BIF501 – BIOINFORMATICS-II

ALL OBJECTIVEs & SUBJECTIVEs FROM PAST FILES FOR MID TERM

[redacted] | : ARHAM (+923351328979) | : AIZA WRITES

OBJECTIVEs

1. ORF stands for ---------**Open Reading Frame**
2. IUB stands for ---------**International Union of Biochemistry**
3. PAM stands for ---------**Pointed accepted mutation**
4. FTP stands for---------**File Transfer Protocol**
5. LINE stands for ---------**Long Interspersed Nuclear Elements**
6. SINE stands for---------**Short Interspersed Nuclear Elements**
7. PIR stands for---------**Protein information resource**
8. HMM stands for---------**Hidden Markov Model**
9. GCG stands for---------**Genetic computer group**
10. TIGR stands for---------**The institute of genome research**
11. XML stands for---------**Extensible Markup Language**
12. PDGF stands for---------**Platelet derived growth factor**
13. NCBI stands for---------**National Center for Biotechnology Information**
14. EMBL stands for---------**European Molecular Biology Lab**
15. DDBJ stands for---------**DNA Data Bank of Japan**
16. GEO stands for---------**Gene Expression Omnibus**
17. STS stands for---------**Sequence Tag Sites**
18. BLOSUM stands for---------**Block substitution Matrix**
19. PSSM stands for---------**position specific scoring matrices**
20. Model plant is. ---------**Arabidoposis thaliana**
21. human genome project started in---------**1886**
22. Endosymbiots hypothesis purposed in---------**1883**
23. SINES are... Bp long ---------**80-300BP**
24. RNA is stranded molecule ---------**Single**
25. In order to predict exon we need to get --------- **ORF or Splice Site**
26. Bypass is the type of algorithm____ **Branch and bound**
27. Median string problem is one of the way to find------- **Motifs**
28. Aminoacids in MS2 genes------- **129 Amino acids**
29. MS2 genome has codons ------ **49**
30. The accuracy of GENSCAN decrease__for genes with many_**short_exons**
31. Paralogs genes are originated from -------- **Gene duplication**
32. Human mitochondria_ kb----------**17Kb**
33. Waardenburg's syndrome Gene present on chromosome ------- **2**
34. Genes present in different organisms ------- **Horizontal Transfer**
35. Motif facilitate for binding.------ **NF-kB sites**
36. Uniport finds---- **Protein sequences**
37. AG and GT are----Region ----- **Exon flanking**
38. Pre order algorithm is a-------**Recursive**

39. -------% human genome consist of gene consecutive----- **3%**
40. Three leotides code for---- Amino acid----**1**
41. Genebank is a _____database-----**DNA**
42. H.Influenza was the published genome----**1st**
43. Gene mapping is determining the-------on Chromosome----**-Location**
44. Which RNA is converted in to proteins-------**mRNA**
45. In RNA thymine is replaced by---------**Uracil**
46. Regulatory Motifs are generally found in --------**Promoter region of DNA**
47. Cystic fibrosis gene is present on chromosomes no ---------- **7**
48. A drug can target about ------ Proteins ---------- **500**
49. Orthologs genes have same ---------- **Function and Evolutional history.**
50. Median string is used to find-------**-Motif**
51. EMBL begins with_-----**seqeunce/ ID**
52. Less edit graph means ------- **Relation is closely related**
53. Restriction enzymes break DNA molecules at -------- **specific sites**
54. Brute force approach is good for _____data. ------ **Small data**
55. Insertion in an alignmnet shows.....in edit graph-------- **space in the top**
56. Each common subsequence belong to an alignment that ------- **no mismatches**
57. Symbols of the pattern and the searched text are chosen from a predetermined finite set, called---------- **alphabet**
58. Next vertex algorithm provides as an output ---------- **Next subtree**
59. simple motif search algorithm is based on --------- **Brute force**
60. Consensus score is defined as ---------- **score(s.DNA)**
61. Gel electrophoresis may be used to measure of restriction fragments -------- **Size**
62. Waardenburg's syndrome is disorder in ----------**Humans,**
63. Hamming distance is not typically used to compare DNA or protein sequences it is used for ---------- **DNA Similarity**
64. Genomic rearrangement results in a change of ---------- **Gene Ordering**
65. HGP revealed that.... Disease are hereditary ---------- **3000-4000**
66. Brute force is called ----- **Exhaustive Search Algorithm**
67. Insulin was sequenced by sanger in ------- **1951**
68. Uniprot is a database of protein --------- **All**
69. Gene mapping is determining....on chromosomes ------- **location and relative distance**
70. Genetic map distances are measured in -------- **Centi-Morgans (CM)**
71. Salamdar genome Is... time larger than human genome --------**10 (ten times)**
72. When transfer of gene occurs from parents to offspring is called....**Vertical Transfer**
73. TCGGGATTTCC sequence in Drosophila in the activation of....gene ----- **Immunity Gene**
74. Ab Initio method is based on ---------- **Gene prediction and repeat masker**
75. Next vertex algorithm is very..... algorithm ---------- **Precise.**
76. Less edit distance shows..... solution --------- **Better**
77. Brute force Algorithm was devised by ---------- **Stephon**

