# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2021
Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

https://gi.cebitec.uni-bielefeld.de/teaching/2021summer/sa

**Übungsblatt 5 vom 20.5.2021**
**Abgabe am 27.5.2021 bis 12:00 Uhr (mittags)**

**Aufgabe 1 (Optimale globale Alignments)** (5 Punkte)

Gegeben seien die Sequenzen $x =$ `AGAT` und $y =$ `ACGT`. Wir definieren die folgenden Scores für die Editoperationen Insertion und Deletion: $\mathcal{I} = \mathcal{D} = -2$. Für Copy ($\mathcal{C}$) und Substitution ($\mathcal{S}$) gilt die Scorematrix $\mathcal{M}^{\text{SC}}$:

| | A | C | G | T |
|---|---|---|---|---|
| A | $+1$ | $-2$ | $-1$ | $-2$ |
| C | $-2$ | $+1$ | $-2$ | $-1$ |
| G | $-1$ | $-2$ | $+1$ | $-2$ |
| T | $-2$ | $-1$ | $-2$ | $+1$ |

1. Zeichne den globalen Alignmentgraphen für $x$ und $y$ mit *allen* Kanten und Knoten.
2. Füge deiner Zeichnung die Kantenbeschriftungen $\lambda$ und die Kantengewichte $w$ hinzu.
3. Berechne alle Knotenwerte $S(v)$.
4. Gib alle optimalen globalen Alignments an und markiere deren Pfade im Graphen.

**Aufgabe 2 (Berechnung von Alignments)** (7 Punkte)

Gegeben seien die Sequenzen $x =$ `AATTGCC` und $y =$ `ATGTGAATG` sowie die folgende Score-Funktion: Match $= 4$, Mismatch $= -3$, Indel $= -2$.

1. Berechne alle optimalen globalen Alignments von $x$ und $y$.
2. Berechne alle optimalen *free-end gap* Alignments von $x$ und $y$.
3. Berechne alle optimalen lokalen Alignments von $x$ und $y$.

Mache in deiner Lösung deutlich, wie du zu dem Ergebnis gekommen bist, z.B. indem du die berechneten Alignment-Matrizen angibst.

**Aufgabe 3 (Affine Gapkosten)** (8 Punkte)

1. Warum sollten bei affinen Gapkosten die Gap-open-Kosten $d$ nicht niedriger gewählt als die Gap-extension-Kosten $e$, also $d \geq e$?
2. Zeige, dass affine Gapkosten subadditiv sind, also dass gilt: $g(\ell_1 + \ell_2) \leq g(\ell_1) + g(\ell_2)$
3. Berechne ein optimales globales Alignment mit affinen Gapkosten der Sequenzen $x =$ `ATCCTAG` und $y =$ `ATTGCCCT` effizient mit Hilfe des Gotoh-Algorithmus (berechne die Matrizen $S$, $V$ und $H$) und gib dessen Gesamtscore an. Verwende dabei: Score für Match $= 3$, Score für Mismatch $= -2$, Kosten für Gap-open $d = 3$, sowie Kosten für Gap-extension $e = 1$.

Du kannst diese Seite für deine Lösung verwenden:

**Aufgabe 2:** Score-Funktion: $\begin{cases} \text{Match} = 4 \\ \text{Mismatch} = -3 \\ \text{Indel} = -2 \end{cases}$

1. **Global Alignments:**

| | $\varepsilon$ | A | T | G | T | G | A | A | T | G |
|---|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | | |
| A | | | | | | | | | | |
| A | | | | | | | | | | |
| T | | | | | | | | | | |
| T | | | | | | | | | | |
| G | | | | | | | | | | |
| C | | | | | | | | | | |
| C | | | | | | | | | | |

2. **Free-end-gap Alignments:**

| | $\varepsilon$ | A | T | G | T | G | A | A | T | G |
|---|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | | |
| A | | | | | | | | | | |
| A | | | | | | | | | | |
| T | | | | | | | | | | |
| T | | | | | | | | | | |
| G | | | | | | | | | | |
| C | | | | | | | | | | |
| C | | | | | | | | | | |

3. **Local Alignments:**

| | $\varepsilon$ | A | T | G | T | G | A | A | T | G |
|---|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | | |
| A | | | | | | | | | | |
| A | | | | | | | | | | |
| T | | | | | | | | | | |
| T | | | | | | | | | | |
| G | | | | | | | | | | |
| C | | | | | | | | | | |
| C | | | | | | | | | | |

Du kannst diese Seite für deine Lösung verwenden:

**Aufgabe 3:** $\begin{cases} \text{Score: Match} = 3\,, \ \text{Mismatch} = -2 \\ \text{Kosten: Gap-open } d = 3\,, \ \text{Gap-extension } e = 1 \end{cases}$

| **V** | $\varepsilon$ | A | T | T | G | C | C | C | T |
|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | |
| A | | | | | | | | | |
| T | | | | | | | | | |
| C | | | | | | | | | |
| C | | | | | | | | | |
| T | | | | | | | | | |
| A | | | | | | | | | |
| G | | | | | | | | | |

| **S** | $\varepsilon$ | A | T | T | G | C | C | C | T |
|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | |
| A | | | | | | | | | |
| T | | | | | | | | | |
| C | | | | | | | | | |
| C | | | | | | | | | |
| T | | | | | | | | | |
| A | | | | | | | | | |
| G | | | | | | | | | |

| **H** | $\varepsilon$ | A | T | T | G | C | C | C | T |
|---|---|---|---|---|---|---|---|---|---|
| $\varepsilon$ | | | | | | | | | |
| A | | | | | | | | | |
| T | | | | | | | | | |
| C | | | | | | | | | |
| C | | | | | | | | | |
| T | | | | | | | | | |
| A | | | | | | | | | |
| G | | | | | | | | | |