Universität Bielefeld
Technische Fakultät

AG Genominformatik
Dipl.-Inform. Klaus-Bernd Schürmann

Lecture: Spezielle Algorithmen der Sequenzanalyse
Summer semester 2006

Exercises

Exercise 2, Discussion: 04/19/2006.

1. **Cost functions.**
   Calculate the aligment cost of the alignments given below:

$$A_1 = \begin{pmatrix} A & A & T & C & A & T & - & - & - & T \\ - & - & T & - & A & C & A & T & G & T \end{pmatrix}$$

$$A_2 = \begin{pmatrix} A & A & T & C & A & T & - & - & - & T \\ - & - & - & T & A & C & A & T & G & T \end{pmatrix}$$

   (a) for unit costs;

   (b) for affine gap costs where the cost function is:

$$\delta(c, c') = \begin{cases} 0 & \text{if } c,\ c' \in \Sigma \text{ and } c = c' \\ 1 & \text{if } c,\ c' \in \Sigma \text{ and } c \neq c' \end{cases}$$

   with gapinit $= 3$ and gapext $= 1$;

   (c) and for general gapcosts where δ is as above and the cost function is defined by the table

l	1	2	3	4	5	6
g(l)	2	2	1	3	3	2

2. **Subadditiv gap costs.**
   A gap cost function g ist subadditive, if $g(l_1 + l_2) \leq g(l_1) + g(l_2)$ for all possible gap lengths l_1 and l_2.
   Prove or disprove:

   (a) Homogeneous gap costs are subadditive.

   (b) Affine gap costs are subadditive.

   (c) The gap costs

$$g(l) = \begin{cases} l & : \ 0 = l \mod 3 \\ l + 10 & : \ otherwise \end{cases}$$

   are subadditiv.

   (d) The gap costs $g(l) = l^2$ are subadditiv.

   (e) The gap costs $g(l) = \log l$ are subadditiv.

3. **Log-Odds Score.**
   Given an alignment database over the alphabet $\{A, C, G\}$ with $100,000$ total sites in these alignments and 1% mismatches. A occurs $70,000$ times and C $40,000$ times. We assume that the number of substitutions of A and G is $m(A, G) = 500$. Compute the log-odds score $\sigma^{(1)}(A, G)$.