

# Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2007

Dr. Sven Rahmann · Dipl.-Inf. Peter Husemann · Dipl.-Biol. Constantin Bannert

<http://gi.cebitec.uni-bielefeld.de/teaching/2007summer/sequenzanalyse/>

**Blatt 8 vom 08.06.2007**

**Abgabe am 15.06.2007 vor der Vorlesung um 8:30 in H14**

**Aufgabe 1** In der Vorlesung wurden die Carillo-Lipman-Schranken für Score-basiertes Alignment angegeben. Fasse die wichtigen Ideen kurz für Kosten-basiertes Alignment zusammen.

**Aufgabe 2** Charakterisiere Sequenzen, auf denen die Carillo-Lipman-Heuristik gut funktioniert, und solche, auf denen sie nicht gut funktioniert.

**Aufgabe 3** Analysiere die Laufzeit der center-star Methode.

**Aufgabe 4** Zur Wiederholung: Beim Alignment mit linearem Platzbedarf kann man so vorgehen, dass man eine Zelle in der mittleren Zeile findet, durch die ein optimales Alignment hindurchgeht. Dazu merkt man sich, für jede Zelle  $(i, j)$  unterhalb der mittleren Zeile, einen Zeiger (Spaltennummer)  $c(i, j)$  auf die mittlere Zeile, so dass  $c(i, j) = x$  bedeutet, dass ein optimales Alignment, das in  $(i, j)$  endet, durch  $(\lfloor m/2 \rfloor, x)$  geht.

Für die Sequenzen GAATTC und GATTA, Matchscore +2, Mismatchscore -1, lineare Gapkosten von -2, berechne die DP-Matrix und die  $c(i, j)$ -Einträge.