

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2013/2014

Prof. Dr. Jens Stoye · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/SequenzAnalyse>

## Übungsblatt 6 vom 26.11.2013

Abgabe in einer Woche vor Beginn der Vorlesung.

### Aufgabe 1 (Gotoh)

(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  $e = 0.5$ .

### Aufgabe 2 (Suboptimale Alignments)

(6 Punkte)

1. Beschreibe in eigenen Worten, was man unter *überlappenden* Alignments versteht.
2. Warum ist man bei der Bestimmung von suboptimalen Alignments in erster Linie an *nichtüberlappenden* Alignments interessiert? Welche Probleme möchte man vermeiden?
3. Beschreibe kurz die Funktionsweise des Waterman-Eggert-Algorithmus. Durch welchen Trick kann man in der Praxis die Laufzeit verkürzen?
4. Gegeben seien  $x = \text{GTAA}$  und  $y = \text{GCTA}$ . Berechne ein lokales Alignment mit dem Smith-Waterman-Algorithmus und das erste nicht überlappende, suboptimale Alignment nach *Waterman-Eggert*. Verwende dazu: Score für Match = 3, Score für Mismatch = 1, Kosten für Indels = 1.

### Aufgabe 3 (Approximatives Stringmatching)

(4 Punkte)

1. Finde die Endpositionen aller Vorkommen des Patterns  $x = \text{ATAT}$  im Text  $y = \text{GGTGATATGTAAAC}$  mit maximal  $k = 1$  Fehlern. Verwende die Cutoff-Variante von *Sellers' Algorithmus* mit Einheitskosten und markiere die *last essential indices*.
2. Gib zu jeder gefundenen Endposition alle zugehörigen Alignments an.
3. Warum macht es Sinn, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? (Stichwort: *Runs*.) Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?