Name: _____
Score: _____ / _____

Exam 2

Dear All,

Here is your exam for Bioinformatics (CS300). This exam is closed notes and closed book; you are not allowed to run any code on any computer to answer any questions. Unless you have made arrangements with the instructor, it is assumed that you will have 75 minutes to complete this exam. The exam is to be completed in Alden hall only. By the submission of your exam, you are agreeing to adhere to the honor code pledge.

Best of luck to all,
Dr. Bonham-Carter

Part 1

1

Which list represents the correct order for the steps of genome sequencing and assembly? Explain.

- A. Genomes are cloned, genomes are broken into fragments, fragments are sequenced, overlapping fragments are assembled to construct the genome sequence.
- B. Genomes are cloned, fragments are sequenced, genomes are broken into fragments, overlapping fragments are assembled to construct the genome sequence.
- C. Overlapping fragments are assembled to construct the genome sequence, genomes are cloned, genomes are broken into fragments, fragments are sequenced.
- D. Genomes are broken into fragments, fragments are sequenced, genomes are cloned, overlapping fragments are assembled to construct the genome sequence.

2

Below are four sequence fragments that can be assembled to create a single contig sequence. Show your alignment and the final contig sequence.

Fragments:

CGGACCAGA
ATCGGA
AGACTTTTTTACCAA
GACCAGACTTT

3

Annotating a newly sequenced genome by searching for known patterns such as ORFs is an example of:

- A. probabilistic annotation
- B. sequence-based annotation
- C. content-based annotation
- D. alignment-based annotation

4

Name and briefly describe two challenges faced when attempting to annotate a eukaryotic genome relative to prokaryotic genome annotation. (1 sentence description for each challenge)

5

What are PAM 120 and BLOSUM 62? Explain.

- A. Substitution matrices used for sequence alignment of DNA.
- B. Substitution matrices used for sequence alignment of a protein to its own DNA sequence.
- C. Substitution matrices used for sequence alignment of protein.
- D. Substitution matrices used for sequence alignment of RNA.
- E. Substitution matrices used for sequence alignment of a protein to its own RNA and associated DNA sequence.

6

What is the purpose of the tool, BLAST. Justify your answer.

- A. It is an open reading frame discovery tool.
- B. It is used to find regions of local similarity between sequences.
- C. It is a database whose principle task is to store the countless reads from an assembly task.
- D. It is a database which stores genes and supplies meta data concerning its origins, in addition to who was responsible for curating this information.
- E. It is a system to determine what products are transcribed and translated from original DNA.

7

What determines protein's shape and function? Explain.

- A. Amino acids determine a protein’s shape and function
- B. Open reading frames determine a protein's shape and function
- C. Nucleotides determine a protein’s shape and function
- D. DNA configurations determine a protein’s shape and function
- E. The methionine composition determines a protein’s shape and function

8

What are open reading frames (ORFs)? Explain.

- A. An ORF is a continuous stretch of proteins that never contains start codons (such as, AUG) and often has missing stop codon (such as, UAA, UAG or UGA).
- B. An ORF is a continuous stretch of codons that sometimes contains a start codon (usually AUG) and occasionally a stop codon (usually UAA, UAG or UGA).
- C. An ORF is a long stretch of continuous amino acids that are read in the 3' to 5' order and contains a start nucleotide which is immediately followed by a stop codon (sometimes AUG or, all together, UAA, UAG and UGA).
- D. An ORF is a continuous stretch of codons that contains a start codon (usually AUG) and a stop codon (usually UAA, UAG or UGA).

9

What is the Shine-Dalgarno Sequence? Justify.

- A. a ribosomal binding site found principally in cows, whales and some types of buffalo. This binding site has a length of about one amino acid and can serve as a landmark to help determine relations between these organisms.
- B. A ribosomal binding site in bacterial and archaeal messenger RNA, generally located around 8 bases found after (downstream of) a stop codon. This genetic landmark serves to mark the exact end of a gene region.
- C. A ribosomal binding site in bacterial and archaeal messenger RNA, generally located around 8 bases found after (downstream of) a start codon AUG and is a landmark inside a gene sequence.
- D. a ribosomal binding site in bacterial and archaeal messenger protein, which has been observed in the upstream sections of the starts of the codons in gene transcription TATA regulator expressions.
- E. A ribosomal binding site in bacterial and archaeal messenger RNA, generally located around 8 bases found before (upstream of) a start codon AUG. This sequence may serve as a genetic landmark outside of the gene for predicting the beginning of a region of gene code.

10

Imagine that you have just used BLAST to process an unknown protein sequence that you were given. When the results come up after processing, you note that there are several suggested protein sequences. Describe how you would use those sequences to determine the function of your unknown protein. In your discussion, describe what the E-value would tell you as a piece of meta-data for each of the suggested sequences.