Subject: bif401 (past papers, objective subjective notes)

CR: Hafiza Mubeen

VU Medical Zone (biotechnologists)

Admins: HAfiza Mubeen, Muhammad Nouman, Tasha kHan, VU Medical zone BIOtechnologists

1) Genomics, evolutionary studies and system biology are application of **bioinformatics**
2) Exact matching require which type of nucleotide-**same number of nucleotide**

3) Accurate solution of protein structure is one of the ___ **toughest problems**

4) **Tract back** is overlap matches can start from any position in scoring matrix

5) The matrix has ……scores. **Positive and negative**

6) **MASCOT**___can search people mass finger printing and shotgun proteomics data set

7) Spectrometer measures the protein by_____ **Mass/charge**

8) Simulation of the folding process depends_____**Energy function**

9) Information about **DNA/RNA i**s available on gene bank

10) **TDP** is used to measure the molecules weight of intact proteins.

11) Coordinates of **alphacarbon** in the protein backbone can be used for structural visualization.

12) in structure visualization c-alpha atoms are traced to recreate a _____**3Dproteinstructure**

13) in UPGMA distance is calculated btw _____ **twoclustersandbetweentwotrees**

14) pk value of Aspartic acid **3.9**

15) Largest amino acid ……**Tryptophan**

16) 2nd STRUCTURE of proteins can be obtained by using_____ **ChouFasmanalgorithm**

17) NJ Algorithm strategy is used to predict __**RNA2`structure**

18) The loop of hair pin must be atleast___ in length **2aminoacids**

19) Fold recognition is also called …….**threading**

20) Ab initio method in contrast ,base their prediction on **lowenergymodel**.

21) Bioinformatics require…….smart mind, and connected to internet, **both**,

22) Sequence alignment tool is…….**PROSIGHTandMASCOT**

23) MS1,MS2 provide us data identifying unknown…….**Proteinssequence**.

24) **MGF** files develop as an open standard for……. **proteomicsdata**.

25) **PDB** coordinates alpha carbon in protein back bone can be used for comparison .

26) Alpha carbon atom can be obtained from……**PDB**

27) in formation of 2nd structure of protein C & N can make……**Hydrogenbonds**

28) information for protein folding is into its native structure is in…….. **protein'saminoacidsequence**!

29) Energy **released** during bond formation.

30) hydrogen bonding occur in ...**secondary** ..Protein structure

31) Bond between C and H in alpha helices____**Hydrogenbonding**

32) Amino acid having 3 codons_____**isoleucine**

33) MALDI typically adds a …… to protein or peptide. **Proton**

34) low quality match gets_____**lowerscore**

35) ____ are sequences of amino acids produced during MS2. **PST** (**peptidesequencetags**)

36) MS2 data can be used to extract…… **peptidesequencetags**.

37) The alpha helices propensity should be more than____(**1.0**)

38) How many types are of protein sequencing techniques__ (**2**)

39) Blast can find sequences of _____**nucleotidesaminoacid**

40) If a protein sequence of 26 amino acids is fragmented at 11 amino acid**C11**, **Z25**

41) To find out unknown sequence of nucleotide we use . **NGS or Mass spectrometry**

42) How many types of peptide May be injected in mass spectrometry chamber.. . **three hundredthousand to four hundred thousand**

43) .**MASCOT**.can search peptide mass finger printing and shotgun proteomics dataset.

44) Stabilizing and destabilizing energy give us **…….. quality of 2nd structure** .

45) The Fasman algorithm is based on **statistical occurrence** ofAmino Acids in known structures.

46) DP recombinant the nucleotide recombination through process of ……. **Traceback**

47) **MALDI**.....add proton to a protein and a peptide..

48) Positively charged amino acids are..... **3**

49) Aromatic amino acid include ……. . **phenylalanine, tyrosine tryptophan**

50) Homology modelling fail to predict …… **quality structure**

51) How many forces are involved in protein folding. **4**

52) Pairwise alignments tells the similarity between sequence**... by maximizing the matches.**

53) Dot plots employ dot matrix with two sequence plotted represents……**pairwise alignment andcomparison**.

54) Differentiate b/w DNA and RNA sequence…… **RNA has more variety of sequence**

55) Cell molecule are produced after transformation of …… **DNA to protein**

56) Dot plot cannot deal with……. **Insertion, deletion and gaps.**

57) Exons are may be more …… **conserved**.

58) In dynamic algorithm we can do comparison …… of more sequence. **Three**

59) Mass/Charge ratio helps calculate the mass of the ….**Protein**

60) In scale tree branches lengths are equal to themagnitude of change in the ……. **nodes**.

61) How many types of peptides mix in MS chamber……. **300,000 – 400,000**

62) mRNA is a …… structure. **Planer**

63) When unpaired bases of 2´ stricture join to form……. **3´ structure**

64) N Jackobson use to pridict **2´structure.**

65) Aromatic amino acid ……. **Tyrosine**

66) Positive charge amino acids……. **Lysine, arginine, histidin**

67) MS begins with the measure of ……. **intact protein**

68) Amount of amino acids in alpha helix…….(**4**)

69) Intact mass of protein can be found by….. **MS1**

70) MSA can be done by……. **CLUSTAL**

71) Which is used as sequence alignment tool……. **PROSIT**

72) MS1 and MS2 help to identify ……. (**PST**)

73) MGF is used for …….. **proteomics**

74) X,Y,Z positions of alpha carbon available online on …… (**PDB**)

75) Method to determine RNA structure …….. **NMR, X-Ray crystallography**

76) CLUSTLA runs……. **Slow accurate/ fast appropriate**

77) Which give info about precursor protein……. **PSTS**

78) Raw data files format can be converted to ……. **Open format**

79) Which of the following is cartoonic figure……. **Ribbon diagram**

80) In CATH protein recognize according to their…….. **structural similarity**

81) Method for obtaining 1´ structure…….. **(Edman Degradation Tandem Mass Spectrometry).**

Explain the relationship between Bioinformatics and modern world Biological sciences.[2]

Bioinformatics gettingfamousbecause of interdisciplinary field it connects many biological sciences like maths, biology, evolution, genetics and molecular biology. It is rapidly growing field in the biological world. Bioinformatics deals digitalized biological information and data obtained from experiments it used math's techniques like algorithms for representation of biological data. Modern Biological science bioinformatics used for different Purpose like writing software and tools for data base and finding the evolutionary relationship of organisms using different techniques of Bioinformatics.

In Modern Biological Sciences Bioinformatics is used for following activities

- Developing Algorithms
- Creating Software's and tools for data analysis
- To determine the Role of Biomolecules
- Evaluations of Data

Explain the applications of Bioinformatics. [2]

Use of Bioinformatics is very complex it is used almost in every branch of biology most important applications are under the following Headings.

- **Genomics**
  - Arranges DNA sequencing by Markers.
  - Nucleotides alignments.
  - Transcribe gene
  - Gene Assembly
  - Gene Finding
  - Data base can be created by the data
- **Proteomics.**
  - Decoding Proteins sequences sounderstanding become easy about protein structure.

- Post translational changes in proteins.
- Protein-Protein Interaction

- **Evolution**
  - Phylogenic tree
- **Systems Biology**
  - Helps development in drugs and their effects on body functions

Application of Bioinformatics in Sciences of Bioinformatics

1. **Molecular Medicine**
   - Human genome will have profound effects on the fielss of biomedical research and clinical medicine.
   - Every infection has ainheritedelement stress which reasonsmodifications in the genome e.g cancers, heart disease and diabetes.
2. **Microbial Genome Applications includes the following**
   - Waste cleaning
   - Climate variation studies
   - Different energy sources
   - Biotechnology
   - Antibiotic Resistance
   - Forensic analysis of microbes
   - Bio-Weapon Creation
   - Evolutionary Studies
3. **Agriculture**
   - The sequencing of the genomes of plants and animals must have hugewelfares for the agricultural community and can be used forexploration for the genes within these genomes and their functions, making them improved, further disease resistant and more industrious.
   - Crops
   - Insect resistance
   - Improve nutritional quality
   - Grow in worse soils and drought resistant
4. **Animals**

- Sequencing developments of many domestic animals including cows, goats and sheep are now well under way in the courage that a superior understanding of the biology of these organisms will have vasteffects for civilizing the manufacture and health of livestock and eventually have advantages for human nutrition.