78. Which notation of spectral similarity shared peaks counts -------- **P1**
79. Growth of sequence in database is ---------- **exponential**
80. 95% of human genome is -------- **intron**
81. if two gene are equal they are called ---------- **homologus**
82. MSA is used to find ------- **Comparison**
83. Next vertex output is ------ **efficient**
84. Which theory help us to find DNA motif -------- **Gold Bug**
85. Generally ---------- ORF is Preferred. --------- **Longer**
86. Random Mutagenesis change --- Sequence ---------- **Amino Acid**
87. Aspartate is more ----------**mutable**
88. A pattern is an ----- sequence of symbols -------- **ordered**
89. Columns of an alignment containing spaces in top row are called --------- **Insertions**
90. Genes present in diff organisms having diff structure may have----- function ------ **similar**
91. gene for embryonic development -------- **Homeobox**
92. To carry out the steps of an algorithm humans are generally ---------- **Slow**
93. A linear time algorithm for string matching avoid useless shift of the ------ **Pattern**
94. Motif Are found in…… **DNA or Proteins**
95. Wardenburg's Syndron is disordered in …… **Human**
96. Systic Fibrosis Gene found on Chromosome No ----- **02**
97. Choufasman Algorithm based on ……. **Statistical Occurrence of Amino Acids.**
98. Probe Length……. **8-32**
99. There are ……. Steps in homology modeling loops formation. ----- **4th**
100. 3D1d alogorithm form in ……. **1991**
101. Which RNA converted into protein -------- **Primary**
102. Genomic Rearrangement results in change of……. **Gene Ordering**
103. Uniprot is data of protien……. **Sequence**
104. A drug target …….protien.------- **About 50**
105. HGP Revealed that …….Diseases are hereditary. **3000-4000**
106. ……is good for gapped alingment . **BLOSUMS50**
107. When senger Sequnce insulin……. **1955**
108. Neddleman alignment published in …….. **1970**
109. ……..Size Fragment can be lighted into BAC Vector… **Large**
110. Brute Force is Called……. **Exaustive Search Algorithms**
111. 3D1d Bowie Algorithm at el in ……**1991**
112. ……..Help in finding global minimum ------ **Global Energy Optimization**
113. BLOSUM Means…… **BLOCK Substitution Matrix**

114. Log Odd PAM Matric…… Substitution matrices are the log-odds matrices used for scoring amino acid substitutions in pairwise alignments. **250**
115. BLOSUM62 is good for……Alignment….. **Ungapped Alignment**
116. Loops are small consisting ……..Amino Acids ------ **3-4**
117. The NeedlemanWunsch algorithm finds……Alignment…. **Best Global**
118. Substitution scores can be derived from…….Model … **Probabilistic**
119. Dayhoff PAM matrices…… **1978**

SUBJECTIVEs

1. Why Sequencing Are Submit In Database? Write Precaution Of Sequencing?

Sequences are submitted to the databases in order to share them with the scientific community Generally sequences are submitted at the time of publication and are reviewed by peers.

Caution: It is important to ensure that sequence files do not contain any special characters • Control characters in addition to standard ASCII characters must be removed else they might mess up the analysis or data transfer.

2. Delete The One, Two And Three Nucleotide From Strand With Example?

To delete specific nucleotides from a DNA strand, we need to specify the position or range of positions where the deletions should occur.

Let's consider the DNA strand sequence: ATCGATCGATCG. If we want to delete the first three nucleotides (A, T, and C), the resulting sequence would be GATCGATCG.

If we want to delete nucleotides at specific positions, let's say positions 5, 6, and 7 (G, A, and T), the resulting sequence would be ATCCGATCG.

3. Translation Of Protein?

Protein synthesis is accomplished through a process called translation. After DNA is transcribed into a messenger RNA (mRNA) molecule during transcription, the mRNA must be translated to produce a protein. In translation, mRNA along with transfer RNA (tRNA) and ribosomes work together to produce proteins.

4. What Are MITES?

MITES (Miniature Inverted repeat TEs) – Features of both Class I and II – 400 bp. **(PPTS)**

5. Genequiz And MAGPIE?

MAGPIE (Multipurpose Automated Genome Project Investigation Environment) It's an automated genome analysis tool that is used for structural annotation.

GeneQuiz Focuses on deriving a predicted protein function based upon available evidence, including evaluation of similarity to the closest homologue in database.

