MBB 342 – Bioinformatics and Genomics - SAMPLE Exam Questions - *Answers*

These questions reflect the overall format and tone of the final exam. There will be a mix of multiple choice, definitions, short answer and a couple of longer answer questions. Questions where you match answers together may also be present. A reminder that the exam will be timed to take no longer than 2 hrs max. Note: For written questions you obviously don't have to have exactly what is written in the answers to be correct. Usually you will write a more detailed answer in some cases – I've just given the bare minimum to guide you.

1. Circle true or false (*2 marks each correct answer, –1 mark for each incorrect choice*)

a) Both maximum parsimony and neighbour joining use a distance matrix to create a phylogenetic tree from a multiple sequence alignment.
TRUE FALSE

b) Sequences with less than 15% identity are considered to be unrelated.
TRUE FALSE

c) All entries in GenBank are indexed by a unique accession number.
TRUE FALSE

2. Choose the most CORRECT answer. BLAST….

a) stands for Boring Long Anti Student Test (or is that EXAM?...)

b) and FASTA are both heuristic algorithms

c) cannot run locally on your computer unless you have internet access to a database to compare your query sequence to

d) is the first sequence comparison method to use statistics to identify all matches to a given sequence

3. Define the following terms (2 marks each)

a) Homolog
SHARE COMMON ANCESTRY. HAVE COMMON ORIGINS BUT MAY OR MAY NOT HAVE COMMON FUNCTION (DEPENDING ON HOW SPECIFIC A FUNCTION YOU ARE EXAMINING).

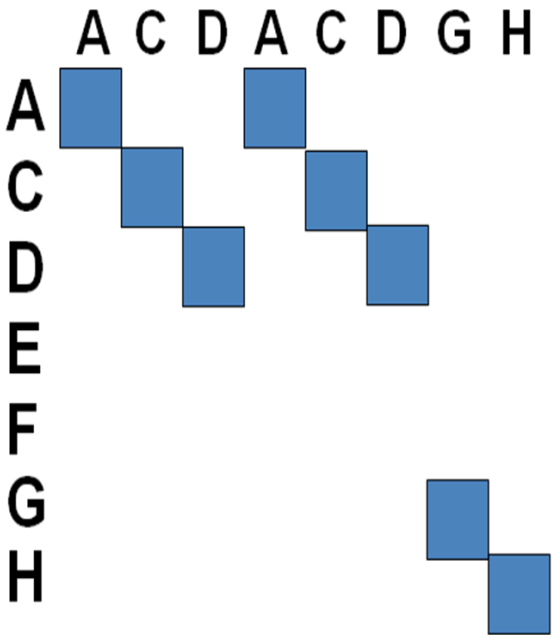
b) Paralog
HOMOLOGS PRODUCED BY GENE DUPLICATION. THEY USUALLY HAVE DIFFERING FUNCTIONS.

d) Sequence Contig
SET OF OVERLAPPING DNA SEGMENTS THAT TOGETHER REPRESENT A CONTIGUOUS, CONSENSUS REGION OF DNA.

4. What are the general benefits of next generation sequencing over Sanger sequencing?

TAKES ADVANTAGE OF MINIATURIZATION TO ENGAGE IN MASSIVELY PARALLEL ANALYSIS, ESSENTIALLY CARRYING OUT MILLIONS OF SEQUENCING REACTIONS SIMULTANEOUSLY. THIS ALLOWS SEQUENCING TO BE PERFORMED MUCH FASTER, AND MORE CHEAPLY.

5. Draw a dot plot illustrating how *direct* sequence repeats would look (3 marks)



6. Describe how a common property profile algorithm works, such as when plotting "secondary structure propensity" for a protein using the GOR algorithm.

ASSIGN EACH RESIDUE A NUMERIC VALUE CORRESPONDING TO THE PHYSICAL PROPERTY.

CHOOSE AN ODD NUMBERED WINDOW (USUALLY 5 OR 7) AND CALCULATE THE AVERAGE VALUE.

ASSIGN THE AVERAGE VALUE TO THE MIDDLE RESIDUE IN THE WINDOW.

MOVE THE WINDOW DOWN BY ONE RESIDUE AND REPEAT THE STEPS UNTIL FINISHED.

PLOT THE RESULTS.