5. **Comparative Studies**
   - Investigating and comparing the genetic material of diverse species is ansignificant method for reviewing functions of genes, mechanisms of hereditary diseases and species evolution.
   - Bioinformatics tackles can be used to markvaluationsamong the numbers, positions and biochemical functions of gene in diverse organisms.
   - Organisms that are appropriate for use in experimental investigation are called Model organisms

Explain different databases and their Specificity [2]

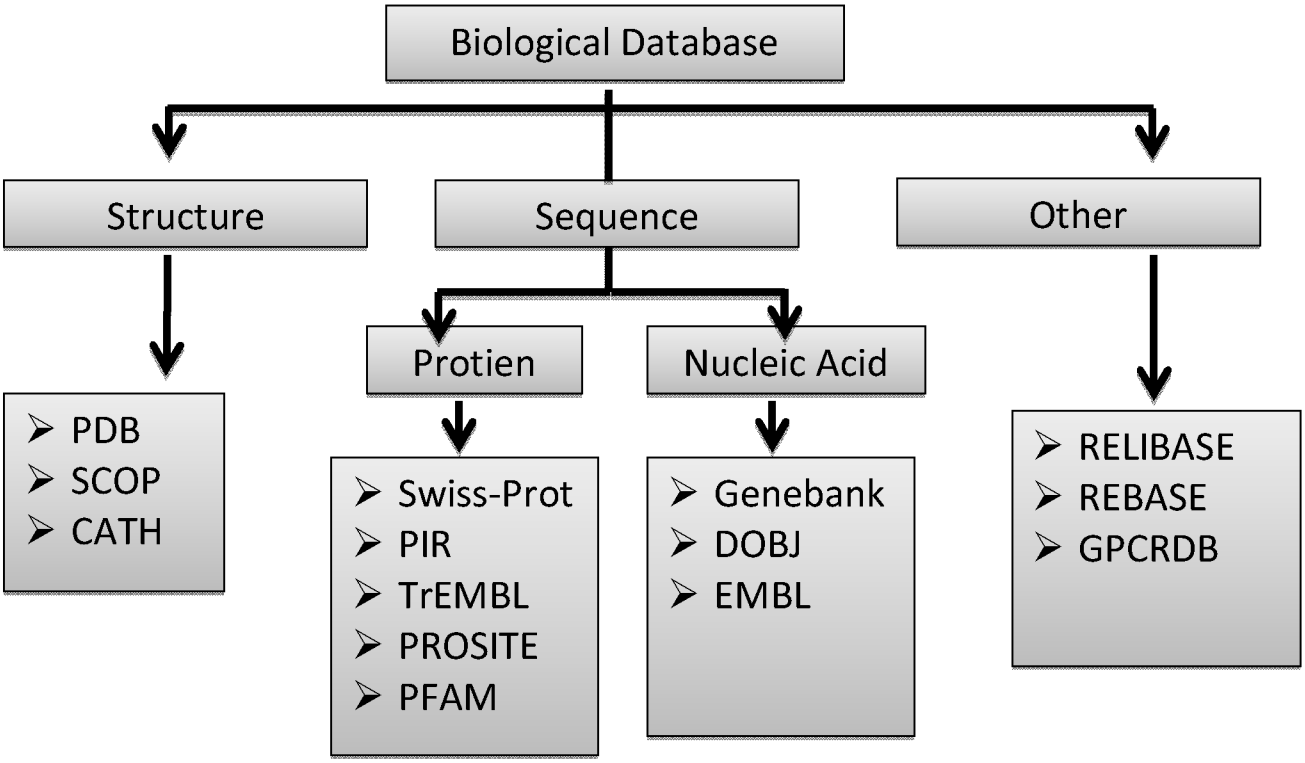
Biological databases are public library of life disciplinesinfo, collected from scientific experiments, published literature, high through put experiment technology, and computational analysis. The data base is subdivided into Primary and Secondary Data base.

- **Primary Biological Databases**

  A primary biological database gives information about sequences or structure information alone.
  - Primary Nucleotide Sequence Database: GeneBank- EMBL, -DDBJ
  - Primary Protein Sequence Database: PIR-PSD, Swiss-prot, TrEMBL.
  - Primary Structure database: PDB
- **Secondary Structure Databases**

  Secondary structure database gives information on classification of proteins based on their structure.
  - Secondary structure database: CATH, -SCOP, -DSSP, -FSSP-DALI.
  - Nucleotide sequence database: GeneBank
  - Protein sequence database: Swisspro

Specificity can be explained by this diagram.



Explain different phylogenetic analysis algorithms. [2]

Phylogentics means to show the evolutionary relationship between the organisms we can obtains this knowledge by comparing difference sequences by creating Rooted and Unrooted Trees that show Phylogentics relationships.

Phylogenic Tree Basics

- Leaves represent things (genes, individuals, strains, species) being compared. Word taxon is used to denote to this.
- Inner nodes are assumedfamily units
- In a rooted tree, track from root denotes an evolutionary pathway (root denotes the communal ancestor)
- An Unrooted tree statesinteractionsbetween things, but no evolutionary path.

Phylogeny Tree Algorithms

- Distance-Based (UPGMA, -Neighbor Join)
- Maximum Parsimony (Character-Based)
- Maximum Probability (Character-Based)

Distance VS. Character State Methods

- **Distance method**
  - UPGMA, Neighbor Joining, Minimum Evolution.
  - Max. Involvesspace measures among sequences.
- **Character State Method**
  - Parsimony, -Max, -Likelihood.
  - Needs discrete characters

UPGMA

- Unweighted pair group method with Arithmetic Mean.
- Accept a molecular timer (constant evolution ratio).
- Create a rooted Tree. Ultra metricstate: for any three taxa (a,b,c),d_{ac}=max (d_{ab}, d_{bc})

Neighbor-Join Method

- Do not assume molecular clock
- Assume additivity of distance matrix (ideal)
- Work for non-addictively of distance matrix (non-ideal situation)
- Most reliable when the branch lengths of trees are allowed to vary
- Goal is to find a tree that minimize the square errors of pairwise distances

Maximum Likelihood Method

- Use probability calculations to find a tree that best accounts for variation in a set of sequences.
- Analysis is performed on each column of a multiple sequence alignment.
- All probable trees are measured. Only can be used for a minor number of sequences

Explain genome of HBV. [2]

The genome of hepatitis B infection is - RT, includes a solitary particle of roundabout and isn't portioned. The genome is to some degree twofold stranded DNA that structures a covalently shut hover with 5' end of the full length short strand which is connected to the viral DNA polymerase. The entire genome is 3020-3320 nucleotides in length (for the total length strand) or 1700-2800 nucleotides in length (for the little length strand). The genome has a guanine + cytosine substance of 48 %. The genome grouping has ends with firm finishes that match the particularly found 5'- closures of the two strands which cover by around 240 nucleotides and keep up the roundabout design of the DNA. Corresponding to the viral mRNA, the negative-sense or non-coding strand is full length; the viral mRNA, the positive sense strand, is shorter than full-length. The twofold stranded genome has a scratch at an interesting site on full length negative strand inverse at a position 242 nucleotides downstream from the 5' end of the positive strand. The 5' end of the negative-sense strand has a covalently appended terminal protein; positive sense strand has a 5' topped oligoribonucleotide pinner. The 3'- end has moderated nucleotide arrangements; of 1 nucleotide long

What is scope of bioinformatics?

Use of computational algorithms and techniques for: Storage, Organization, Analysis, and Representation of biological information

Mention some activities of Bioinformatics.

1. Developing algorithms
2. Writing software
3. Data processing pipelines
4. Statistical evaluation of data
5. Data visualization

Write names of three instruments used for obtaining digitalized biological information on genes and proteins?

1. Next Generation Sequencers
2. High Resolution Mass Spectrometers
3. Nuclear Magnetic Resonance Spectroscopy

Where can bioinformatics be applied specifically?

Genomics, Transcriptomics , Proteomics, Metabolomics, Structural Proteomics Drug Design, Systems Biology, Medicine

what does century of bio informatics holds for us?

An era of personalized medicines and eradication of several common diseases! U, G, C T

DNA and RNA molecules are built up from four types of molecules, each called_____?
Nucleotide

Write names of purines and pyrimidine bases.

Adenine and guanine collectively called purines and cutosine uracil and tyamine are called pyrimidines.

Write names of public databases for DNA/RNa and protein sequences

For DNA/RNA public sequence database is gene bank(by NIH) and for proteins it is Uniport (by uniport consortium)

How to find sequence from gene Bank?

To find any sequence we go online to NCBI GenBank website..... which is : www.ncbi.nlm.gov//genbank

What is Swiss Prot?

In Uniprot home page there is a box named Swiss Prot.**Swiss-Prot contains human curated protein information and annotations**

- Accession number, unique identifier
- Sequence, Molecular mass
- Observed and predicted modifications

Sequences can differ from each other in how many ways?

Sequences can differ from each other in three different ways:

- Substitutions A**C**GA → A**G**GA
- Insertions ACGA → AC**C**GA
- Deletions A**C**GA → AGA

What is benefit of comparing sequences?

By comparing sequences, we can get:

- Similarity
- Specific Differences
- Relationship
- Evolutionary Insight..... About genomic sequence

What is MAP kinase?

MAPKs are involved in directing cellular responses to a diverse array of stimuli, such as mitogens, osmotic stress, heat shock and proinflammatory cytokines. They regulate cell functions including proliferation, gene expression, differentiation, mitosis, cell survival, and apoptosis

What is PROSITE PATTERN?

Ans ..... this is a specific pattern for accommodation of in exact matcing.

PROSITE Pattern for MAP Kinase is F-x(10)-R-E-x(72,86)-R-D-x-K-x(9)-C

WHAT IS THE DIFFERENCE BETWEEN BASIC AND ACIDIC AMINO ACIDS?

the difference is in their side chains, the acidic ones have acidic side chain while basic one have basic side chain at neutral PH.

- Arginine, lysine and histidine are the basic ones.
- Aspartate and glutamate are the acidic ones.

Definealighnment.

The process of inexact matching while keeping in view the conserved residues is called sequence "Alignment"

What is dot plot?

- Dot plots employ dot matrix with two sequences plotted on top and left of the matrix
- Matches are represented by dots
- Dots on diagonals are connected and represent alignments

What is formula for identity?

Identity = Num. of matches/Smaller Legth * 100

What are the approaches to aligning sequences?

