

Qualifying Exam – Bioinformatics
Spring 2009
Closed books and lecture notes, open paper 1-paper 5

Problem 1 (25 points)

- (a). Describe how hashing and chaining is used in FASTA to find a word match for two protein sequences.
- (b). Describe an alternative method used in BLAST to find a match of similar words.
- (c). Suppose that an HSP is located at (i, j) where i is the starting position of the HSP in sequence 1 and j is the starting position of the HSP in sequence 2. Describe in details the “X-drop” method used in BLAST to obtain an alignment in the region of the HSP.

Problem 2 (20 points)

- (a). List the nature of genomes that will make the pair-wise alignment methods such as dynamic programming and BLAST useless in aligning two genomes.
- (b). Given two genomes, describe **how** a unique match is found from a suffix tree. Describe how a MUM can be found.
- (c). Describe how the alignment of the genomes is generated after the MUMs are found in the MUMer method.

Problem 3 (30 points) (Paper 5 by Barenboim et al, “Statistical geometry based prediction of nonsynonymous SNP function effects using random forest and neuro-fuzzy classifiers”)

- (a). On page 1932, it says Q_{wt} is calculated using the “scores of quadruplets”. What is the score of a quadruplet? How is it derived?
- (b). Describe the method of determination of the most important three factors ΔQ , V , and sT among the 17 factors.
- (c). Give at least one advantage and one disadvantage of using the Delaunay tessellation as the measurement of the structure potential.
- (d). Propose a measurement of the structure potential that improves from the Delaunay tessellation. You need to build your proposal with sufficient supports.

Problem 4 (25 points) (Paper 3 by Altschul et al, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”)

- (a). On page 3390 at the line above equation 3, why does it say that the “aligned pair of letters (i, j) tends to occur with the target frequency q_{ij} that is “ $P_i P_j e^{\lambda u_{Sij}}$ ”? You may give your own justification about this.
- (b). Suggest an experiment to verify equation 3. Your experiment should have sufficient details for implementation.
- (c). Define each notation in equation 4. Explain how each of the notations can be obtained.