6. Genome And Genome Informatics? 2

Genome: Genome sequencing provides the sequences of all the genes of an organism • A major application of Bioinformatics is analysis of full genomes that have been sequenced. Challenge is to identify those genes that are predicte d to have a particular biological function.

Genome Information:

Genome informatics is the field in which computational and statistical techniques are applied to derive biological information from genome sequences. Genome informatics includes methods to analyze DNA sequence information and to predict protein sequence and structure.

7. Restriction Enzyme?

Restriction enzymes are DNA-cutting enzymes used in genetic engineering. They recognize specific DNA sequences and cut the DNA at or near those sites. These enzymes play a vital role in molecular biology techniques such as DNA cloning and gene editing. They can produce either blunt ends or sticky ends, which enable the insertion, deletion, or modification of DNA sequences. By targeting and cutting DNA at precise locations, restriction enzymes allow scientists to manipulate genetic material, create recombinant DNA molecules, study gene function, and generate genetically modified organisms. Their specificity and versatility make them essential tools in genetic research and biotechnology applications.

8. Exon Changing?

Exon changing refers to the alteration of exons, which are the coding regions of genes that are transcribed into messenger RNA (mRNA) and ultimately translated into proteins. Modifying exons can have significant impacts on protein structure and function. This can be achieved through techniques such as exon skipping, where specific exons are intentionally excluded during mRNA processing, or exon replacement, where one exon is substituted with another. Exon changing methods are used in gene therapy and genetic engineering to correct genetic mutations, restore protein functionality, or introduce desired modifications. These approaches hold promise for treating genetic disorders and developing novel therapeutic strategies.**)**

9. Three Steps Of Genome Annotation?

Genome annotation is an important first step in **genome analysis**, based on; Finding significant alignment to sequences of known function in databases. There are two steps in genome annotation

- Structural annotation
- Functional annotation.

10. Structural Annotation

Identification of gene features like;

- Promoters
- Terminators
- Shine-Dalgarno sites
- DNA motifs
- Co-transcription units
- Operons in microbes.

11. Note On ORF?

ORF, or Open Reading Frame, is a section of DNA or RNA that has the potential to be translated into a protein. It is defined by the presence of a start codon, usually AUG, and a stop codon. Identifying ORFs helps in predicting protein-coding genes and understanding gene structure and function. Bioinformatics tools are used to search for ORFs in genetic sequences, aiding in genome annotation and analysis. ORFs play a crucial role in genomics, gene prediction, and unraveling the functional aspects of genetic information.

12. Define Structure of Annotation? Structural Annotation? What can we do by using this technique?

Structural Annotation Identification of gene features like,

- Promoters
- Terminators
- ShineDalgarno sites
- DNA motifs
- Co -transcription units
- Operons in microbes. MAGPIE (Multipurpose Automated Genome Project Investigation Environment) It's an automated genome analysis tool that is used for structural annotation. **(PPTS)**

Structural annotations are physical regions of a genome that encode a genomic feature. Examples of such annotations are genes, mRNA, transcript, repeat sequences, **(Internet)**

13. What Kind of Gene Expression OMNIBUS? Kind of Records of Gene Expression OMNIBUS (GEO)?

A public repository for the archiving and distribution of gene expression data submitted by the scientific community. MIAME compliant data – Minimum Information About a Microarray Experiment. Convenient for deposition of gene expression data, as required by funding agencies and journals

⇒ Curated resource for gene expression data
⇒ Browsing, querying, analysis and retrieval.

GEO has four kinds of records

- **Sample (GSM)** preparation and description of the samples
- **Platform (GPL)** technology used and the features detected – microarray or RNASeq.
- **Series (GSE)** defines a set of samples and how they are related
- **Datasets (GDS)** sample data collections assembled by GEO

14. EM BK Note?

EM BK, short for Electron-Magnetic Break, is a braking system used in some vehicles. It utilizes the principle of electromagnetic induction to generate braking force. When the driver applies the brakes, an electric current is passed through a coil, creating a magnetic field. This field interacts with the rotating metallic disc or drum, inducing eddy currents that generate resistance and slow down the vehicle. EM BK offers advantages such as quick response, precise control, and low wear compared to traditional friction-based braking systems. It is commonly used in hybrid and electric vehicles, where regenerative braking is employed to convert kinetic energy into electrical energy for battery charging.

15. Brute Runtime Algorithm?

In actual cases, the performance (Runtime) of an algorithm depends on n, that is the size of the input or the number of operations is required for each input item. Runtime grows quicker than previous all based on n. Runtime grows even faster than polynomial algorithm based on n.

16. Collinear Nucleotide?

Charles Yanofsky proved that a gene and its protein product are collinear. Yanofsky's experiment was so influential that nobody even questioned about codons and for almost 2 decades biologists believed that a protein was encoded by a long string of adjacent triplets.

