MIS No.

Comp + IT

COLLEGE OF ENGINEERING PUNE

(An Autonomous Institute of Govt. of Maharashtra)

END -SEM EXAM

Computational Biology

Program: **Final Year B.Tech. (Computer Engg. / Information Tech.)**

Year: 2012-13 **Semester: Spring** Date: 23-04-13

Duration: 3 hrs. Max. Marks: 50.

Instructions:

1. **Make appropriate assumptions wherever necessary.**
2. **Give examples and draw neat diagrams wherever necessary.**

Q.1. A. Compare and contrast the probability mass function (PMF), probability density function (PDF) and cumulative distribution function (CDF). **(05)**

B. It has been estimated that about 30% of frozen chicken contain enough salmonella bacteria to cause illness if improperly cooked. A consumer purchases 12 frozen chickens. What is the probability that the consumer will have more than 6 contaminated chickens? **(05)**

OR

B. When robins' eggs are weighed, it turns out that they vary according to approximately the normal distribution with mean 17.5 mg and standard deviation 3.5 mg. What proportion of robins' eggs weigh between 10 mg and 25 mg? How large are the largest 2% of all robins' eggs? **(05)**

Q.2. A. **(05)**

a) Why do we use dynamic programming algorithms for pair wise sequence alignment problems but not for multiple pair wise alignments?

b) What is the difference between local alignment and global alignment

c) Given uses for global alignment and local alignment

B. Define the following (any Five): **(05)**

a. Cell
b. Nucleus
c. Genes
d. Alleles
e. Chromosomes
f. Genome

Q.3. a) What is the purpose of BLAST tool? **(01)**

b) Explain the algorithm step wise. **(05)**

c) Explain the impact of word size on the results. **(02)**

d) What is the typical value of work size for protein comparison and DNA comparison? **(01)**

e) Explain importance of e-value **(01)**

Q.4. **A.** Build the tree from the following distance matrix between species A,B,C,D using the UPGMA (Un-weighted Pair Group Method using arithmetic Averages) method **(05)**

	A	B	C	D
A		0.26	0.34	0.29
B			0.42	0.44
C				0.44
D				

B. Compute the global alignment between the two strings s1 = ACCGTT and s2 = AGTTCA, considering the following scoring parameters: +1 for match, −1 for mismatch, and −1 for a gap. **(05)**

(i) What is the maximum similarity score between the two sequences s1 and s2?

(ii) Find an alignment with this similarity score.

(iii) Is the alignment you found unique, or are there multiple alignments achieving the maximum similarity score?

Q.5. **A.** What is Next Generation sequencing? What are the applications of NGS technologies? **(04)**

B. Align the following sequence using Smith Waterman Algorithm, similarity=1, mismatch=1, gap=2, using S1=GCTGGAACCAG, S2=ACTGGATCAG. **(04)**

C. Describe the limitations of Phylogenetic trees in brief. **(02)**