

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2025/26

Prof. Dr. Jens Stoye

Leonard Bohnenkämper

Tutorien: Lennart Finke, Sofie Jans

<https://gi.cebitec.uni-bielefeld.de/teaching/2025winter/sa1>

Übungsblatt 4 vom 20.11.2025

Abgabe bis 27.11.2025 bis 9:30 Uhr per Mail an den Tutor/die Tutorin

Aufgabe 1 (Alignments)

(5 Punkte)

Welche Art von Alignment (global, semi-global, free-end gap, lokal) ist jeweils am besten für die Anwendung in folgenden fiktiven Situationen geeignet? Begründe deine Antwort.

1. Gegeben sind überlappende Sequenzfragmente mit Sequenzierfehlern einer längeren Sequenz, aus denen die ursprüngliche Sequenz bestimmt werden soll. Hierzu muss der paarweise Überlapp der Sequenzen berechnet werden.
2. Gegeben sind zwei Proteinsequenzen. Es soll herausgefunden werden, ob sich diese eine Domäne, d.h. einen mehr oder weniger gleichen Abschnitt, teilen.
3. Ein unbekanntes Gen wurde sequenziert. Aus einer Gruppe bekannter Gensequenzen soll die dem unbekannten Gen ähnlichste Sequenz gefunden werden.
4. Gegeben ist eine unbekannte Gensequenz und eine bekannte Genomsequenz. Es soll herausgefunden werden, ob dieses Gen oder ein sehr ähnliches in diesem Genom vorkommt.

Aufgabe 2 (Berechnung von Alignments)

(10 Punkte)

Gegeben seien die Sequenzen $x = \text{AGTCCCTA}$ und $y = \text{CGTACCGGTAA}$ sowie die Score-Funktion: Match = 2, Mismatch = -1, Indel = -2.

1. Berechne ein optimales globales Alignment von x und y .
2. Berechne ein optimales semi-globales Alignment von x und y .
3. Berechne ein optimales *free-end gap* Alignment von x und y .
4. Berechne ein optimales lokales Alignment von x und y .

Aufgabe 3 (Gotoh-Algorithmus)

(5 Punkte)

1. Warum sollten die Kosten für eine *gap extension* nicht höher als die Kosten für ein *gap open* gewählt werden, also $e \leq d$?
2. Zeige, dass affine Gapkosten subadditiv sind.
3. Berechne ein optimales globales Alignment mit affinen Gapkosten von den Sequenzen $x = \text{CTATGTGA}$ und $y = \text{CTGA}$ effizient mit Hilfe des Gotoh-Algorithmus und gib dessen Gesamtscore an. Verwende dabei: Score für Match = 2, Score für Mismatch = 0, Kosten für Gap-open $d = 2$, sowie Kosten für Gap-extension $e = 0.5$.