17. Four Cases Of Pattern Finding?

Pattern recognition is the automated recognition of patterns and regularities in data. Pattern recognition is closely related to artificial intelligence and machine learning, together with applications such as data mining and knowledge discovery in databases (KDD), and is often used interchangeably with these terms. However, these are distinguished: machine learning is one approach to pattern recognition, while other approaches include hand-crafted (not learned) rules or heuristics; and pattern recognition is one approach to artificial intelligence, while other approaches include symbolic artificial intelligence

18. Pre Order Algorithm?

PREORDER(v)

- output v
- if v has children
- PREORDER(left child of v)
- PREORDER(right child of v)

19. Ms2?

Infects E.Coli and Enterobacteriaceae

- Single stranded sense RNA
- Encodes MS2 coat protein. MS2 sequencing
- The entire nucleotide sequence was established by nuclease digestion and characterization of fragments. MS2 genome
- 49 different codons in the genetic code specify the sequence of the 129 aminoacids long coat polypeptide

20. KMP Matcher?

Knuth Morris Pratt (KMP) is an algorithm, which checks the characters from left to right. When a pattern has a subpattern appears more than one in the sub-pattern, it uses that property to improve the time complexity, also for in the worst case. The time complexity of KMP is O(n).J **(Internet)**

21. Edit The MANHATTEN Graph?

To edit the Manhattan graph, several approaches can be taken to modify its structure or attributes.

One possible edit is to add or remove vertices or nodes to the graph. This can involve adding new locations or landmarks as vertices or removing existing ones. Additionally, edges or connections between vertices can be modified or added to represent different routes or pathways in the Manhattan area.

Another editing option is to update the attributes or properties of the graph. For example, the weight or distance associated with each edge can be adjusted to reflect accurate travel distances. Other attributes such as the type of transportation (e.g., subway, bus, walking) can also be included to provide more comprehensive information.

Overall, editing the Manhattan graph involves making changes to its structure, vertices, edges, or attributes to better represent the transportation network and facilitate accurate analysis or navigation within the area.

22. Application Of Bioinformatics? Five Application Of Bioinformatics?

Drug Development, Molecular Medicine, Crop Improvement, Personalized Medicine, Microbial Genome. Preventive Medicine, Gene Therapy. Biotechnology, Comparative Study.

23. Swiss Port, Fasta, Gene Bank, Orf?

SWISS-PROT

SWISS-PROT is an annotated protein sequence database established in 1986 and maintained collaboratively, since 1987, by the Department of Medical Biochemistry of the University of Geneva and the EMBL Data Library (now the EMBL Outstation -The European Bioinformatics Institute.

FASTA Sequence Format

- ⇒ DNA sequences are stored in specified formats
- ⇒ Different softwares need sequences in different formats. FASTA is the most frequently used format to present DNA and Protein sequences
- ⇒ It is recommended that all lines of text be shorter than 80 characters in leng th

Example:

>gi|568815581:c7687550-7668402 Homo sapiens chromosome 17, GRCh38 Primary Assembly

GATGGGATTGGGGTTTTCCCCTCCCATGTGCT
CATCTAGAGCCACCGTCCAGGGAGCAGGTAGC TGCTGGGCTCTCCACGACGGTGACACGC----
----

Genebank Format

- A sequence file in GenBank format can contain several sequences
- One sequence starts with a line containing the word LOCUS and a number of annotation lines. The start of the sequence is marked by a line containing "ORIGIN" and the end of the sequence is marked by two slashes ("//")

Example:

ORIGIN 1 aacctgcgga aggatcatta ccgagtgcgg gtcctttggg cccaacctcc catccgtgtc 61 tattgtaccc tgttgcttcg gcgggcccgc cgcttgtcgg ccgccggggg ggcgcctctg 121 ccccccgggc ccgtgcccgc cggagacccc aacacgaaca ctgtctgaaa gcgtgcagtc 181 tgagttgatt gaatgcaatc agttaaaact ttcaacaatg gatctcttgg ttccggc //

24. Bypass Algorithm?

NEXTVERTEX Algorithm BYPASS Algorithm Skip the subtree rooted at vertex (a, i) Increment ai (unless ai = k BYPASS(a, i,L, k)

- for j i to 1

- if aj < k
- aj aj + 1
- return (a, j)
- return (a, 0) (Proper Answer PPTS mn dykhain mujy ye sahi trha samj nh aaya)

25. Brute Force?

Brute force approach-find all local similarities Each substring from the genomic sequence that exhibits sufficient similarity to the target protein could be considered a putative exon Exon-flanking dinucleotides AG and GT Overlapping.

