

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

Application of Computer and Statistics in Biology BIO G510

FIRST SEMESTER 2016 – 2017

COMPREHENSIVE EXAM

TOTAL WEIGHTAGE 35% Date: 09.12.2016 DURATION: 3Hrs. (Part A & Part B)

Total Marks (46+24) =70

- Answer **Part A** and **Part B** in separate answer sheets.
- Irrelevant answer may attract penalty.

PART – A (CLOSED BOOK) (Max. duration: 2 Hrs., Max. Marks 46)

1. Write short note on following topic [2X5=10]

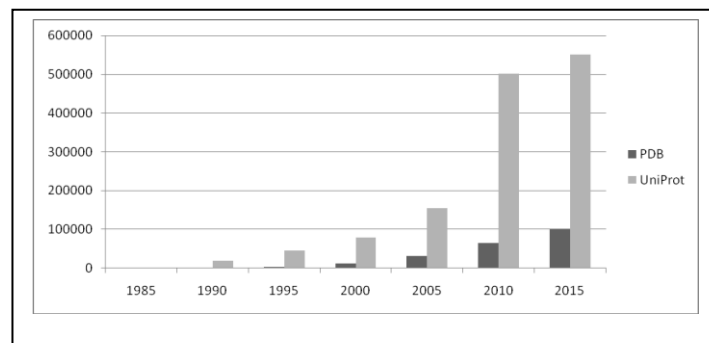
i) Biological database ii) *In-silico* gene prediction iii) Scoring matrix iv) Multiple sequence alignment v) Bootstrap

2. Compare the following pairs [2X5=10]

i) Global alignment and local alignment ii) Homology and similarity iii) Maximum Parsimony method and Neighbor joining method iv) Heuristic method of sequence alignment and deterministic method of sequence alignment v) rooted tree and unrooted tree.

3. a) Explain the observed data pattern in the following table and plot regarding growth of protein sequence (Uniprot/swissprot) and protein structure (PDB) database. [2]

Year	PDB Entry	Uniprot/Swissprot Entry
1985	173	1000
1990	450	20,000
1995	3449	45,000
2000	12136	80,000
2005	31149	155,000
2010	64681	501,000
2015	100,716	550,000



b) Write the Unix commands for the following questions: [1]

i) Save detailed list of all folders and files present in current folder as a file 2.txt.

ii) Make a file Z.txt only accessible to yourself and inaccessible to anybody else.

4. The average lifetime of a light bulb is 3000 hours with a standard deviation of 696 hours. A simple random sample of 36 bulbs is taken. [2.5X4=10]

a) What is the probability that the average life time in the sample will be between 2670.56 and 2809.76 hours?

b) What is the probability that the average life time in the sample will be equal to or greater than 3219.24 hours?

c) What is the probability that the average life time in the sample will be equal to or less than 3180.96 hours?

d) How large of a sample needs to be taken to provide a 0.01 probability that the average life time in the sample will be equal to or greater than 3219.24 hours?

5. A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-square test, did she get the predicted outcome? **[10]**

Q2. Write the sampling method for the following cases: **[1X3=3]**

a) In an in-store observational study assessing the impact of NLEA-mandated labels on consumers' search for nutrition information, Balasubramanian and Cole positioned observers in the aisles for three product categories in 3 chain grocery stores. Data collection was distributed across the time of day and day of the week.

b) In a study examining longitudinal trends in use of nutrition information among Canadians. Goodman and colleagues used a plus-digit, random-digit dialling process to select the households to take part.

c) All people in US were divided in 4 time zones (Eastern,Central, Mountain,Pacific). 500 people from each time zone were chosen and all these included in sample.

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
Application of Computer and Statistics in Biology BIO G510
FIRST SEMESTER 2016 – 2017
COMPREHENSIVE EXAM

TOTAL WEIGHTAGE 35% Date: 09.12.2016 DURATION: 3Hrs. (Part A & Part B)

Total Marks (46+24) =70

-
- **Answer Part A and Part B in separate answer sheets.**
 - **Irrelevant answer may attract penalty.**
-

PART – B (OPEN BOOK) (Max. duration: 1 Hr., Max. Marks 24)

1. a) Give probable reasons for variation in match/mismatch score in BLOSUM62 scoring matrix. [A= Ala; W= Trp; E=Glu; D=Asp] **[4]**

	A	W	E	D
A	4	-3	-1	-2
W		11	-3	2
E			5	2
D				6

b) For nucleic acid sequences, it is common to use a simple scheme for substitutions, +1 for a match, -1 for a mismatch where as for protein sequences more complicated substitution matrix like BLOSUM62 are used. Explain with reasons. **[2]**

c) From the given multiple sequence alignment of part of DNA sequence of four species construct best phylogenetic tree using NJ method. Show entire analysis. **[8]**

Species-1 ACGCGTTGTGCGATGGCAAC
Species-2 ACGCGTTGTGCGACGGTAAT
Species-3 ACGCATTGAATGATGACAAT
Species-4 ACACATTGAGTGATAATAAT

2. Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university? **[4]**

3. Does SD govern shape of normal distribution curve? Justify your answer. **[2]**

4. How would you relate the variance and sum of squares by a formula? In a study on growth of children, 2 groups were taken. Group 1 had sd of 2.5 and sample size 100 and group 2 had sd of 3 and sample size of 150. Using the formula given below determine Standard error of difference. **[4]**

Combined variance= Sum of sum of squares of both the samples/(n1+n2-2)