Two approaches are:

i. **Global Alignment**
ii. **Local Alignment**

.What is difference between global and local alighnment?

....

GLOBAL ALIGHNMENT	LOCAL ALIGHNMENT
• Maximizing sequence matches over the	• Find regions between two sequences

entire length of two sequences, by introducing gaps.	that have the strongest similarity, excluding less similar regions.
• Used to determine overall similarity, conservation	• Finding similar domains, motifs, detecting distant homology
• maximizes the number of matches between the query and source sequences along the entire length of both the sequences.	• Local alignment - gives the highest scoring local match between both query and sequences.

.What is order of allogrithms?

.... Order of allogrithm is that in how many ways a pair of sequence can be aligned.

.How to calculate order of allogrithm?

a. Comparing two sequences of length 'n' requires n^2
b. Hence its order is $O(n^2)$

.What is dynamic programming?

..... Dynamic programming is an algorithm technique used commonly in sequence analysis.

.Score =?

..... score= #matches + #mismatches + #gaps

.What is traceback?

- ... **Traceback is such that we start from the bottom right and select the element from top, left or diagonal which led to the starting element**

.Which strategy allows us to differentiate between a local and a global alignment

... traceback strategy...

.writeabbrivations for PSA and MSA?

.PSA=pairwise sequence alignment,MSA=multiple sequence alignment

.which software is used for MSA?

:Clustal

.name the types of scoring matrices in pairwise sequence alignment?

There are two types of scoring matrices namely pam matrices and blosum matrices.

.differentitate between pairwise sequence alignment and multiple sequence alignment?

:aligment between two sequences by sliding across each other is known as pair wise sequence alignment while multiple sequence alignment involves the aligmnet of more than 3 sequences or multiple sequences sliding across each other from top to bottom.

.what do you meant by blosum62?

:it means that 62% similarity is concluded in comparison of scoring matrices.

.name two commonly used matrices?

:blosum 62 and pam 120.

.Write the names of algorithms that we use for Local and Global alignment?

: For Local pair wise alignment we use Smith and Waterman algorithms, While for Global alignment we often use Needleman and Wunch algorithms.

.11 Write the basic properties of freshman algorithms that make it different from Needleman.

No option,

:

- Top row and column are set to zero
- Alignment can end anywhere
- Traceback start from highest score

write the function of clustal ? no option

- Creat multiple alignments
- Optimize existing alignments
- Profile analysis
- Creat phylogenetic trees

write types of basic blast? No option

There are two main types of BLAST.

Nucleotides

- **Blastn:** Compares a nucleotide query sequence against a nucleotide database.

Proteins

- **Blastp:** Compares an amino acid query sequence against a protein database.

There are also many other types of BLAST:

- **Blastx:**

Compares a nucleotide query sequence against a protein sequence database.

Helps find potential translation products of unknown nucleotide sequences

- **tblastn:**

Compares a protein query sequence against a nucleotide sequence database

Nucleotide sequence dynamically translated into all reading frames

- **tblastx:**

Compares the six-frame translated proteins of a nucleotide query sequence against the sixframe translated proteins of a nucleotide sequence database.

Write uses of blast? no option

- Blast can be used to search for local alignment of protein and introduction
- Smith waterman can align complete sequence
- BLAST does it in an approximate way
- Henece, BLAST is a fester BUT does not ensum optimal alignment
- Nucleotide sequence
- Its available online

WhAT IS CLUSTAL? NO OPTION

- **Clustal** is a series of widely used computer programs used in Bioinformatics for

What is BLAST ? NO OPTION

- **BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical

significanceIn bioinformatics, **BLAST** for **B**asic **L**ocal **A**lignment **S**earch **T**ool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

- Different types of BLASTs are available according to the query sequences. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence.

What is protemes sets? No option

- A proteome is the set of proteins thought to be expressed by an organism. Uniprot provides proteomes for species with completely sequence genomes .
- Some proteome have been (manually and algorithmically) selected as refrence proteomes. They cover well-studied model organisms and othe organisms of interest for biomedical research and phylogeny.

write two main types of blasts? No option

- Nucleotides; Blastn: compares a nucleotide query sequence against a nucleotide database.
- Proteins; Blastp: Compares an amino acid query sequence against a protein database.

Write the name of basic blasts ? no option

- Blastn(nucleotide BLAST)
- Blastp(PROTEIN BLAST)
- Blastx(translated BLAST)
- Tblastn(Translated BLAST)
- Tblast(translated BLAST)

what is FASTA? No option

- FASTA can perform quick comparison of protein and nucleotide sequence
- Can also perform genome and proteome similarities search
- Itsalso available online like BLAST.

. Why we use FAST Algorithm?

FASTA can search sequence databases and identify unknown sequences by comparing them to the known sequence databases. This can help obtain information on the parent organism, function and evolutionary history.

. What are the steps involved in FASTA Algorithm?

STEP1: Local regions of identity are found

STEP2: Rescore the local regions using PAM or BLOSUM matrix

STEP3: Eliminate short diagonals below a cutoff score

STEP4: Create a gapped alignment in a narrow segment and then perform Smith Watermann alignment

. Write down the types of FASTA?

There are six types of FASTS:

- **fasts35**

Compare unordered peptides to a protein sequence database

- **fastm35**

Compare ordered peptides (or short DNA sequences) to a protein (DNA) sequence database

- **Fasta35**

Scan a protein or DNA sequence library for similar sequences

- **Fastx35**

Compare a translated DNA sequence (6 ORFs) to a protein sequence database

- **tfastx35**

Compare a protein sequence to a DNA sequence database (6 ORFs)

- **fasty35**
  - Compare a DNA sequence (6ORFs) to a protein sequence database

DDBJ & AND EMBL Stand for? No option

- 1= DNA Databank of japan (DDBJ)
- 2= European molecular biology laboratory (EMBL)

Define phylogenetics . no option

- Phylogentics is the science of studying evolutionary relationships.

What are phylogenetic trees? How can we visualize them?

. Phylogenetics involves processing sequence information from different species to find evolutionary relationships.
Output from such studies include Phylogenetic Trees.

Phylogenetics specifies evolutionary relationship with the help of trees. Trees can be rooted or unrooted. Rooted trees can show temporal evolutionary direction.

writetypes of phylogenetic trees ? no option

- **Scaled trees.;** Branch lengths are equal to the magnitude of change in the nodes.
- **Unscaled trees** ; Only representing the relationship between sequences.

what is UPGMA? NO OPTION

UPGMA ; unweighted pair- group method using arithmetic averages . two sequence with the shortest evolutionary distance between them are considred. These sequence will be the last to diverge ,, nd represented by the most recent internal node .

how will u find two values for making clusters or in other words what method u would take to make clusters from sequences?

fst u will find the shortest distance between two sequence from the table of sequences u have given to compute phylogenetic tree.u will see the smallest digit in all columns ,after u have found the smaller distance u will see in which column and row it is located and then u will form cluster after finding these two sequences.

write types of RNA ?

m RNA t RNA r RNA miRNA si RNA

Write two categories of RNA ? no option

- **Coding RNAs**
- Coding RNAs as is obvious from their name , code for proetein
- **Non-coding RNAs**
- Non-coding RNAs regulate/assist in the process of translation.Q

How RNA structure arise ? no option

- Remember that RNAs are composed of sugar, phosphate &nucleotides .
- Each nucleotide is covalenty bonded to the backbone.
- The sequences is then represented as a combinations of A,C,U & G.

write hydrogen bonding between nucleotide bases . no option

- A can make hydrogen bonds with U
- 'G' can make hydrogen bond with ''c''
- 'g' make hydrogen bond with ''u''(wobble pair)

define open reading frame?

In molecular genetics, an **open reading frame** (**ORF**) is the part of a **reading frame** that has the potential to code for a protein or peptide. An **ORF** is a continuous stretch of codons that do not contain a stop codon (usually UAA, UAG or UGA).

write role of RNA strcture ? no option

- DNA information transfer
- Regulatory roles
- Catalytic roles
- Defense & immune response
- Strcture –based special roles

What does role of 5´ cap and 3´ poly A tail?

Ans: the 5´ end of molecule caped with 7 methyl guanosine tri phosphate and 3´ end poly A tail cap, both caps play major role to translate mRNA to ribosome, also protect mRNA.

What is the difference between DNA and RNA helix? No option

RNA helix formed by folding of RNA single strand upon itself, while DNA helix simply formed by pairing of base pairs between its two strands.

How many types of structures involved in making secondary structure? No option

4 types:

1. Helix 2. Loops 3. Bulges 4. Junctions

how can we store reported structure of RNA in data form?

for this purpose we use two online data bases,

1. RNA bricks
2. RMDB

How many steps are involved in proteomics? Write its name?There are four steps are involved in proteomics in MS based.

1. **Separation**
2. **Ionization**
3. **Mass analysis**
4. **Detection**

How many major types of MS – based Proteomics and how they deal?

There are two major types of MS-based proteomicse.g mass spectrometry proteomics,

Bottom up proteomics deals with peptides

Top down proteomics can handle whole proteins

- Bottom up proteomics measures the peptide masses produced after protein enzymatic digestion. And Top down proteomics measures the intact proteins followed by peptides after fragmentation.

How many approaches to perform BUP? Write its name?

There are two approaches to perform BUP.