26. KMP Algorithm?

KMP algorithm: searches for occurrences of a "word" W within a main "text string" S by employing the observation that when a mismatch occurs, the word itself embodies sufficient information to determine where the next match could begin

27. Why We Cannot Use Beret Force To Approach Motif?

Using brute force to approach motif finding is not feasible due to the exponential growth of possibilities. Motif finding involves searching for a common pattern or motif in a set of DNA or protein sequences. Brute force would require exhaustively checking all possible combinations of motifs, which becomes computationally intractable as the sequence length or number of sequences increases. The number of possible motifs grows exponentially with the length of the motif and the number of sequences, making it impractical to explore all combinations. Therefore, more efficient algorithms and techniques are needed for motif finding.

28. Why Brute Force Not Use For Motif Finding?

Brute force is not used for motif finding due to its inefficiency. Motif finding involves searching for a common pattern in a set of DNA or protein sequences. Brute force would require checking all possible motif combinations, which becomes computationally infeasible as the sequence length or number of sequences increases. It is not a practical approach due to the exponential growth in possibilities, demanding more efficient algorithms for motif discovery.

29. Six Model Of Organism?

E.Coli- Bacteria S.Cerevisiae- Yeast C.elegans- Worm, D -Melangaster- Fly Daniorerio-Zebrafish Musmusculus-Mouse Homo sapien- Human Arabidopsis- Plant

30. Difference B/W EMBL And DDJB? 5

- **European Molecular biology Lab** (EMBL established1980): The European Molecular Biology Laboratory is a molecular biology research institution supported by 25 member states, four prospect and two associate member states. EMBL was created in 1974 and is an intergovernmental organisation funded by public research money from its member states.
- **DNA databank of Japan** (DDBJ) established in Mishima japan (1984): The DNA Data Bank of Japan is a biological database that collects DNA sequences. It is located at the National Institute of Genetics in the Shizuoka prefecture of Japan. It is also a member of the International Nucleotide Sequence Database Collaboration or INSDC.

31. Difference Indel. Insertion And Deletion In An Alignment?

The columns of the alignment containing one space are called **indels**, with the columns containing a space in the top row called **insertions** and the columns with a space in the bottom row **deletions**.

32. Note On Splice Signal? 5

A **splice site mutation** is a genetic mutation that inserts, deletes or changes a number of nucleotides in the specific site at which splicing takes place during the processing of precursor messenger RNA into mature messenger RNA. Splice site consensus sequences that drive exon recognition are located at the very termini of introns. The deletion of the splicing site results in one or more introns remaining in mature mRNA and may lead to the production of abnormal proteins.

33. Lines And Sine, Chromatin And Heterochromatin? 2

- **SINES** (Short Interspersed Nuclear Elements – 80-300 bp – Alu repeats
- **LINES** (Long Interspersed Nuclear Elements) – 6-8 KB long.
- **Chromatin** is a mass of genetic material composed of DNA and proteins that condense to form chromosomes during eukaryotic cell division. Chromatin is locate d in the nucleus of our cells.
- **Heterochromatin** is a tightly packed form of DNA or condensed DNA, which comes in multiple varieties. These varieties lie on a continuum between the two extremes of constitutive heterochromatin and facultative heterochromatin. Both play a role in the expression of genes.

34. What Is Personalized Medicine?

Personalized medicine, precision medicine, or theranostics is a medical model that separates people into different groups—with medical decisions, practices, interventions and/or products being tailored to the individual patient based on their predicted response or risk of disease.

35. Describe Region Of DNA?

DNA. The vast majority of organisms encode their genes in long strands of DNA (deoxyribonucleic acid). DNA consists of a chain made from four types of nucleotide subunits, each composed of: a five-carbon sugar (2-deoxyribose), a phosphate group, and one of the four bases adenine, cytosine, guanine, and thymine.

36. Difference Local Alignment Input And Output? 5

Local Alignment Problem : Find the best local alignment between two strings

Input : Strings v and w and a scoring matrix δ

Output: Substring of v and w whose global alignment, as defined by δ, is maximal amoung all global alignment of all substrings of v and w.

37. Difference B/W Gene Bank And Fasta Format?

GenBank to **FASTA** accepts a **GenBank file** as input and returns the entire DNA sequence in **FASTA format**. Use this program when you wish to quickly remove all of the non-DNA sequence information from a **GenBank file**. Paste the contents of one or more **GenBank files** into the text area below.

38. Five Task Of Genome Analysis?

The five tasks of genome analysis include:

- Genome sequencing: Determining the complete DNA sequence of an organism's genome.
- Genome assembly: Piecing together short DNA fragments to reconstruct the complete genome sequence.
- Annotation: Identifying and labeling functional elements within the genome, such as genes, regulatory regions, and non-coding sequences.
- Comparative genomics: Comparing genomes across different species to understand evolutionary relationships and identify conserved regions.
- Functional genomics: Studying gene expression, protein interactions, and other functional aspects to decipher the roles of genes and their products in biological processes.

