# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022
Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga
`https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa`

**Übungsblatt 10 vom 9.6.2022**
**Abgabe am 23.6.2022 bis 12:00 Uhr (mittags)**

**Aufgabe 1 (Burrows-Wheeler-Transformation)** (9 Punkte)

1. Gegeben sei der String $t =$ `DDACCBDDACBDDCCB`.
   (a) Berechne die Burrows-Wheeler-Transformierte $s = \text{BWT}(t\$)$. Gib alle Zwischenschritte an.
   (b) Demonstriere den String-Matching-Algorithmus unter Verwendung der Sequenz $s = \text{BWT}(t\$)$ beispielhaft an der Suche des Musters $p =$ `ACBDD` in $t$.
   (c) Schreibe $r = \text{RLE}(s)$ als komprimierten String mit Hilfe von *run-length encoding* auf. Fasse dabei nur Buchstaben zusammen, die mindestens dreimal hintereinander vorkommen.
2. Gegeben sei der String $r =$ `BD4CDA$CA3DBB`.
   (a) Dekomprimiere $r$ in $s = \text{RLE}^{-1}(r)$.
   (b) Rekonstruiere den String $t\$ = \text{BWT}^{-1}(s)$. Gib alle Zwischenschritte an.

**Aufgabe 2 (Sum-of-Pairs Multiple Alignment)** (3 Punkte)

Gegeben sei das folgende Alignment $A$ von 4 Sequenzen:

$$A = \begin{pmatrix} \texttt{A} & \texttt{C} & \texttt{-} & \texttt{-} & \texttt{-} & \texttt{G} & \texttt{T} \\ \texttt{-} & \texttt{C} & \texttt{-} & \texttt{-} & \texttt{-} & \texttt{G} & \texttt{T} \\ \texttt{A} & \texttt{-} & \texttt{A} & \texttt{A} & \texttt{G} & \texttt{G} & \texttt{-} \\ \texttt{-} & \texttt{C} & \texttt{A} & \texttt{C} & \texttt{G} & \texttt{G} & \texttt{C} \end{pmatrix}$$

Betrachte die folgende Score-Funktion: match $= 4$, mismatch $= -1$ sowie affine Gapkosten mit gap-open $=$ 2 und gap-extension $= 1$. Berechne den Sum-of-Pairs-Score von $A$. Gib alle Zwischenschritte deiner Berechnung an.

**Aufgabe 3 (Sum-of-Pairs Multiple Alignment)** (8 Punkte)

Gegeben seien die Sequenzen $s_1 =$ `CAG`, $s_2 =$ `AGT` und $s_3 =$ `CAT`. Betrachte die folgende Score-Funktion: match $= 2$ und mismatch $=$ gap $= -1$.

1. Berechne für jedes der drei Paare $(s_1, s_2)$, $(s_1, s_3)$ und $(s_2, s_3)$ die zweidimensionale Alignment-Matrix und ein optimales globales Alignment.
2. Lässt sich durch die Kombination zweier dieser Alignments ein optimales globales multiples Alignment $A$ von $s_1$, $s_2$ und $s_3$ bilden? Wie lautet es und welchen Sum-of-Pairs-Score hat es?

   Berechne die drei Projektionen $\pi_{\{1,2\}}$, $\pi_{\{1,3\}}$ und $\pi_{\{2,3\}}$ des Alignments $A$ und ihre Scores. Welche der Projektionen korrespondieren zu dem oben berechneten optimalen Alignment und welche nicht?

**Aufgabe 4 (NP-vollständige Probleme)** (1* Punkte)

Gegeben sei ein NP-vollständiges Problem $Q$. Was kann über die Komplexität eines anderen Problems $Q'$ gesagt werden, wenn ...

1. ... $Q' \leq_p Q$?
2. ... $Q \leq_p Q'$?

**Aufgabe 5 (Umgang mit NP-vollständigen Problemen)** (4* Punkte)

Angenommen, ein Problem ist NP-vollständig.

1. Wie kannst du vorgehen, um zu einer Lösung zu kommen? Zähle fünf Möglichkeiten auf, die du je mit einem kurzen Satz beschreibst.
2. In welchen Fällen ist dein Ergebnis noch korrekt? Was kannst du über eventuelle Abweichungen vom korrekten Ergebnis sagen?
3. Wenn du in der Lage wärst zu zeigen, dass sich dein NP-vollständiges Problem in polynomieller Zeit lösen lässt, welche Folge hätte das für andere NP-vollständige Probleme?

**Aufgabe 6 (Polynomielle Zeitverifikation)** (5* Punkte)

Ein Hamilton-Kreis in einem ungerichteten Graph $G = (V, E)$ ist ein einfacher Kreis, der jeden Knoten in $V$ genau einmal enthält. Ein Graph, der einen Hamilton-Kreis enthält, wird Hamilton-Graph genannt. Das Hamilton-Kreis-Problem lautet: Enthält ein Graph $G$ einen Hamilton-Kreis?
Dieses Problem ist NP-vollständig.

1. Du weißt, dass das Hamilton-Kreis-Problem NP-vollständig ist. Welche Aussage kannst du dann über einen Verifikationsalgorithmus treffen, der, gegeben eine mögliche Anordnung von Knoten im Graph $G$, herausfindet, ob $G$ ein Hamilton-Graph ist?
2. Formuliere selber solch einen Verifikationsalgorithmus und analysiere seine Laufzeit.