

KATHMANDU UNIVERSITY  
End Semester Examination  
August/September, 2017

Mark Scored:

Level : B. Tech.

Course : BIOT 305

Year : III

Semester : II

Exam Roll No. :

Time: 30 min

F. M. : 20

Registration No.:

Date

SEP 10 2017

SECTION "A"

[20 Q × 1 = 20 marks]

1. SRS is a search engine like google for \_\_\_\_\_ database:  
a. Residue                      b. EMBL                      c. DDBJ                      d. SRS
2. The database, which is derived by processing or annotating raw data is:  
a. Prosite    b. Secondary database  
c. SwissProt    d. Specialized database
3. Which of the following describes affine penalty the best?  
a. The gap opening penalty is -2  
b. The gap opening penalty is -2 and the extension penalty is -1  
c. The gap opening penalty is -2 and the gaps to be inserted are 10.  
d. The gap opening penalty is -2, the extension penalty is -1 and gaps to be inserted are 10.
4. SBML is expanded as \_\_\_\_\_.
5. The database and query required by TBLASTN is \_\_\_\_\_ and \_\_\_\_\_, respectively:  
a. Protein, Nucleotide translated                      b. Translated nucleotide: both query and nucleotide  
c. Translated nucleotide, protein                      c. Protein, Protein
6. The default word size of BLASTn is:  
a. 11                                      b. 3                                      c. 8                                      d. 28
7. W, in Clustal alignment is color-coded:  
a. Orange                                      b. Blue                                      c. Green                                      d. Red
8. A statistical model, where each state is dependent upon previous one state is known as:  
a. First order Markov model                                      b. Second order markov model  
c. Hidden Markov model                                      d. Zero order markov model
9. While doing a phylogenetic analysis, it is found that within a clade of prokaryotes, there lie protozoa; of which the possible reason may be horizontal gene transfer. The protozoa and the bacterium are collectively known as:  
a. Orthologs                                      b. Paralogs                                      c. Xenologs                                      d. Ohnologs
10. A group of organisms or genes that include the most recent common ancestor of all of its members and all of the descendants of the most recent common ancestor of all of its members. All of the descendants of that most recent common ancestor is termed as:  
a. Taxon                                      b. Branch                                      c. Clade                                      d. Node

11. In the maximum likelihood method of constructing phylogenetic trees:
- Different models are fitted to single tree and the best model is selected
  - The evolutionary tree is predicted that minimizes the number of steps required to generate variation in sequence.
  - Tree topology with highest probability is generated that signifies our raw data, given a substitution model
  - The consensus data is chosen and the minimum number of models is selected.
12. While doing a phylogenetic analysis, which of the following will not affect our result?
- Assuming each four nucleotides has equal probability of mutation.
  - Shuffling the order of sequence, in which the multiple sequence alignment is evaluated
  - Using the DNA sequences from different databases
  - Using different methods of tree construction.
13. In a peptide chain, phi designates:
- The angle around  $C\alpha - C'$
  - The angle around  $N - C\alpha$
  - The angle around  $C' - N$
  - Rotation around  $C_n - N_{n+1}$
14. In  $\pi$  helix,
- There are 3.6 residues per turn
  - There are 4.1 residues per turn
  - There are 4.6 residues per turn
  - There are 5 residues per turn
15. Chou-Fasman method says that Methionine and Tryptophan tend to attain \_\_\_\_\_ & \_\_\_\_\_ configuration.
- Sheets and Sheets
  - Sheets and coils
  - helix and sheets
  - helix and helix
16. A gene prediction analysis program plots a coding score vs the 3' and 5' splice sites. A curve is drawn that separates the coding and non-coding genes based on the coding score. This analysis is known as:
- Linear regression analysis
  - Logistic discriminant analysis
  - Polynomial discriminant analysis
  - Linear discriminant analysis
17. The following can be a problem when cells are used for amplification of templates
- Non-specific amplification
  - Loss of genetic information
  - Restriction digestion
  - DNA methylation
18. Using only one of the primers, i.e. forward only OR reverse only in SOLEXA sequencing, would lead to produce:
- Amplified template DNA sequence
  - Unamplified template DNA sequence
  - Adapter DNA sequence
  - Terminator sequence
19. In SOLiD sequencing, the total numbers of primer resets are \_\_\_\_\_ in order to interrogate each base \_\_\_\_\_.
- 5 and 2
  - 4 and 1
  - 4 and 2
  - 5 and 1
20. The scores of 454 sequencing are given in \_\_\_\_\_ file format
- ASCII
  - .csfasta
  - .qual
  - .fastq format

KATHMANDU UNIVERSITY  
End Semester Examination  
August/September, 2017

SEP 10 2017

Level : B. Tech.  
Year : III  
Time : 2 hrs. 30 mins.

Course : BIOT 305  
Semester : II  
F. M. : 30

SECTION "B"

Attempt *ALL* questions. Illustrate your answer with figures whenever applicable.

1. What would the following code in R do? **Seqinr** package is similar to **bio3d** that we did in lab. **s2c** converts a **number** array to **character** array, **nrow** lists the length of a row (same for **ncol**), **rownames** assigns row names (same for **colnames**), **seq** returns a sequence which is specified by the first entity till the second entity with intervals of the third entity in the code below. Write the output for matrix **s** and **F**. [4]

```
library(seqinr)
x <- s2c("GAATTAGTA");
y <- s2c("GGATCGA");
d <- 2
s <- matrix(data=NA,nrow=length(y),ncol=length(x))
for (i in 1:(nrow(s)))
for (j in 1:(ncol(s)))
{if (y[i]==x[j]) s[i,j]<- 5 else s[i,j]<- -2 }
rownames(s) <- c(y);
colnames(s) <- c(x)

F <- matrix(data=NA,nrow=(length(y)+1),ncol=(length(x)+1))
rownames(F) <- c("",y);
colnames(F) <- c("",x)
F[,1] <- -seq(0,length(y)*d,d); F[1,] <- -seq(0,length(x)*d,d)
for (i in 2:(nrow(F)))
for (j in 2:(ncol(F)))
{F[i,j] <- max(c(F[i-1,j-1]+s[i-1,j-1],F[i-1,j]-d,F[i,j-1]-d,0))}
```

2. Using FM algorithm of phylogenetic tree construction, draw a phylogenetic tree based on the following distance matrix. [4]

Table 2: Distance matrix for phylogenetic analysis

	Spinach	Rice	Mosquito	Monkey	Human
Spinach	-	84.9	105.6	90.8	86.3
Rice	-	-	117.8	122.4	122.6
Mosquito	-	-	-	84.7	80.8
Monkey	-	-	-	-	3.3
Human	-	-	-	-	-

### SECTION "C"

Attempt *ALL* questions. Answer to the point. Be very specific.

3. How does BLAST return you alignment? What will happen when you increase the neighborhood word threshold? [3]
4. What is Principal Component Analysis (PCA)? What is it used for? Explain the variances accordingly the principal components! [3]
5. What does Naïve Bayes theorem states? How would you perform k-fold cross validation? [4]
6. How are the DNA templates amplified in 454 pyrosequencing? Assume you have the product after emulsion PCR. [2]
7. How are templates sequenced in SOLiD sequencing? [3]
8. What is k-means clustering? [2]

### SECTION "D"

Attempt *ALL* questions. Answer to the point. Be very specific.

9. What is an annotation file? How do you use it? [1]
10. What is a MA plot? What can be inferred from it? [1]
11. What is Euclidean distance? [1]
12. What is bootstrapping? [1]
13. What is a ROC curve? What is it used for? [1]