1 : Peptide Mass Fingerprinting
2 :Shotgun Proteomics

Name experimental method we use to determine RNA structure? No option

a) X-ray crystallography

b) Atomic force microscopy

c) Nuclear magnetic resonance imaging (NMR)

What are mutations in sequences?

Mutations are:

- Replacement of amino acids with certain other amino acids in proteins
- Replacement of nucleotides with certain other nucleotides in DNA/RNA

Explain the relationship between bioinformatics and modern world biology.

"Bioinformatics primarily deals with the digitalized biological information as well as data reported from biological experiments".

In modern biological science, bioinformatics is used for developing algorithms for biological data, processing, statistical evaluation of data and for digitalized visualization of data.

In such way, we can say that,

´Modern world biological science is undergoing rapid change due to onset of BIOINFORMATICS.

Explain applications of bioinformatics.

There are some major applications of bioinformatics:

In this field bioinformatics help us in gene finding, in assemblage of genes and forming of databases.

2.Proteomics:

It helps us to decoding protein sequences for better understanding about protein structure, protein to protein relationship, post translational changes and in generating databases for their sequence and structure.

3.Evolutionary study:

It can make phylogenetic trees to find evolutionary relationship between species, also show ancestry between them,

4.System biology:

Bioinformatics also helps us in regulatory mechanism in genes and proteins. So that we can better understand about regulators and treat them by evaluate drugs.

Bioinformatics also help us in further many aspects like,

Drug development

Waste cleanup

Gene therapy

Preventative medicine ***etc.***

Q.: Explain different databases and their specificity.

There are two basic types of data bases,

Ans: 1. Gene Bank:

It is an online public sequence data base. In Gene Bank we go to online NCBI Gene Bank to find sequence of DNA and RNA using following Website, www.ncbi.nim.nih.gov/genebank

Gene Bank can be research by typing,

1. Sequence name 2. Species 3. ID 4. Locus 5. Author
6. Journal 7. Accession number 8. Name

2.Uniport:

It is also an online public data base. By using Uniport we can find sequence of proteins. If we want to find a protein sequence we must go an online Website given bellow,

www.uniport.org

In home page there is a box named **"Swiss Prot"** which contains human curated protein information, molecular mass, observed and predicted modifications etc.

uniport can be research by typing.

1.Amino acid 2. Name 3. ID 4. Sequence

Further Data Bases and their specificity is mention bellow,

Taxonomic Database ……. For classification

Genomic Database ……. For Gene level information

Q. Explain different phylogenetic analysis algorithms.

Ans:

Phylogenetics:

"Phylogenetics is the study of evolutionary relationship and refers to creation of relationship trees between various species of bacteria, archaea and eukaryote".

Rooted and unrooted trees can be used to show phylogenetic relationship between sequence.

Phylogenetic analysis algorithms:

There are many types of algorithms which are basically classified into two classes,

1.Clustering approaches
2.Objective base method

Clustering ap ıroach	*Objective base method*
UPGMA	LEAST SEQUARE DISTANCE
WPGMA	Maximum Likelihood
Neighbor jo ning	Maximum Parsimony
Single link ıge	
Complete li kage	

UPGMA:

UPGMA is a simple agglomerative method. Which is the abbreviation of `Unweighted pairgroup Method with Arithmetic means`.

In this method two sequences with the shortest evolutionary distance are considered and last to diverge, and represented by the most internal node.

Least Square Distance Method:

It is often used to represent the "Observed" distance between sequence. Applicable branch length.

Q.. Explain genome of HBV.

Ans: HBV is an abbreviation of Hepatitis B Virus. It is belonged to the Hepatdnaviridae family (NCBI taxonomy, ICTV, Viral zone)

Structure:

Structure of HBV virus consist of a small, partially double stranded, relaxed circular DNA of about 3.2 kilo base pairs.

And the Genome is 3020-3320 nucleotide long (for full length strand), and 1700-2800 nucleotide long (for short length strand).

1) **The genes may produce _____ different proteins**
   a. <250,00000
   b. >25,0000000
   c. >250,000
   d. <25,0000000
2) **Humans have over _____ genes**
   b. 30,000
   c. 25,000

d. 35,000
e. 39,000

3) **Data is being accumulated at an _____ rate**
   a. Exponentially increasing
   b. Double increasing
   c. Decreasing
   d. Stable

4) **Modern biological experiments produce data which is stored on _____.**
   a. Internet
   b. Computers
   c. Labs
   d. None of these

5) **Biology easily has _____ years of exciting problems to work on.**
   a. 550
   b. 500
   c. 450
   d. 1000
   e. 850

6) **"Biology easily has 500 years of exciting problems to work on" is saying of_____?**
   a. Donald trump
   b. William Knuth
   c. Donald knuth
   d. William Edwin
   e.

7) **Since genome and _____ information is publically available, you can process this information after downloading it.**
   a. Proteome
   b. Evolutionary studies
   c. System biology
   d. Both b and c

8) **By year 2002 growth of gene bank had risen to _____ base pairs.**
   a. 25 billions
   b. 65 billions
   c. 56 billions
   d. 52 billions

9) **Phylogenetic " tree of life " basically consisits of following three domains:**
   a. Bacteria, archae, monera
   b. Plantae , animilia, fungi
   c. Bacteria, archae, eucarya
   d. Eucarya , viruses, protozoa

10) **Bioinformatics can be specifically applied to genomics through?**
   a. Post translational modification

b. Transcription data
c. Genome assembly
d. Both b and c

11) Bioinformatics can be specifically applied to proteomics through?

a. Protein Sequencing
b. Post-translational Modifications
c. Database Development
d. All of above

12) If we observe the growth of gene bank than from 1l982 it comprised _____ base pairs.

a. 3 billions
b. 2 billions
c. 3 millions
d. 2 millions

13) Bioinformatics can be specifically applied to evolutionary studies through:

a. Evolutionary relationships,Evolutionary distances, Phylogenetics,Tree of life
b. Evolutionary relationships,Evolutionary distances
c. Evolutionary distances, Phylogenetics, Tree of life
d. None of above options is right.

14) Bioinformatics techniques have enabled us to generate _____ data and its utilization!

a. Anic
b. Onic
c. Omic
d. Amic

15) Step by step applications of bioinformatics are expanding from genome level to entire_____?

a. Metabolism
b. Biosphere
c. Biosystem
d. Protein level

16) Bioinformatics helps us to understand the systems from small to big like gene finding to_____?

a. Sequence finding
b. Location finding
c. Function prediction
d. Both a and b

17) _____ integrate biological data

a. Simulation
b. Model
c. Disease prediction
d. Both a and c

18) Simulations help validate testable hypotheses

a. Simulation
b. Model

c. Disease prediction
d. Both a and c

19) Bioinformatics not only organizes, stores and analyzes biological data, but can also validate novel hypotheses

a. Genomic data
b. Proteomic data
c. Biological data
d. Phylogenetic data

20) Recent trends in bioinformatics involve development of personalized therapeutics for_____?

a. Cancer and ulcer
b. Obesity and cancer
c. Diabetes and obesity
d. Cancer and obesity

21) NGS stands for _____?

a. New generating sequence
b. New generation sequencing
c. Next generation sequencing
d. Next genomic sequence

22) Massive amount of biological data is in_____?

a. Megabyte files
b. Terabyte files
c. Gigabyte files
d. All of these

23) Identification of hitherto unknown types as well as roles of RNA is considered as frontier of bioinformatics in_____?

a. Genomics
b. Proteomics
c. Transcriptomics
d. Both a and b

24) Identification of known as well as novel post-translational modifications is considered as frontier of bioinformatics in _____ ?

a. Genomics
b. Proteomics
c. Transcriptomics
d. Both a and b

25) Accurate solution of _____ is one of the toughest problems.

a. Amino acid structure
b. Protein structure
c. Gene structure
d. Chromosome structure

26) _____ century is the century of Bioinformatics

a. 20th
b. 21st
c. 19th
d. 18th

What is scope of bioinformatics?

Use of computational algorithms and techniques for:

1. Storage,
2. Organization,
3. Analysis, and
4. Representation of biological information

Mention some activities of Bioinformatics.

6. Developing algorithms
7. Writing software
8. Data processing pipelines
9. Statistical evaluation of data
10. Data visualization

Write names of three instruments used for obtaining digitalized biological information on genes and proteins?

4. Next Generation Sequencers
5. High Resolution Mass Spectrometers
6. Nuclear Magnetic Resonance Spectroscopy

Where can bioinformatics be applied specifically?

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics
- Structural Proteomics
- Drug Design
- Systems Biology
- Medicine

what does century of bio informatics holds for us?

An era of personalized medicines and eradication of several common diseases!

