BIF-401 CURRENT PAPERS SOLUTION-PART 1

CREATED BY IRTAZA ALI

Q. How junctions are formed?

Junctions are formed when two or more double-stranded regions of RNA converging to form a closed structure.

Q. would you like to please write and describe the formula of (distance btw two sequence) ?

$$Distance = 1 - \frac{Matches}{Comparison}$$

Q. Name famous protein database.

UniProt & SwissProt

Q. objective of comparing sequence

• Similarity of sequences • Evolutionary History • Predict the function of molecules

Q. Difference between DNA and RNA

DNA has A, T, C & G while RNA has A, U, G & C • DNA is double stranded while RNA is single stranded

Q. Calculate scoring function

DP uses a "scoring function" to deal with matches, mismatches and gaps! An example Scoring Scheme:
Match: +a Mismatch: - b Gap: - c

Score = #matches + #mismatches + #gaps

DP uses a "scoring function" to deal with matches, mismatches and gaps!

C T G T C G – C T G C

\- T G C – C G – T G -

-5 10 10 -2 -5 -2 -5 -5 10 10 -5 Total = 11

Q. Write 2 types MS based proteomics protocol

There are two types of proteomics protocols that are usually employed

• Bottom Up Proteomics • Top Down Proteomics

Join VU BIO ANIMALS

Q. How are bulges formed? (2)

• In RNA secondary structure the Bulges, are formed when a double-stranded region cannot form base pairs perfectly.

Q. Three ways of DNA modifications (3)

DNA get modified by three ways that is 1) DNA Methylation 2) Histone Modification and 3) Acetylation

Q. Three applications of Bioinformatics (3)

Genomics , in genomics we find DNA sequencing , Gene finding and Genome assembly

Evolutionary Studies , in evolutionary studies we get idea about Evolutionary relationship their distances , Phylogenetics and Tree of life

Proteomics :- in this category we use protein sequencing study about protein structures PPI protein Protein Interaction

Q. complete name of EMBL and EBI (2)

European Molecular Biology laborartory and European Bioinformatics Institute

Q. Bioinformatics scope??3

Bioinformatics primarily deals with digitalized biological information as well as data reported from biology experiments. Computational methods, data processing techniques and algorithms are employed in addressing the following issues:

 Storage of data Organization data Analysis of many experiments For representation of biological information

Q. Url of uni prot ?2

www.uniprot.org

Q. How score matrics compute??3

Needleman Wunsch Algorithm is the way for alignment. The method is same like dot plot but it computes the scores in different way. We Start with a zero in the second row, second column. Move through the cells row by row, calculating the score for each cell.

Q. Define mutations (2)

A mutation is a change that occurs in our DNA sequence, either due to mistakes when the DNA is copied or as the result of environmental factors such as UV light

Join VU BIO ANIMALS

Q. Acronym of BLAST (2)

BLAST. Basic Local Alignment Search Tool

Q. Three uses of Dot Plot (3)

1) We use it to compare 2 sequences 2) It is used to find the similarities btw sequences 3) Differences btw 2 sequences

Q. Smith Watermann Algorithm (3)

Local alignments can be performed using Smith Waterman algorithm • Traceback can start from any position in the scoring matrix • Local alignments can be extracted by starting from a high score till reaching '0'

Q. Biological Simulation help community (2)

Biological Simulations help community by predicting diseases progression & outcomes

Q. Purine and pyramidine bases(2)

Purine is double ring structure bases while pyrimidine are single ring bases. Purine contains adenine and guanine while Pyrimidine contains Cytosine thymine and uracil in case of RNA

Q. Define BLASTn, BLASTp, BLASTx (3)

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

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

Blastx: Compares a nucleotide query sequence against a protein sequence database

Q. EXPASY (2)

Expasy provides access to a variety of online databases and tools. • Depending upon your requirement, you find sequence information from Expasy

Q. what is the foundation principle for predicting rna 1' And 2' structure. 2m

Folding of RNA make the foundation principle for predicting RN1` and 2` structures. Cause there is no folding in RNA1`

Q. would you plz enlist few points come under the space of the bioinformatics. 3m

Bioinformatics is an exciting, truly interdisciplinary field that brings 'Life Sciences' (biology, genetics, medicine, pharmaceutics) together with 'Computational Sciences'. (computer sciences, mathematics, and statistics

Q. what is the purpose of UniProt data base , highlights its 3 significant features. 3m

uniProt contains all the data base related to proteins , its actual purpose is to keep all the sequences of proteins can be available publically so people can get access.

VU BIO ANIMALS

Significances. It is used to compare the proteins sequences, 2. Comparison result leads to the evolutionary relation of protein .3) we can predict about the protein functions 4) can be able to know about PPI

Q Define gene bank and what can we obtain from gene bank?

A Genbank is data base where all kind of information related to DNA/RNA sequences are available we can compare all known and unknown sequences .

