Mansoura University

Faculty of Agriculture

Biotechnology Department

Bioinformatics

3rd Year (Final Exam)

14-06-2015 (Time: 2h)

Answer the following questions (Total Marks: 60 M):

1. [10 M] True or false. Correct the wrong statement:

a. One of the main goals of bioinformatics is to use of software tools and biological databases to extract new knowledge.

b. Comparison of sequences between individuals can detect changes that are related to diseases.

c. Smith-Waterman technique is a global alignment method, which is used to best score for aligning part of sequences.

d. The S-value is the number of alignments with scores greater than or equal to score E that are expected to occur by chance in a database search.

e. There are different PMA matrices. The matrices are derived from each other by multiplying the PMA1 matrices N times

f. BLOSUM matrices are based on the replacement patterns found in more highly conserved regions of the sequences without gaps, which is equal to local alignment.

g. Sequence motifs represent a **short common sequence** (length **30-50**) which is highly represented in the data.

2. [10 M] Fill in the spaces:

a. Bioinformatics is the field of science in which ……………, ……………, and …………… are merged.

b. The ultimate goal of the bioinformatics is to enable the …………… as well as to …………… from which unifying principles in biology can be discerned.

c. The …………… algorithm allows several segments that are separated by short gaps to be connected together to one alignment.

d. The BLAST-E value is calculated by ……………, whereas the gap scores is calculated by ……………………

e. …………………… gives an alignment of closely related sequences. We can score the relation between amino acids based on how frequently they substitute each other.

f. …………………… matrices based on mutations observed throughout a global alignment, includes both highly conserved and highly mutable regions.

g. …………………… is an approximation strategy (heuristic algorithm) yields a possible alignment, but not necessarily the best one.

3. [5 M] Write the main steps of Bootstrapping tree?

4. [10 M] Finding TF targets using a bioinformatics approach that has binding motif is known. Starting from the following set of aligned motifs and length l=7:

Seq 1	AAAGCCC
Seq 2	CTATCCA
Seq 3	CTATCCC
Seq 4	CTATCCC

Seq 5	GTATCCC
Seq 6	CTATCCC
Seq 7	CTATCCC
Seq 8	CTATCCC
Seq 9	TTATCTG

a. Determine the Position Weight Matrix (PWM).
b. Calculate Pr(CTAATCCG).

5. [10 M] There are four sequences as shown below:

Sequence a: ACGCGTTGGGCGATGGCAAC
Sequence b: ACGCGTTGGGCGACGGTAAT
Sequence c: ACGCATTGAATGATGATAAT
Sequence d: ACACATTGAGTGTGATAATA

Using the UPGMA distance method to build a dendrogram for these four sequences.

6. [5 M] Write the main steps of BLAST algorithm? How to interpret a BLAST search?

7. [10 M] Using Needleman-Wunsch sequence global alignment method to calculate the similarity of these two sequences: A = ACGCTG, B = CATGT.

With my best wishes
Dr. Mohammed Elmogy