27) **_____ provides the blueprint for building cells**
a. DNA
b. RNA
c. Proteins
d. Both a and b

28) **_____ dictates the production of cell's proteins, carbohydrates & vitamins**
a. DNA
b. RNA
c. Proteins
d. Both b and c

29) **The transformation of information in DNA into proteins is termed the _____? No options**
Ans.... Central dogma

30) **DNA codes for _____**
a. Gene
b. Protiens
c. RNA
d. Nucleuse

31) **_____ along with some other molecules form cells including their organelles and membrane.**
a. Lipids
b. Nucleotides
c. Proteins
d. DNA

32) **Molecules like proteins and carbohyderate are produced after _____ of DNA into proteins**
a. Transcription
b. Degradation
c. Upgradation
d. Transformation

33) **DNA is constituted by _____ bases**
a. 2
b. 3
c. 4
d. 5

34) **_____ pairs with T only and _____ pairs with G**
a. A ,G
b. A, C
c. C, A
d. G, A

35) **Process of transcription _____ nucleotides bases in the DNA**

a. Encodes

b. Decodes

c. Intercalates

d. None of above options is right

36) Based on the information received after decoding DNA, an RNA molecule is _____.

a. Translated

b. Decoded

c. Encoded

d. Degraded

1) RNA is constituted by four bases. These are _____?

a. A, C,T,G

b. T, G,A,T

c. A,C,U,G

d. U, G, C T

2)**DNA and RNA molecules are built up from four types of molecules, each called_____? No options.**

nucleotide

3)name of pentose sugar present in DNA.

a. 3-deoxyribose

b. 5-deoxyribose

c. 2-deoxyribose

d. 4. Deoxyribose

4)Wrtie names of purines and pyrimidine bases.

adenine and guanine collectively called purines and cutosine uracil and tyamine are called pyrimidines.

5)**Protein and carbohyderates are produced as a result of transformation of RNA molecules. This transformation is called_____**

a. Trascriptomics

b. Translation

c. Transcriprion

d. Replication

6)**Ribosome reads an RNA transcript and _____ amino acids according to that from the cell's cytosol**

a. Arranges

b. Collects

c. Makes

d. Assigns

7)_____ **nucleotides are read at a time from the RNA**

a. 2

b. 3
c. 4
d. 5

8)**Each codon corresponds to a specific _____**

a. **Protein**
b. **Gene**
c. **Amino acid**
d. **Peptide bond**

9)**Translation involves coding of proteins by RNAs at_____?**

a. Cytosol
b. Nucleus
c. Ribosome
d. Endoplasmic reticulum

10) When _____ of amino acids takes, place water is formed .

a. Isomerization
b. Polymerization
c. Hydrogenation
d. Condensation

11) The amino acids are joined with each other with peptide bonds and then fold with each pther in___ form they make protein structure.

a. 2D
b. 3D
c. 4D

12) **RNA sequences can be a little less than the _____ but they are in a much larger variety**

a. Genes
b. Proteins
c. DNA
d. Both b and c

13) **Proteins have long sequences and are _____ in number**

a. Small
b. Very small
c. Very large
d. Large

14) Write names of public databases for DNA/RNa and protein sequences

Ans.... For DNA/RNA public sequence database is genbank(by NIH) and for proteins it is Uniport (by uniprot consortium)

15) Both genbank and Uniport are online databases and DNA, RNA and Protein sequences are available here online for _____

a. Scientists
b. Researchers
c. Public
d. Both b and c

16) _____ is online database where researchers can get access to sequence of DNA. RNA and proteins..... no option

Ans.....genBank

17) How to find sequence from genBank?

Ans ..... to find any sequence we go online to NCBI GenBank website..... which is : www.ncbi.nlm.gov//genbank .....

GenBank can be searched by:

- Sequence
- ID
- Name
- Species etc.

Other information includes:

- Locus
- Accession number
- Authors
- Journal etc.

18) Uniprot is the public database which is being used to search the sequence of_____?

a. DNA
b. RNA
c. Both a and b
d. Proteins

19) Ubiquitin plays an important role in _____ for recycling the proteins

a. Ribosome
b. Nucleuse
c. Cytosol
d. DNa

20) In Uniprot home page there is a box named_____

a. Blast
b. Prot
c. Swiss prot
d. Swiss blast

21) What is Swiss Prot?

Ans... In Uniprot home page there is a box named Swiss Prot.**Swiss-Prot contains human curated protein information and annotations**

- Accession number, unique identifier
- Sequence, Molecular mass
- Observed and predicted modifications

Q.... What is benefit of comparing sequences?

Ans....**By comparing sequences, we can get:**

- Similarity
- Specific Differences
- Relationship
- Evolutionary Insight..... About genomic sequence

22) _____ **include complete matching in terms of residues and their order**

b. **Inexact matcing**
c. **Exact matching**
d. **Both a and b**
e. **Optimum matching**

37) _____ **includes partial matching in terms of residues and order with room for variations in both.**

a. **Inexact matcing**
b. **Exact matching**
c. **Both a and b**
d. **Optimum matching**

38) What is MAP kinase?

Ans..... MAPKs are involved in directing cellular responses to a diverse array of stimuli, such as mitogens, osmotic stress, heat shock and proinflammatory cytokines. They regulate cell functions including proliferation, gene expression, differentiation, mitosis, cell survival, and apoptosis

39) What is PROSITE PATTERN?

Ans ..... this is a specific pattern for accommodation of in exact matcing.

PROSITE Pattern for MAP Kinase is F-x(10)-R-E-x(72,86)-R-D-x-K-x(9)-C

40) In prosite pattern _____ separate the elements of the pattern

a. Latter
b. X
c. Hyphen
d. Dot

41) Bracketed numbers denote

a. Specific amino acid
b. Any amino acid
c. the elements of the pattern
d. the repeat length of a residue,

42) If the repeat length varies, the _____ of this variation is quoted in the brackets.

a. Range
b. Numbers

c. Domain
d. Both a and b

43) How many types of matching?
a. 5
b. 4
c. 3
d. 2

44) **In exact matching includes _____ matching in terms of residues and order with room for variations in both.**
a. Complete
b. Partial
c. Half
d. Impartial

45) **The process of inexact matching while keeping in view the _____ residues is called sequence "Alignment"**
a. **Experimental**
b. **Conserved**
c. **Under observation**
d. **Both a and b**

46) **Pairwise sequence alignment is therefore alignment of a pair of _____ sequences**
a. 2
b. 3
c. 2 or 3
d. More than 3

47) **Pairwise sequence alignment takes into consideration _____ matching**
a. **Exact**
b. **Inexact**
c. **Optimum**
d. **Both b and c**

48) In pairwise sequence alighnment _____ are coloured.
a. Mismatches
b. Indels
c. Insertions
d. Matches

49) In pairwise sequence alighnment _____ are denoted by '.'
a. Matching nucleotides
b. Mismatching nucleotides
c. Substituted nucleotides
d. None of above

50) Define alighnment.

Ans.... **The process of inexact matching while keeping in view the conserved residues is called sequence "Alignment"**

51) _____ **could be inserted to account for insertions and deletions**
   a. gaps
   b. hyphens
   c. dots
   d. both a and b

52) **Gaps ('.') may carry a penalty for _____ scores from unreasonable alignments**
   a. Reducing
   b. Increasing
   c. Equalizing
   d. Increasing as well as equalizing

53) How many types of pairwise alighnment?
   a. 4
   b. 3
   c. 2
   d. 1

54) The alighnment with higest score is_____?
   a. Global
   b. Local
   c. Optimal
   d. Both a and b

55) _____ maximizes the number of matches between the query and source sequences along the entire length of both the sequences
   a. Global alighnment
   b. Local alighnment
   c. Optimal alighnment
   d. Both a and b

56) _____ includes **Finding similar domains, motifs, detecting distant homology**
   a. Global alighnment
   b. Local alighnment
   c. Optimal alighnment
   d. Both a and b

57) **Gaps are inserted in a sequence being aligned to _____ missing amino acids or nucleotides**
   a. Add
   b. Exclude
   c. Accommodate
   d. Delete

58) Sequences can differ from each other in how many ways?

Ans..... Sequences can differ from each other in three different ways:

- Substitutions ACGA → AGGA
- Insertions ACGA → ACCGA
- Deletions ACGA → AGA

59) **_____ lead to Gaps in alignments**

a. Mismatches
b. Matches
c. Indels
d. Substitutions

60) **Sequences are written on _____ side of a dot matrix grid**

a. Top and right
b. Top and left
c. Bottom and left
d. In middle

61) What is dot plot?

- Dot plots employ dot matrix with two sequences plotted on top and left of the matrix
- Matches are represented by dots
- Dots on diagonals are connected and represent alignments

62) **_____ is the number of nucleotides or amino acids which match exactly between two biological sequences"**

a. Similarity
b. Dissimilarity
c. Identity
d. In some cases identity and in some cases similarity

63) What is formula for identity?

Identity = Num. of matches/Smaller Legth * 100

64) **For computing similarity between sequences, you will need to first _____ the two sequences using pairwise sequence alignment!**

a. Copy
b. Remove the gaps
c. Alighn
d. Remove the shorter sequence of

65) **_____ is the comparison between sequences calculated by using alignment approach**

a. Similarity
b. Identity
c. Global alighnment
d. Local alighnment

66) **What are the approaches to aligning sequences?**

Two approaches are:

1)**Global Alignment**

2)**Local Alignment**

67) What is difference between global and local alighnment?

.

GLOBAL ALIGHNMENT	LOCAL ALIGHNMENT
• Maximizing sequence matches over the entire length of two sequences, by introducing gaps.	• Find regions between two sequences that have the strongest similarity, excluding less similar regions.
• Used to determine overall similarity, conservation	• Finding similar domains, motifs, detecting distant homology
• maximizes the number of matches between the query and source sequences along the entire length of both the sequences.	• Local alignment - gives the highest scoring local match between both query and sequences.