Q. Substitution and Indel difference.

Subsitution and Indel both are the types of mutations in subsituion one amino acid is replace with another aminoacid within a protein sequences. While indel means insertions or deletions if an amino acid is removed from the sequence will be called deletions and if the amino acid is introduce within the sequence called insertions and in combine it is called indel

Q. Psuedonots

Ans: when in RNA, the unpaired bases of 3´ end fold up to itself within tertiary structure, it may result in formation of abnormal folding called pseudoknots

Q What is central dogma in bioinformatics? (2)

According to the central dogma DNA codes information for RNA and RNA makes the Protein and that protein along with some organelles make cells and its systems.

Q. Write down the URL of NCIB.

https://www.ncbi.nlm.nih.gov

Q. What is nucleotide Pairing

Molecule of DNA contains only four base pairs (A, T, C, and G) which are repeated thousands of time and Adenine "A" pairs with Cytosine "C", While Thymine "T" binds with Guanine "G" and all pairings are with the help of Hydrogen bonding

Q. write down the three names of Hydrophobic Amino acids

hydrophobic amino acids are those with side chain that do not like to reside in an aqueous medium. There are some hydrophobic amino acids,

Alanine …. Valine …. Leucine …. Isoleucine …. proline …. Methionine … tryptophan.

Q. What are the types and categories of RNA and define their function?

There are actually several types of ribonucleic acid or RNA, but most RNA falls into one of three categories:

- mRNA or Messenger RNA
  mRNA transcribes the genetic code from DNA into a form that can be read and used to make proteins.
  mRNA carries genetic information from the nucleus to the cytoplasm of a cell.

Join VU BIO ANIMALS

- rRNA or Ribosomal RNArRNA is located in the cytoplasm of a cell, where ribosomes are found. rRNA directs the translation of mRNA into proteins.
- tRNA or Transfer RNA
  Like rRNA, tRNA is located in the cellular cytoplasm and is involved in protein synthesis. Transfer RNA brings or transfers amino acids to the ribosome that correspond to each three-nucleotide codon of rRNA. The amino acids then can be joined together and processed to make polypeptides and proteins.

Q. Identity and Similarity?

"Identity is the number of nucleotides/amino acids which match exactly b/w 2 biological sequences"

Similarity is the comparison between sequences calculated by using alignment approach

Q.. write down the steps of PAM matrices.

1. Align the protein sequence which are 1-PAM Unit diverge.
2. Let Ai,i be the number of times Ai is substituted by Ai.
3. Compute the frequency fi of amino acid Ai.

Then, PAM1=pii= $\frac{A_{ij}}{\sum_k A_{ik}}$

PAM 'n'= $(PAM1)^n$

Q Contribution of Bioinformatics in societies

Bioinformatics can be applied to life in many ways it helps us to understand the sequence and function of biomolecules and their relationships. Recent trends in bioinformatics involve development of personalized therapeutics for cancer and diabetes.

Q.. what is BLOSSOM matrice? Also mention its steps.

BLOSUM matrices is also called the Block substitution matrix without any gap although it has mismatches in sequences.

There are three steps to compute the BLOUSM Matrices.

Step 1: Eliminate sequences that are identical in x% positions

Step 2: Compute observed frequency $f_{i,j}$ of aligned pair A_i to A_j. Hence, $f_{i,j}$ becomes the probability of aligning A_i and A_j in the selected blocks.

Step 3: Compute f_i which is the frequency of observing A_i in the entire block.

Q. What is the most common data bases (2)

Most common Data bases are Genebank for DNA/RNA sequences and UniProt and SwissProt for Protein sequences.

Q. Write the salient feature in block diagram of MS (mass spectrometer) (3)

Mass spectrometers are used to measure the molecular weight of proteins and peptides • Given that several proteins may have the same mass, how can we identify the protein sequence

Join VU BIO ANIMALS

Q. Difference between PSA and MSA

Pairwise Alignment	Multiple Sequence Alignment (MSA)
An alignment procedure comparing two biological sequences of either protein, DNA or RNA	An alignment procedure comparing three or more biological sequences of either protein, DNA or RNA
Pairwise alignments can be generally categorized as global or local alignment methods.	MSA is generally a global multiple sequence alignment
Comparatively simple algorithm is used	Complex sophisticated algorithm is used
A general global alignment technique is the Needleman–Wunsch algorithm. A general local alignment method is Smith–Waterman algorithm.	A technique called progressive alignment method is employed. In this approach, a pairwise alignment algorithm is used iteratively, first to align the most closely related pair of sequences, then the next most similar one to that pair, and so on.

Q. Enlist 3 features of MSA and PSA

MSA	PSA
a) Primarily to find out conserved regions between the two sequences. b)Similarity searches in a database	a) To detect regions of variability or conservation in a family of proteins b) Phylogenetic analysis (inferring a tree, estimating rates of substitution, etc.) c) Detection of homology between a newly sequenced gene and an existing gene family prediction of protein structure d) Demonstration of homology in multigene families

Join VU BIO ANIMALS