

8. Übung zur Vorlesung "Sequenzanalyse"

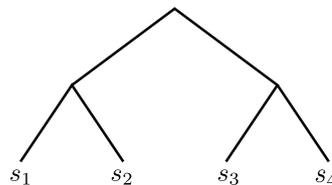
Sebastian Böcker, Markus Fleischauer

Aufgabe 1 (5 Punkte)

Gegeben seien vier Sequenzen $s_1 = TACA$, $s_2 = CTAC$, $s_3 = GTAG$, $s_4 = ATGC$. Berechnen Sie das multiple Sequenzalignment dieser vier Sequenzen anhand der Center-Star-Methode mit Einheitskosten.

Aufgabe 2 (5 Punkte)

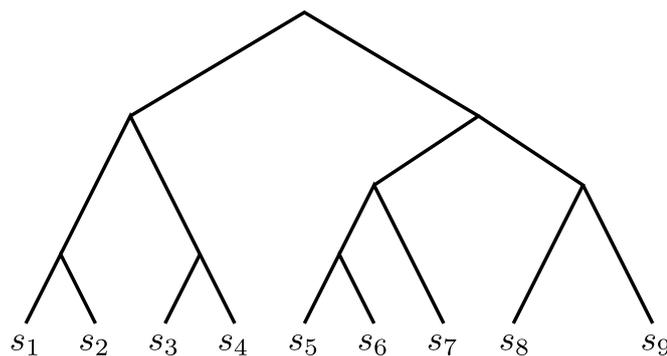
Berechnen Sie das multiple Sequenzalignment der Sequenzen $s_1 = AGTCAT$, $s_2 = GTACT$, $s_3 = ATTATC$, und $s_4 = GGCCT$ mit Einheitskosten und anhand des unten gegebenen Leitbaums. Alignieren Sie dabei das Alignment (s_1, s_2) mit (s_3, s_4) an der Wurzel des Baumes.



Erklären Sie, warum das Problem schwieriger wird, wenn anstatt Einheitskosten die affinen Gapkosten verwendet werden.

Aufgabe 3 (5 Punkte)

Gewichten Sie die Sequenzen s_1 bis s_9 anhand des unten stehenden Leitbaums. Die Kantengewichte seien dabei 1.



Aufgabe 4 (10 Punkte) Gegeben die 4 Sequenzen *TATAGGCAGGT*, *TATACAGG*, *ATAGGCAGGT* und *TAGGTAG*. Erstellen Sie die erweiterte Alignment-Sammlung für Sequenzen 3 und 4. Verwenden Sie anschließend diese um die beiden Sequenzen zu alignieren.

- a) Erstellen Sie die erweiterte Alignment-Sammlung für Sequenzen 3 und 4, indem Sie die Sequenzen 3 und 4 end-gap-free paarweise **und** über jede mögliche dritte Sequenz alignieren.
- b) Verwenden Sie die Alignment-Sammlung aus a) um die Sequenzen 3 und 4 end-gap-free und ohne Gapkosten zu alignieren. Ermitteln Sie dafür die Ähnlichkeiten für die gematchten Positionen durch Zählen (ohne zusätzliches Gewichten anhand der Sequenzen).