1)**Non-diagonal broken dots are _____ matches**

a. Exact
b. Mismatches
c. Random matches
d. Non random

2)**Aligned portions of sequence can be considered in varying orders and the process is called _____?**

a. Domain alignment
b. Domain shuffling
c. Domain alteration
d. Domain consideration

3)**Local alignments have the power to detect small regions of high _____ between two sequences**

a. **Identity**
b. **Similarity**
c. **Consideration**
d. **Alignment**

4)**_____ can be elicited from proteins belonging to different families**

a. Query domain
b. Conserved domain
c. Alternating domain
d. Experimental domain

5)**Alignment of insertions and deletions can be performed by insertion of _____ in either the template or the target sequence**

a. Matches
b. Mismatches

c. Gaps

6)_____ **were addition or removal of amino acids from protein sequences OR nucleotides from DNA or RNA sequences**

a. Matches
b. Mismatches
c. Panalties
d. Indels.

68) What are mutations in sequences?

Mutations are:

- Replacement of amino acids with certain other amino acids in proteins
- Replacement of nucleotides with certain other nucleotides in DNA/RNA

69) **No _____ is inserted in template or target**

a. Indel
b. Matrix
c. Gap
d. Mismatch

70) Mutations are treated with _____ penalties

a. Gap penalty
b. Substitution penalty
c. Varying penalty
d. Point penalty

71) _____ are treated using gaps and gap penalties

a. Indels
b. Mutations
c. Substitutions
d. Alighnments

72) To perform an alignment by sliding sequences across each other, we used _____ to find matching nucleotides and amino acids

a. Grid
b. Matrix
c. Dot plot
d. Both b and c

73) Matches are labelled by _____ instead of a dot

a. 0
b. +1
c. -1
d. i

74) _____ can be included in the dot plot as -1

a. Indels
b. Substitutions
c. Both a and b

d. None of above

75) What is order of allogrithms?

Ans.... Order of allogrithm is that in how many ways a pair of sequence can be aligned.

76) How to calculate order of allogrithm?

a. Comparing two sequences of length 'n' requires n^2
b. Hence its order is $O(n^2)$

77) Calculation steps determine the time taken by an _____?

a. Alighnment
b. Computation
c. Algorithm
d. Both a and c

78) What is dynamic programming?

Ans..... Dynamic programming is an algorithm technique used commonly in sequence analysis.

79) _____ is a costly process.

a. Extracting DNA
b. Extraction of RNA
c. Web programming
d. Sequence analysis

80) _____ helps to reduce costs of sequence analysis..... no options

Ans is dynamic programming

81) DP stands for_____?

a. Desighning primer
b. Dynamic pathadology
c. Dynamic programming
d. Dynamic progression

82) DP uses a _____ to deal with matches, mismatches and gaps!

a. Scoring functionality
b. Scoring function
c. Scoring penalty
d. Possibility function

83) Alignments are represented by _____ in the dot matrix plot

a. Vertical components
b. Horizontal components
c. Diagonals
d. Dots

84) Score =?

Ans..... score= #matches + #mismatches + #gaps

85) Which cell is diagonal to any given cell i.e (i , j)

a. (i, j-1)
b. (i-1, j)

c. (i-1, j-1)
d. (1-i, j)

86) Which option is correct is there is gap on y-indices

a. (i, j-1)
b. (i-1, j)
c. (i-1, j-1)
d. (1-i, j)

87) Which option is correct is there is gap on x-indices

a. (i, j-1)
b. (i-1, j)
c. (i-1, j-1)
d. (1-i, j)

88) In NEEDLEMAN WUNSH ALOGRITHM _____ is match score.

a. Alpha
b. Beta
c. Gamma
d. None of these

89) In NEEDLEMAN WUNSH ALOGRITHM _____ is mis-match penalty

a. Alpha
b. Beta
c. Gamma
d. None of these

90) In NEEDLEMAN WUNSH ALOGRITHM _____ is gap penalty.

a. Alpha
b. Beta
c. Gamma
d. None of these

91) NEEDLEMAN WUNSH ALOGRITHM finds optimal combination of alighned_____.

a. Left position eliment
b. Upper position element
c. Left position element
d. Diagonals

92) Acording to NEEDLEMAN WUNSH ALOGRITHM diagonal move is a _____?

a. Match
b. Mis match
c. Gap
d. mutation

93) After completely calculating the matrix, we need to do _____

a. Alighnment
b. Programming
c. Traceback
d. Computing values

94) What is traceback?

- Ans... **Traceback is such that we start from the bottom right and select the element from top, left or diagonal which led to the starting element**

95) Traceback allows us to extract the _____ alignment

a. Global
b. Local
c. Optimal
d. Both a and b

96) Which strategy allows us to differentiate between a local and a global alignment

Ans... traceback strategy...

1.Blosum matrics are used to align:

a.DNA sequences b.protein sequences c.amino acids sequences d.none

2.pair wise sequence alignment consists comparison of-----sequences:

a.2 b.3 c.4 d.5

3.pair wise alignment may be:

a.local b.global c.somewhat specific d.both a and b

4.dynamic programming cannot work after:

a.2 sequences b.4 sequences c.6 sequences d.9 or 10 sequences

5.for multiple sequence alignment we use:

a.dynamic programming b.progressive alignments c.slective alignments
d.none of these

6.write abbrivations for PSA and MSA?

Ans.PSA=pairwise sequence alignment,MSA=multiple sequence alignment

7.which software is used for MSA?

Ans:Clustal

8.MSA involves the comparison of:

a.less than 3 sequences b.more than 3 sequences

c.more than 5 sequences d.more than 10 sequences

9.there are --- types of scoring matrices:

a.2 b.3 c.4 d.5

10.name the types of scoring matrices in pairwise sequence alignment?

There are two types of scoring matrices namely pam matrices and blosum matrices.

11.blosum matrices were first purposed in:

a.1772 b..1862 c.1992 d .1872

12.differentitate between pairwise sequence alignment and multiple sequence alignment?

Ans:aligment between two sequences by sliding across each other is known as pair wise sequence alignment while multiple sequence alignment involves the aligmnet of more than 3 sequences or multiple sequences sliding across each other from top to bottom.

13.gaps are not allowed in:

a.pam sequences b.blosum sequences c.both a and b d.none

14.in computiong of blosum matrices if i=j then it means:

a.simillar amino acids b.different amino acids c.mutational substitutions d.none

15.what do you meant by blosum62?

Ans:it means that 62% similarity is concluded in comparison of scoring matrices.

16.name two commonly used matrices?

Ans:blosum 62 and pam 120.

1=charactersize protein families based on ..... regions

a)=MULTIPLE gene b)= homologousc)= heterzyous d)= none of these

2=Evaluate evolutionary order of species or .....

a)= Phylogenyb)= phasmids c)= bacteria d)= both b and c

3=Pairwise alignment is the alignment of .....

a)= Multiple sequence b)= single sequence c)= two sequenced)= all of these

4=..... can help to perform MSA!

a)= Progressive alignmentb)= multimple alignment c)= both

5=In multiple sequence of Alignment -------- is required to remove sequence

a)= 30% b)= 40% c)= 57% d)=80%

6= pairwise alignment include the sequence which is …

a)=S1 &S2 B)= S3 &S4 C)= All

7=similar materics =……

a)= matches/ comparisons b)= camparisons / matches c)= matches/ frequence
d)= Camparisons/frequence

8=---------- is the free tool which do all the process of MSA for our body .

a)= Gene b)= DNA c)=clustald)=RNA

9=Clustal is developed by ……. Molecular lab

a)= Russians b)= American c)= assians d)=european

10=Clustal perform alignment sequence in

a)= slow and accurate b)=fast/ approximate c)= fast/accurate d)= both a and b

11=write the function of clustal ? no option

- Creat multiple alignments
- Optimize existing alignments
- Profile analysis
- Creat phylogenetic trees

12=Clustalw can use multiple file formats including
a)= EMBL b)= swissprot c)= pearson(fasta) d)=All of these
13=Blast developed in
a)=1980 b)=1970 c)=1990d)1956
14=blast stand for? No option
Basic local alignment search tool
15=Blast is used for…
a)= search databases for query protein and nucleotide sequence
b)= search for translational products
c)= online search to blast.ncbi.nlm.nih.gov/blast.cgi
d)= All of these
16=write name of basic blast? No option

1= Nucleotide blast ;search a nucleotide database using a nucleotide query, Algorithms; blastp, psi-blast,phi blast , delta-blast
2= protein blast ;search protein databases using a protein query , Algorithums; blastp,psi-blast, phi-blast, delta-blast
3= Blastx; search protein database using a translated nucleotide
4= Tblastn; search translated nucleotide database using a protein
5=Tblastx; search translated nucleotide databases using a protein
17=write uses of blast ? no option

- Blast can be used to search for local alignment of protein and introduction
- Smith waterman can align complete sequence
- BLAST does it in an approximate way
- Henece, BLAST is a fester BUT does not ensum optimal alignment
- Nucleotide sequence
- Its available online
- Can perform search across species and organisms .

WhAT IS CLUSTAL? NO OPTION

Clustal is a series of widely used computer programs used in Bioinformatics for multiple sequence alignment.

What is BLAST ? NO OPTION

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significanceIn bioinformatics, **BLAST** for **B**asic **L**ocal **A**lignment **S**earch **T**ool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

Different types of BLASTs are available according to the query sequences. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see if humans carry a similar gene; BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of sequence.

