

# 7. Übung zur Vorlesung "Sequenzanalyse"

Sebastian Böcker, Markus Fleischauer

## Aufgabe 1 (6 Punkte)

- Gegeben seien ein Query-String  $s_1 = GAGCTA$ , ein Datenbank-String  $s_2 = GATCGAGCAA$  und ein spaced seed  $seed = 1101$ . Geben Sie sowohl für  $s_1$  als auch für  $s_2$  eine (alphabetisch) sortierte Liste aller q-Gramme und die zugehörigen Indizes der Vorkommen an.
- Entwickeln Sie einen Algorithmus mit dessen Hilfe Treffer in Linearzeit gefunden werden können und geben Sie diese Treffer für das Beispiel aus a) an.

## Aufgabe 2 (8 Punkte)

Gegeben seien drei Sequenzen  $s_1 = TACA$ ,  $s_2 = CTAC$ ,  $s_3 = GTAG$ . Benutzen Sie den in der Vorlesung angegebenen dynamischen Programmierungsalgorithmus um das Sum-of-Pairs optimale multiple Sequenzalignment mit Einheitskosten zu berechnen.

## Aufgabe 3 (4 Punkte)

Gegeben seien zwei Sequenzen  $s_1 = AGATC$ ,  $s_2 = TACATA$ . Berechnen Sie die Beste-Kosten-Matrix  $M_{1,2}$  mit Einheitskosten.

## Aufgabe 4 (7 Punkte)

Zusätzlich zu den Sequenzen aus Aufgabe 3 sei die Sequenz  $s_3 = GAGAT$  und der multiple Alignmentscore 10 für ein heuristisches Alignment dieser drei Sequenzen gegeben. Berechnen Sie die Beste-Kosten-Matrizen  $M_{1,3}$  und  $M_{2,3}$  mit Einheitskosten. Markieren Sie die Bereiche, die in der Rückprojektion in der dreidimensionalen Edit-Matrix nicht berechnet werden müssen.