39. How Mutation In DNA?

Mutations in DNA can occur due to errors during DNA replication, exposure to mutagenic agents like radiation or chemicals, or spontaneous changes. These mutations can involve substitutions, insertions, deletions, or rearrangements of nucleotides, leading to genetic variations within an organism's DNA sequence.

40. Write Note On Swiss Port? 2

Swiss-Prot is a manually curated protein sequence database that provides comprehensive and accurate information about proteins from various organisms.

41. What Are Suffix? 2

It is a compressed tree containing all the suffixes and allows many problems on strings to be solved quickly.

- Exact searching or pattern matching methods
- Approximate searching or pattern matching methods
- Position weight matrices.
- Suffix trees.

42. Write Note On Transposable Elements?

Transposable elements, also known as "jumping genes," are DNA sequences capable of moving or "transposing" within a genome. They can insert themselves into new locations, causing genetic rearrangements. They are found in the genomes of organisms ranging from bacteria to humans and have played a significant role in shaping genome structure, evolution, and genetic diversity.

43. Difference B/W Mouse Synteny And Human X Chromosome? 3

Synteny between species means not only that orthologous (functionality and ancestry identical). Genes are resent but that they are present in the same order on the genome, thus indicating common 4pc uence of the mouse genome and its comparison with the previously sequenced human genome reveal that 90.2% of the human genome and 93.3% of the mouse genome lie in the conserved egment. So not only we can find the homologous sequence between the two genomes; the gene Ily lie on the chromosome in the same order.This is powerful evidence for common ancestry.

44. What Are Primary Medical Database? Enlist Four Medical Database?

NIVEL Primary Care Database collects data that is routinely recorded in the health care provider's electronic health record system. This includes data on health problems and treatment. **(Database are not found)**

45. What Is Long Common Sequence?

The simplest form of a sequence similarity analysis is the Longest Common Subsequence (LCS) problem, where we eliminate the operation of substitution and allow only insertions and deletions. A subsequence of a string v is simply an (ordered) sequence of characters (not necessarily consecutive) from v.

46. Write Note On Intron, Exon And Splicing Process?

Introns: intron is a long stretch of noncoding DNA found between exons (or coding regions) in a gene. Genes that contain introns are known as discontinuous or split genes as the coding regions are not continuous. Introns are found only in eukaryotic organisms.

Exons: An exon is any part of a gene that will encode a part of the final mature RNA produced by that gene after introns have been removed by RNA splicing. The term exon refers to both the DNA sequence within a gene and to the corresponding sequence in RNA transcripts.

Splicing Process: RNA splicing is a process that removes introns and joins exons in a primary transcript. An intron usually contains a clear signal for splicing (e.g., the beta globin gene). In some cases (e.g., the Tau gene), a splicing signal may be masked by a regulatory protein, resulting in alternative splicing.

47. What Is Alignment How We Calculate In Give Example?

Alignment and Alignment Score. (Here "alignment score" is "matching score" used in the definition of the sequence identity in problemE.html.) Remark: The program scoreout.c is designed for computing the alignment score of a given pair of amino acid sequences.

48. Write Detail On Sequence LOGOS?

A sequence logo is a graphical representation of the sequence conservation of nucleotides (in a strand of DNA/RNA) or amino acids (in protein sequences) . A sequence logo is created from a collection of aligned sequences and depicts the consensus sequence and diversity of the sequences. **(Internet)**

49. Type Of Complexity? 2

Three types of complexity could be considered when analyzing algorithm performance. These are worst -case complexity, best-case complexity, and average-case complexity. ... Average-Case running time for a function, f(n) such that where a,b and c are constants can be described as an average between best-case and worstcase. **(Internet)**

50. Note On ETS?

ETS stands for the E26 transformation-specific (ETS) family of transcription factors. ETS proteins are a diverse group of transcription factors that regulate gene expression and play critical roles in various cellular processes such as development, differentiation, proliferation, and apoptosis. They have been implicated in numerous diseases, including cancer, making them important targets for research and therapeutic interventions.

51. Exon Predication?

GeneAlign is a coding exon prediction tool for predicting protein coding genes by measuring the homologies between a sequence of a genome and related sequences, which have been annotated, of other genomes. ... The rates of missing exons and wrong exons are smaller than 1%.

52. Synteny Its Important And Example?

Synteny refers to the conserved arrangement of genes or genetic loci on chromosomes across different species. It is important in evolutionary studies and comparative genomics, as it allows researchers to infer evolutionary relationships and identify conserved genomic regions.

An example of synteny is the conservation of gene order between humans and mice, where large blocks of genes are found in the same order on corresponding chromosomes.

53. List Of Method Finding Pattern Of Genome?

Extac Searching Method, Approximate Searching Method, Position Weight Matrices, Suffix Tree.