1=---------- can align complete sequence of Blast.

a)= Edward b)= Smith watermanc)= Frenklin

d)= Griffth

2=BLAST use in .....

a)= For approximate sequence matching

b)= Input to BLAST is fASTA formatted sequence

c)= In search of parameters

d)= All of these

3=Output of BLAST is .....

a)=Result are shown in HTML, plan text, and XML formats

b)=Atable list the sequence hits found along with scores

c)= Users can read this table off and evaluate results

d)= All of these

4=What is protemes sets? No option

A proteome is the set of proteins thought to be expressed by an organism. Uniprot provides proteomes for species with completely sequence genomes .

Some proteome have been (manually and algorithmically) selected as refrence proteomes. They cover well-studied model organisms and othe organisms of interest for biomedical research and phylogeny.

5=..... can search sequence databases and identify unknown sequence by comparing them to the known sequence .

a)= BLASTB)= Algorithm BLAST C)= INPUT BLAST

D)= OUTPUT BLAST

6=Blast help in identify the .....

a)= parent organism b)= function & evolutionary history

c)= both a and b

d)= none of these

7= the protein have ..... list of all possible words

a)= PQG(score 15) b)= QGE (Score 9)

C)= GEL(Score 12) & PEG(score 13) d)= both a and c

8= Blast performs searches by ..... alignments on sequence.

a)= Slow b)= quickc)= normal c~)= All

9=write two main types of blasts? No option

Nucleotides; Blastn: compares a nucleotide query sequence against a nucleotide database.

Proteins; Blastp: Compares an amino acid query sequence against a protein database.

10= Compares the six-frame translated proteins of a nucleotide query sequence against the six – frame translated proteins of a nucleotide sequence database. Is called .....

a)= Tblastxb)= Tblastn c)= Blastx d)= none of these

11=Multiple types of BLAST exsit which assist in .....

a)= Aligning nucleotide b)= Amino acid sequence c)= protein d)= both a nd b

12=For known sequence we use ..... website

a)= NCBI B)= UCSC C) BINC D)= Both a nd b

13=For unknown sequences we use .....

A)= NGS or Mass spectrometryb)= UCSC C)= NCBI D)= NONE OF THESE

14= Write the name of basic blasts ? no option

- Blastn (nucleotide BLAST)
- Blastp(PROTEIN BLAST)
- Blastx(translated BLAST)
- Tblastn(Translated BLAST)
- Tblast (translated BLAST)

15= Fasta developed in .....

a)= 1988b)=1984 c)=1989 d)=1982

- 16=what is FASTA? No option
- FASTA can perform quick comparison of protein and nucleotide sequence
- Can also perform genome and proteome similarities search
- Its also available online like BLAST.

1=Structure of FASTA is obtained from.....

a)=X-ray crystallography b)= microscopy & nuclear Spectrocopy

c)= both a & bd)= none of these

2=The famous data store is called .....

a)= uniprot & swiss protb)= blast c)= fasta

d)= none of these

3=---------- contains human curated protein information .

a)= Uniprot b)= Swiss protc)= FASTA D)= BLAST

4=Over ..... protein structures are reported and present in this database .

a)= 80,000 b)= 60,000 c)= 30,000 d)= 50,000

5=Protein sequence contain .....

a)= Mass spectrometry b)= Edman Degradation c)= online website
d)= both a and b

6=Sequence data can be used to obtain

a)= similarity of sequences b)= evolutionary history c)= predict the function of molecules d)= All of these

7=In gene bank public database of nucleotide sequences for over ..... Organisms.

a)= 200,000c)= 300,000 c)= 400,000 d)= 100,000

8=DDBJ & AND EMBL Stand for? No option

1= DNA Databank of japan (DDBJ)

2= European molecular biology laboratory (EMBL)

9=Define phylogenetics . no option

Phylogentics is the science of studying evolutionary relationships.

10= phylogenetics has led to the creation of relationship trees between various species of .....

a)= Bacteria b)= Archaea c)= eukaryote d)= All

11=write two types of phylogenetic trees ? no option

Scaled trees.; Branch lengths are equal to the magnitude of change in the nodes.

Unscaled trees ; Only representing the relationship between sequences.

12= DNA gets modified by .....

a)= mutation & substitution b)= insetion c~)= deletion d)= all of these

13=what is root node? No option

Root node is the ancestor of all other nodes! The direction of evolution is from ancestor to the terminal nodes.

14= rooted and unrooted trees can be used to show .....realtionship between sequence.

a)=phylogeneticb)= genetic c)= genomic d)= none of these

15=what is UPGMA? NO OPTION

UPGMA ; unweighted pair- group method using arithmetic averages . two sequence with the shortest evolutionary distance between them are considred. These sequence will be the last to diverge ,, nd represented by the most recent internal node .

16= least squares distance method.....

a)= D(HUMAN , CHIMP)=0.3 B)= D(HUMAN GORILLA)=0.4

C)= D(CHIMP, GORILLA)=0.5 D= ALL OF THESE

1.UPGMA stands for:

a.unweighted pair_group method using arithmetic averages

b.undifferentiated pair—group method arithmetic averages

c.unweighted pair –general method arithmetic averages

d.unidirectional pair—general method arithmetic averages

2.the distance matrix is obtained by:

a.multiple sequence alignment b.single sequence alignment

c.pair wise sequence alignment d.all

3.UPGMA is used to calculate:

a.maximum linkage b.least square distance

c.maximum likelihood d.maximum parsimony

4. $$d_{XY} = \frac{1}{N_X N_Y} \sum_{i \in X, j \in Y} d_{ij}$$

This formula is to find:

a.distance of clusters b.distance between trees

c.distance between sequences d.distance between cluster and sequence

5.in the above formula I have shown you the N_X is:

a.number of clusters b.number of trees

c.number of sequences d.none

6.component of UPGMA are:

a.2 b.3 c.4 d.5

7. $$d_{XY} = \frac{1}{N_X N_Y} \sum_{i \in X, j \in Y} d_{ij}$$

In d_{zw} the w is:

a.cluster b.new sequence c.mixed cluster d.tree

8.in the second component of UPGMA we add:

a.two clusters b.cluster+sequence

c.two sequnecs d.none of these

9.how will u find two values for making clusters or in other words what method u would take to make clusters from sequences?

Ans:fst u will find the shortest distance between two sequence from the table of sequences u have given to compute phylogenetic tree.u will see the smallest digit in all columns ,after u have found the smaller distance u will see in which column and row it is located and then u will form cluster after finding these two sequences.

10.cluster+sequence will make:

a.sequence b.cluster

c.may be a cluster or sequence d.phylogenetic tree

Lecture 86- 90

1= Application of UPGMA resulted in formation of ..... Sub trees

a)= 2 b)= 4 c)= 6 d)=8

2=Un-weighted pair group method using ..... is a clustering method to construct phylogenetic trees.

a)=Average b)= Airthmetic average c)= frequency d)= ALL

3=In earlier days, RNA was only considred as a passive intermediary between ..... and .....

a)= DNA AND RNA B)= RNA and protein c)= DNA and protein

d)= None of these

4=write types of RNA ?

1= m RNA

2=t RNA

3= r RNA

4= miRNA

5= si RNA

5=More so, that they perfom irtical functions in ,,,,,,,,, and ,,,,, expression

a)= Dna and gene b)= RNA and GENE C)= Gene and protein

d)= Protein and DNA

6=The two hydroxyl groups make RNA less stable than DNA bcoz it''s more prone to .....

A)= Hybirdization b)= Hydrolysis c)= none of these

7=..... has a short life span and is more prone to degradation .

A)= DNA B)= RNA C)= BOTH A &B D)= GENE

8=Diverse form of RNA ...

a)= RNA molecules are the same b)= RNAs differ in nucleotide sequence

c)= RNA have multiple role in cells d)= All of these

9=Write two categories of RNA ? no option

- Coding RNAs
- Coding RNAs as is obvious from their name , code for proetein
- Non-coding RNAs
- Non-coding RNAs regulate/assist in the process of translation.

10=Only ..... of RNA content in a cell

a)= 5-80% B)= 5-10%C)= 5-64% D)= 5-30%

11=..... cARRy genetic information from DNA to ribosomes where proteins are being assembled.

a)= m RNAB)= r RNA C)=t RNA D)= NONE OF THESE

12=THE 5'can helps identify mRNAs at the ,,,,,,,

a)= Chromosome b)= Ribosomesc)= DNA D)=RNA

13=The 3'end of mRNA has a polyA tail around ..... adenylate residues which help shield against 3' exonucleases.

a)= 30-200b)=30-400 c)=20-200 d)=20-300

14=THE 5' -3' regions are both cushioned with.....regions

a)= RTRs b)= UTRsc)= DSAs d) all

15=RNAs can adopt complex .....structures.

a)= 3DB)= 2D C)=4D D)=6D

16=How RNA structure arise ? no option

- Remember that RNAs are composed of sugar, phosphate & nucleotides .
- Each nucleotide is covalenty bonded to the backbone.
- The sequences is then represented as a combinations of A,C,U & G.

