

# 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 7 vom 01.06.2007**

**Abgabe am 08.06.2007 vor der Vorlesung um 8:30 in H14**

**Aufgabe 1** Gegeben sind die vier Sequenzen  $s_1 = \text{ACCA}$ ,  $s_2 = \text{TCCA}$ ,  $s_3 = \text{ATC}$ ,  $s_4 = \text{ATCCCA}$ , sowie linear-homogene Gapkosten von  $(-)$ 2 und folgende Substitutions-Scores von +1 für einen match und  $-2$  für einen mismatch (gap vs. gap gibt 0 Punkte).

Weiterhin sei ein ungewurzelter Baum gegeben, in dem  $s_1$  und  $s_2$  benachbarte Blätter sind, und ebenso  $s_3$  und  $s_4$ .

Ermittle sinnvolle (idealerweise: scoremaximierende) Sequenzen für die gemeinsamen Vorfahren  $x$  von  $s_1$  und  $s_2$ , sowie  $y$  von  $s_3$  und  $s_4$ . Berechne dann den Baum-Alignment-Score.

**Aufgabe 2** Gegeben sind  $N$  Sequenzen der Länge  $L = 500$ . Ein paarweises Alignment dauert ca. eine Sekunde. Wie lange, in etwa, dauert dann das exakte dynamic programming Alignment von  $N$  Sequenzen? Wie viele Sequenzen kann man alignieren, bevor in 5 Milliarden Jahren die Sonne verlöscht? Spezifiziere alle Annahmen zum Lösen dieser Aufgabe.