1-Feb-05

BioE/MCB/PMB C146/246, Spring 2005

Problem Set 3

Due 7 February 2005 by midnight to lfl@compbio.berkeley.edu

1. 15 points

Sequence Alignment with Affine Gaps

Revise your alignment program from the last problem set using the following parameters:

Identity	+4
Transition	–2
Transversion	–4
Gap	–8 (first position), –1 (each gap position)

Given the sequences

ATTTTAAGCGCATACCGC
TCGCAAATATAC

Perform a global alignment on the two sequences and report the score. Attach all dynamic programming matrices (no tracebacks) to your email as *`jrandom`*`_ps3_1.txt`

2. 20 points

A. Perform linear-space alignment on the following sequences: HECYDWH and HEWGH. Extra credit for implementing the Hirschberg/Myers/Miller algorithm.

B. Perform any ONE of the following alignments, using the scoring matrix from Problem Set 2. You may use a program to assist you, though implementations are not required.

(i) Repeated matches, threshold 10 HECYDWH and HEWGH
(ii) Banded alignment HECYDWH and HEWGH
(iii) Sub-optimal global alignment HEAGAWGHE and PAWHEA
(iv) Global alignment with three-parameter gaps (–8, –1, –2) for the sequences
SSFTLT and SCHKDIL

Include dynamic programming matrices, traceback paths, alignments and scores in your answer. Attach the dynamic programming matrix to your email as *`jrandom`*`_ps3_2.txt` (tracebacks can be added by hand in the paper version).

3. 5 points

Why are sub-additive gap penalties used? Give (at least) two reasons.

4. 5 points

Compare and contrast the construction and features of the BLOSUM and PAM series of matrices. Mention the strengths and weaknesses of each.

5. 5 points

A 1-PAM matrix changes on average of 1% of amino acids. Does a 2-PAM matrix change on average 2%? Explain.

6. 5 points

For alignments performed with PAM matrices, explain the meaning of a substitution score and the score of the alignment.

7. 5 points

Why are gap parameters NOT estimated the same way as substitution matrix parameters?

8. 15 points

Given the following BLOCK (multiple sequence alignment of proteins):

```
MMKE
MKKE
IKIE
MEME
IMKI
IKKE
MKME
IKKE
MKME
IKKE
```

(A) (10 points) Compute the joint probabilities q_{ij} and the marginal probabilities p_i for each i, j in the amino acid alphabet.

(B) (5 points) Compute the BLOSUM matrix for this BLOCK.

9. 15 points

Given the initial mutability matrix below, calculate the corresponding 3-PAM matrix. Normalize your answer such that each row and column sums to 1000.

	S	T	V
S	990	7	3
T	7	993	0
V	3	0	997