17=Nuleotide can make ..... Bonds themselves

a)= carbon b)= oxygen c)= hydrogend)= All

18=write hydrogen bonding between nucleotide bases . no option

- A can make hydrogen bonds with U
- 'G' can make hydrogen bond with ''c''
- 'g' make hydrogen bond with ''u''(wobble pair)

19=write role of RNA strcture ? no option

- DNA information transfer
- Regulatory roles
- Catalytic roles
- Defense & immune response
- Strcture –based special roles

20=..... Is the energy of an RNA molecules available for reaction. RNA have low energy .

a)= kinetic energy b)= potential energy c)= gibbs free energy

d)= ALL

1. If nucleotides could not make pair together there will be formation of.....

a) Helix b) disjunction c) loops d) pseudoknots

2. Loop of hair pin must be at least ..... bases longe

a) 3 b) 4 c) 20 d) 8

3. What is the difference between DNA and RNA helix? No option

Ans: RNA helix formed by folding of RNA single strand upon itself, while DNA helix simply formed by pairing of base pairs between its two strands.

4. ..... include two or more double stranded regions converging to form a closed structure.

a) Bulge b) loop c) junction d) a+b

5. When unpaired bases formed H bonds with other unpaired bases which structure originates

a) Primary b) secondary c) tertiary d) all

6. How many types of structures involved in making secondary structure? No option

Ans. 4 types:

2. Helix 2. Loops 3. Bulges 4. Junctions

7. When RNA molecule fold upto itself within tertiary structure it may result formation of...

a) Double stranded RNA b) stable RNA c) pseudoknots d) loops
8. ……. helps us to detect pseudoknots,
a) Linear plot b) dot plot c) circular plot d) pseudo plot
9. In circular plot …… are represent pseudoknots,
a) Circular arc b) vertical arc c) perpendicular arc d) intersecting arc
10. Name experimental method we use to determine RNA structure? No option
d) X-ray crystallography
e) Atomic force microscopy
f) Nuclear magnetic resonance imaging (NMR)
11. Which is quite expensive procedure to determine the structure of RNA?
a) NMR b) RNA bricks c) x-ray crystallography d) a+b
12. Commonly RNA has extremely …… structure.
a) Stable b) strong c) sturdy d) fragile
13. formation of H bond b/w nucleotides result in release of energy,
a) -7. K cal/mol b) -17.0 K cal/mol c) -12.0 K cal/mol d) -5.0 K cal/mol

14.which work according to principal of diffraction?

a) atomic force microscopy b) x-ray crystallography c) NMR d) A+B

15. how can we store reported structure of RNA in data form?

Ans: for this purpose we use two online data bases,

3. RNA bricks
4. RMDB

1 : Several algorithms exist for predicting ……structure?
A) One B) Two C) Three D) Four

2 : ……implements an ungraded version of Zuker,s Algorithms?

A) MFOLD B) MOFLD C) MFODL D) MOFDL

3: MDOLD ……set of possible structure? NO option?

Ans) Outputs

4: Order…….complexity for sequence of length N?

A) N^2 B) N^3 C) N^4 D) N^5

5: MFOLD used to predict structure of sequences less than ……?

A) 8000 nuclei B) 7000 nucleotides C) 8000 nucleotides D) 8000 nucleolus

6: MFOLD helps fold an……Nucleotides sequences?

A) DNA B) RNA C) mRNA D) SRNA

7: RNA nucleotides fold to form .....structure?

A) Two B) Three C) Four D) Five

8: which occur in wide verity of.....?

A) CUCUGG , RNA B) CUCGG , DNA C) CUUCGG ,DNA D) CUUCGG , RNA

9: CUUCGG mostly form a stable.....loop?

A) D B) hairpin C) G D) T

10: RNA bricks is a database of RNA .....structure motifs?

A) 4D B) 3D C) 2D D) 6D

11: DNA comprises of .....types of nucleotides bases?

A) Two B) Three C) Four D) Five

12: Which is correct about translation?

A) DNA < RNA B) RNA < Protein C) Protein < Modification D) DNA < Protein

13: Which protein chains fold to take 3D structure?

A) H1 B) H2 C)H3 D) All of these

14: Nucleotides codons comprises of.....code for protein?

A) A ,C & G , T B) A , U & G ,C C) A , C & G , C D) A , U & G , T

15: How many start and stop codon? No option?

A) 1 start codon and 3 stop codons

16: ORF is stand for.....? No option?

A) Open Reading Frame

17: How many ORFs exist for each DNA sequences?

A) Four B) Five C) Six D) Seven

18: The valid ORFs will have the longest.....read?

A) RNA B) DNA C) SRNA D) a and b

19: Which stop codon will mark the end of the protein sequences?

A) First B) Second C) Third D) Fourth

20: PhNCS attaches to the free amino group at .....terminal residue?

A) B B) N C) A D) a and b

21: How many amino acid is removed as a PhNCS?

A) One B) Two C) Three D) Four

22: Restricted to chain of .....residues?

A) 40 B) 50 C) 60 D) 70

23: Protein can be charged with .....?

A) Proton B) Electron C) a and b D) Neutron

24: Write formula of Lorentz force law? No option?

A) $F = Q (E + v \times B)$

25: Each protein is deflected in proportion to its.....?

A) Molecules B) Molecular weight C) a and b D) none

26: Give example of protein sequences? No option?

A) Uniprot , Swissprot

27: MS- based proteomics helps sequences .....protein?

A) Larger B) Smaller C) Bigger D) a and c

28: How many steps are involved in proteomics? Write its name?

Ans: There are four steps are involved in proteomics in MS based.

5. **Separation**
6. **Ionization**
7. **Mass analysis**
8. **Detection**

29: How many peptides is selected at a time for sequencing?

A) One B) Two C) Three D) Four

How many major types of MS – based Proteomics and how they deal?

- **There are two major types of MS-based proteomics**
- **Bottom up proteomics deals with peptides**
- **Top down proteomics can handle whole proteins**

1 : Which enzyme are used to cleavage proteins into peptides?

A) Trypsin B) Pepsin C) Rennin D) Ligase

2: Number of .....depend on number of sites at which enzyme are cut?

A) Protein B) Peptides C) Amino acid D) Sequences

3: Each peptide is measured for its mass.....?

A) MS4 B) MS3 C) MS2 D) MS1

4: Each peptide is searched in .....sequences?

A) Amino acid B) MS3 C) Protein D) a and b

5: Bottom up proteomics measure the .....of peptides?

A) Mass B) Length C) Sequences D) protein

6: BUP stand for.....?No Option?

A) Bottom up proteomics

7: How many approaches to perform BUP? Write its name?

- A ns: **There are two approaches to perform BUP**.

3 : Peptide Mass Fingerprinting
4 :Shotgun Proteomics

8: Peptide mass fingerprint followed by .....protein peptide analysis?

A) Single B) Double C) Triple D) a and b

9: Protein carry typically Post.....modification?

A) Transcriptional B) Translational C) Transformation D) a and b

10: After.....,one protein is selected at a time?

A) MS1 B) MS2 C) MS3 D) MS4

11: Peptides are measured for their mass……?

A) MS4 B) MS3 C) MS2 D) MS1

12: Which measure the mass of intact protein?

A) BUP B) TDP C) a and b D) peptides

13:Mass spectrometer are used to measure the molecular weight of……?

A) Protein B) Peptides C) a and b D) None

14: Scoring Schemes are required to…….represent the quality of result?

A) Qualitatively B) Quantitatively C) Protein sequences D) a and b

15: Ionization result in an decrease or increase of the…….mass?

A) Protein B) Peptide C) a and b D) None

16: How many Salient ionization technique include….ionization?

A) One B) Two C) Three D) Four

17: MALDI & ESI stand for……?

A) Matrix Assisted Laser Desorption Ionization&Electro Spray ionization

18: MALDI typically adds a ……to the protein or peptides?

A) Electron B) Proton C) Neutron D) a and b

19: The molecular weight of the molecule increase by……?

A) One B) Two C) Three D) Number of proton added

20: MS reports the molecule at ……?

A) Mass + 1 B) Mass + 2 C) Mass + 3 D) Mass + 4

21: In which difficult to find molecules with +1 charge?

A) MALDI B) ESI C) MS D) MS1

22: which data from MALDI ionization is easier to handle as the product ions?

A) ESI B) MS C) MS1 D) MS3

Q.1 DNA coding regions are called

a) Intron b) Exons c) Codons d) a and b

Q.2 In local alignment we actually find?

a) Functional subunits

b) Non functional subunits

c) Gene location

d) position of introns

Q.3 ……. Are more likely involved in mutations

a) Exons b) Introns c) Nucleotides d) Amino acids

Q.4 Write the names of algorithms that we use for Local and Global alignment?

Ans: For Local pair wise alignment we use Smith and Waterman algorithms, While for Global alignment we often use Needleman and Wunch algorithms.

Q.5 Difference between needle man and waterman is that ……. is placed in relationship.

a) +1 b) -1 c) 0 d) i, j

Q.6 Which statement is correct about Exons?

a) exons are only coding regions

b) exons are noncoding regions

c) exons can never be found in noncoding state

d) exons can be found in noncoding state

Q.7 How many types of alignments are there? Mention their names.

Ans: There are two types of alignments

1) Optimal alignments
2) Best alignments

Q.8 In Smith freshman algorithms traceback starts from

a) left b) right c) highest score d) lowest score

Q.9 Traceback help us to find

a) single aligned regions b) double aligned region

c) opposite aligned region d) multiple aligned regio

Q.10 Which threshold we use for matching?

a) "L" b) "T" c) "M" d) "O"

Q.11 Write the basic properties of freshman algorithms that make it different from Needleman.

- Top row and column are set to zero
- Alignment can end anywhere
- Traceback start from highest score

Q.12 Scoring scheme often use to produce

a) Best alignment b) Sequence alignment

c) Optimal alignment c) a and c