidinbase durch eine andere substituiert wird (*Transition*) als dass eine Purinbase durch eine Pyrimidinbase ersetzt wird oder umgekehrt (*Transversion*).

Diese Tatsache berücksichtigen wir in der folgenden Kostenfunktion δ :

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

Berechnen Sie die DP-Matrix von GTGCACA und TCACTTA unter Verwendung dieser Kostenfunktion, und geben Sie wie oben alle optimalen Alignments an.

Hinweis: Wenn keine Einheitskosten verwendet werden, gilt für die DP-Matrix:

$$D[i, 0] = \sum_{k=1}^i \delta(u_k, -) \quad \text{und} \quad D[0, j] = \sum_{k=1}^j \delta(-, v_k)$$

Aufgabe 3 (2 Punkte): Im Verlaufe der Evolution haben sich Mensch und Maus aus einem gemeinsamen Vorfahren entwickelt. Daher wäre es praktischer, Gene von Mensch und Maus nicht miteinander, sondern mit denen des gemeinsamen Vorfahren zu alignieren. Warum vergleichen wir nicht gegen den Vorfahren? Warum ergibt es trotzdem Sinn, Gene von Mensch und Maus miteinander zu alignieren?

Bonus (6 Punkte): Berechnen Sie mit der folgenden Rekurrenz die Anzahl aller globalen Alignments zweier Strings u, v für alle Längen von u und v bis einschließlich vier:

$$N_{\mathcal{A}}[i, 0] = N_{\mathcal{A}}[0, j] = 1 \quad N_{\mathcal{A}}[i, j] = N_{\mathcal{A}}[i - 1, j] + N_{\mathcal{A}}[i, j - 1] + N_{\mathcal{A}}[i - 1, j - 1]$$

Wieso funktioniert diese Rekurrenz?