54. Generalized Algorithm For Pattern Finding?

A generalized algorithm for pattern finding typically involves the following steps:

1. Input: Obtain the sequence data or dataset.
2. Preprocessing: Clean and prepare the data, if necessary.
3. Pattern Definition: Define the pattern or motif to search for.
4. Pattern Finding: Implement an algorithm to search for and identify occurrences of the pattern within the dataset.
5. Output: Report the locations or statistics of the identified pattern.

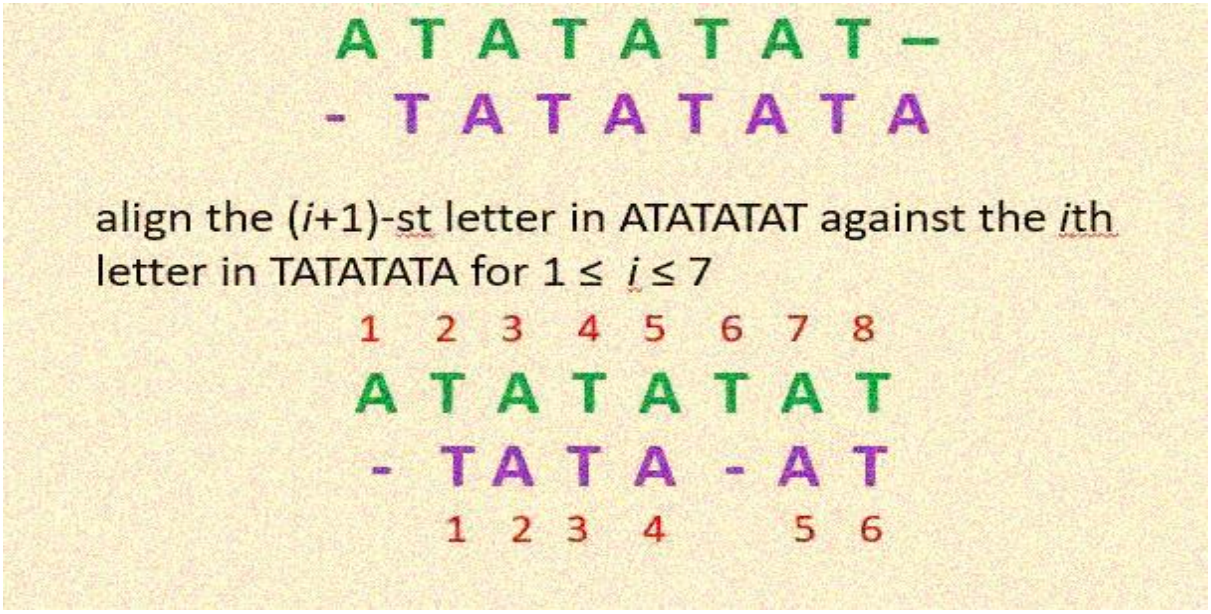
55. Difference B/W DNA And RNA?

DNA is a double- stranded molecule while RNA is a single stranded molecule. DNA is stable under alkaline conditions while RNA is not stable. ... DNA and RNA base pairing is slightly different since DNA uses the bases adenine, thymine, cytosine, and guanine; RNA uses adenine, uracil, cytosine, and guanine.

56. Edit Distance And Example?

Edit distance between two strings as the minimum number of editing operations needed to transform one string into another, where the edit operations are insertion, deletion, and substitution of one symbol for another. It is often the case that the *i*th symbol in one sequence corresponds to a symbol at different position in other. Mutation in DNAevolutionary process: DNA replication- substitutions, insertions, and deletions

of nucleotides, leads to "edited" DNA texts. Whether the *i*th symbol in one DNA sequence corresponds to the *i*th symbol in the other



57. Difference B/W Clustering By Subgroup And Cluster By Linkage?

Clustering By Subgraph

- Each sequence is a vertex
- Significant alignment score is an edge
- Trimming by removing weak edges (High P/E)

Clustering by Linkage

- Each sequence is a vertex
- Significant alignment score is an edge
- Trimming by removing weak edges (High P/E)
- Or remove > e-6
- Remaining subgraph should share 2/3rd of edges.

