Bioinformatics (CAP5937) Midterm Exam

Date: Oct. 9, 2006

Time: 75 minutes

Name:

1. (15 points) Use global dynamic programming to align the following two sequences. Fill the matrix and write down an optimal alignment.

Scoring function:

S(match) = 2, S(mismatch) = -1, S(gap, character) = -2

(1) Fill matrix (10 points)

	-	C	A	G	C
-					
C					
G					
C					
T					

(2) Optimal alignment (5 points):

2. (10 points) Describe **two** main differences between local dynamic programming algorithm (Smith-Waterman algorithm) and global dynamic programming algorithm (Needleman-Wunsch algorithm).

3. (10 points) List three ways to represent sequence profile.

4. (10 points) Why dynamic programming is usually not used for multiple sequence alignment? Describe the idea of the progressive multiple sequence alignment.

5. (10 points) Describe PSI-BLAST algorithm? What distribution does PSI-BLAST alignment score follow?

6. (5 points) For the scoring matrix BLOSUM62, what does “62” stand for? List one difference between the process of constructing BLOSUM matrix and PAM matrix.

7. (10 points) Describe Gibbs sampling algorithm for motif finding (or local multiple sequence alignment) according to the following three steps. (hints: two key terms: the start position of a motif and probability matrix)

Initialization

Repeat

Until

8. (10 points) To construct a 2^{nd} order Markov model for gene sequences composed of four nucleotides (A, C, G, T), how many parameters are required?

9. (10 points) List and explain three major problems of hidden Markov model

Problem 1:

Problem 2:

Problem 3:

10. (10 points) Construct an un-rooted phylogenetic tree using the maximum parsimony algorithm for the following four sequences using site 3 (column 3). **Show** all three possible trees, **report** the number of mutations of each tree, and **select** the best tree.

1.	A	C	**G**	T	A
2.	C	C	T	C	A
3.	C	T	G	A	C
4.	T	A	T	C	A

Site 3