

# Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WiSe 2009/2010

Prof. Dr. Jens Stoye · Dipl.-Inform. Nils Hoffmann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/SequenzAnalyse>

**Blatt 10 vom 17.12.2009**

**Abgabe 7.01.2010 vor der Vorlesung um 8:30 in H5**

## Aufgabe 1 Waterman-Eggert-Algorithmus

(2 Punkte)

1. In welchem biologischen Kontext sind suboptimale Alignments interessant? Wie findet der *Waterman-Eggert*-Algorithmus solche Alignments?
2. Gegeben sei  $x = \text{ACTG}$ ,  $y = \text{ATGT}$ , berechne ein lokales Alignment mit dem Smith-Waterman-Algorithmus und das erste nicht überlappende, suboptimale Alignment nach *Waterman-Eggert*. Verwende dazu Scores mit  $\text{score}(C) = 3$ ,  $\text{score}(S_{a,c}) = 1$  für  $a \neq c$  und  $\text{score}(I_c) = \text{score}(D_c) = -1$ .

## Aufgabe 2 Sellers' Algorithmus

(3 Punkte)

1. Finde die Endpositionen aller semiglobalen Alignments für maximal  $k = 1$  Fehler von  $x = \text{TATA}$  in  $y = \text{ACTATAGATA}$  mit Hilfe von *Sellers'* Algorithmus. Verwende dazu Einheitskosten.
2. Markiere die *last essential indices* für die Cutoff-Variante in der Alignmentmatrix mit Kreisen.

## Bonuszettel

Alle ab hier erreichbaren Punkte werden Eurem Punktekonto als Bonus gutgerechnet.

## Aufgabe 3 Globales Alignment, homogene Gapkosten

(2 Punkte)

Gegeben sind die Sequenzen  $x = \text{RHABARBER}$  und  $y = \text{BARBARA}$ . Es sollen die **SCORES**  $C = 3$ ,  $S_{a,c} = 2$  für  $a \neq c$  und  $I = D = -1$  verwendet werden.

1. Berechne das globale Alignment von  $x$  und  $y$  nach Needleman-Wunsch und gib dessen Wert an.
2. Gib alle optimalen Edit-Sequenzen und das jeweils dazugehörige optimale Alignment an.

## Aufgabe 4 Lokales Alignment, homogene Gapkosten

(2 Punkte)

Gegeben sind die Sequenzen  $x = \text{RHABARBER}$  und  $y = \text{BARBARA}$ . Es sollen die **SCORES**  $C = 3$ ,  $S_{a,c} = 2$  für  $a \neq c$  und  $I = D = -1$  verwendet werden.

1. Berechne das lokale Alignment von  $x$  und  $y$  nach Smith-Waterman und gib dessen Wert an.
2. Gib alle optimalen Edit-Sequenzen und das jeweils dazugehörige optimale Alignment an.

## Aufgabe 5 Globales Alignment, affine Gapkosten

(3 Punkte)

Gegeben sind die Sequenzen  $x = \text{AACGGTAA}$  und  $y = \text{AAAA}$ . Es sollen die **SCORES**  $C = 2$ ,  $S_{a,c} = 1$  für  $a \neq c$  und  $d = 1$  (Gap open), sowie  $e = 0.5$  (Gap extension) verwendet werden.

1. Berechne das globale Alignment mit affinen Gapkosten von  $x$  und  $y$  effizient mit Hilfe des Gotoh-Algorithmus (Matrizen  $S$ ,  $H$  und  $V$ ) und gib dessen Wert an.
2. Gib eine optimale Edit-Sequenz und das dazugehörige optimale Alignment an.

## Aufgabe 6 Globales Alignment, free end gaps

(2 Punkte)

Gegeben sind die Sequenzen  $x = \text{AACGGTAA}$  und  $y = \text{CGGC}$ . Es sollen die **SCORES**  $C = 2$ ,  $S_{a,c} = 1$  für  $a \neq c$  und  $I = D = 0.5$  verwendet werden.

1. Berechne das globale Alignment mit free end gaps von  $x$  und  $y$ .
2. Gib eine optimale Edit-Sequenz und das dazugehörige optimale Alignment an.

**Aufgabe 7** Alignmentgraph

(2 Punkte)

1. Zeichne den Alignmentgraphen schematisch für das globale Alignment.
2. Zeichne den Alignmentgraphen schematisch für das free end gaps Alignment.

**Aufgabe 8** q-gram-Distanz

(2 Punkte)

Gegeben sind die Sequenzen  $x = \text{BALUDERBAER}$  und  $y = \text{BRUDER}$ .

1. Gib die q-gram-Profile von  $x$  und  $y$  an.
2. Berechne die q-gram-Distanz für  $q = 4$  von  $x$  und  $y$ .

**Aufgabe 9** Maximal-Matches-Distanz

(2 Punkte)

Gegeben sind die Sequenzen  $x = \text{BALUDERBAER}$  und  $y = \text{BRUDER}$ .

1. Berechne  $\delta(x||y)$  und  $\delta(y||x)$  und gib die Partitionen an.
2. Gib die Maximal-Matches-Distanz von  $x$  und  $y$  an.

**Aufgabe 10** Lokales Alignment nach Smith-Waterman

(5 Punkte)

Implementiere eine JAVA-Klasse `LocalAlignment`, die das lokale Alignment zweier Strings mit Hilfe des Smith-Waterman-Algorithmus und homogenen Gapkosten berechnet. Die Methode, in der diese Funktionalität realisiert werden soll hat die Signatur

```
public double getScore(String x, String y)
```

Die Methode soll nach Berechnung des lokalen Alignments ein optimales lokales Alignment auf der Konsole ausgeben. Es soll weiterhin möglich sein, über eine Methode

```
public void setScores(double copy, double subst, double indel)
```

die Scores der einzelnen Operationen zu setzen. Eine weitere Methode

```
public void setScoreMatrix(ScoreMatrix sm, double indel)
```

soll das Setzen einer Scorematrix erlauben, sowie die Festlegung der Indelscores.

Frohe Weihnachten und einen guten Rutsch ins neue Jahr!