58. Similarities Approach Of Gene Prediction?

A similarity-based approachPreviously sequenced genes Unknown genes in newly sequenced DNA fragments Combinatorial puzzle-find a set of substrings (candidate exons) whose splicing best fits the target. Brute force approachfind all local similarities Each substring from the genomic sequence that exhibits sufficient similarity to the target protein could be considered a putative exon Exon-flanking dinucleotides AG and GT Overlapping. Model a putative exon with a weighted interval in the genomic

sequence, parameters (l, r, w) l is the left-hand position, r is the right-hand position, and w is the weight of the putative exon w -local alignment score Likelihood that this interval is an exon A chain is any set of non overlapping weighted intervals. Total weight of a chain A maximum chain Weights of all intervals are positive (w > 0). Model a putative exon with a weighted interval in the genomic sequence

59. Write Note On Waardenburg Syndrom?

Hearing loss Two Different colored eyes Gene present on chromosome 2 Splotch gene in mice, Human genomemouse genome Cut into 300 genomic fragments-Synteny blocks Chromosome 2 in humansmouse chromosomes 1, 2, 3, 5, 6, 7, 10, 11, 12, 14, and 17, Genome rearrangement results in a change of gene ordering Analysis of human and mouse genomes-250 genomic rearrangements. **(PPTS)**

OR

Waardenburg syndrome is a genetic disorder characterized by distinctive features such as hearing loss, changes in pigmentation (hair, skin, and eye color), and facial abnormalities. It is caused by mutations in genes involved in the development and migration of neural crest cells. There are different types and variations of Waardenburg syndrome, each with specific clinical features.

60. Human Genome Project?

Pilot project, the Human Genome Initiative begun by Department Of Energy (DOE) in 1986. National Human Genome Research Initiative (NHGRI), a federally funded organization (NIH) started in 1988 (Francis Collin)

- Celera, a commercial vendor (Craig Venter) joined the venture in 1998.
- Both simultaneously announced the completion of the project in 2000
- Total 3.4 billion bases sequenced at a cost of $1/base. **(PPTS)**

OR

The Human Genome Project was an international scientific research project with the goal of determining the sequence of nucleotide base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint. **(Internet)**

61. Edit Graph?

The grid that is achieved after alignment is similar to the Manhattan grid where each entry in the grid looks like a city block The graph called Edit Graph The main difference between the Manhattan and Edit Graph is that we can move along the diagonals in the Edit Graph.

62. Enzyme Commission Numbers With Example

Enzyme Commission (EC) numbers:

It is another scheme which was put forward by the **E**nzyme **C**ommission (EC) that was working under the **IUBMB (International Union for Biochemistry and Molecular Biology).** They say that the enzymes are classified on the basis of the reactions they catalyze and have a 4-digit scheme which is actually the enzyme commission number: **EC a.b.c.d**

- **'a' (**first digit**)** informs that it is from one of the 6 classes of biochemical reactions (enzyme might be coming from one of these classes).
- **'b' (**second digit**)** informs that is from the group of substrate (the thing on which the enzyme attacks**).**
- **'c' (**third digit**)** informs us that it is anaccepter molecule.
- **'d' (**fourth digit**)** gives thedetails of biochemical reaction

For example, tripeptideaminopeptidases **EC 3.4.11.4** Where**,**
3 – tells us that it is a Hydrolase (use water to break substrate).
This **3.4-** tell us thatit is aHydrolase that acts on the peptide bonds.

The **3.4.11-** tells us that it is a Hydrolase that cleaves the amino terminal amino acids of polypeptide.

While putting everything together**, EC 3.4.11.4-** it tells us that it is a Hydrolase that cleaves the amino terminal amino acids of a tri-peptide

63. General Algorithm

A general algorithm is a step-by-step procedure or set of instructions that can be applied to a variety of problems or situations. It provides a general framework for solving a class of problems and typically involves defining inputs, performing computations or operations, and producing desired outputs. General algorithms aim to be versatile and applicable across different scenarios.

64. Hierarchical Agglomerative General Algorithm

- Find the 2 closest objects and merge them into a cluster
- Find and merge the next two closest points, where a point is either an individual object or a cluster of objects
- If more than one cluster remains, return to step 2

65. Six Model Of Organisms ?Marks 3

Model organisms: Most of the times, while we are doing those genome sequencing projects, our objective is to find the cure of some disease, or improving some variety of the crop for enhancing its production, or looking into some drugs against different organisms so it's a good idea to have some model organisms that can be used for studying various processes in labs and there are is a range of model organisms which includes:

1. **E. coli – bacteria**
2. **S. cerevisiae – yeast**
3. **C. elegans – worm**

4. **D.melanogaster – fly**
5. **Daniorerio – zebrafish**
6. **Musmusculus - mouse**
7. **Homo sapiens – you and me 8. Arabidopsis - plant**

66. Self Comparison ?

Self-comparison of proteome (sometimes we are interested in finding genes which are kind of duplicated within the same organism, so in order to do that this selfcomparison of proteome is made, where proteome is the collection of the proteins which are derived from those genomes. Therefore, the whole collection of one organism's proteins can be termed as proteome and we can compare it with itself and can find about those sequences which are being duplicated in it).

اگر کامیابی حاصل کرنا چاہتے ہو تو
مسلسل محنت کرتے رہو۔

عائزہ رائیٹس