BT504 – GENOMICS & PROTEOMICS

FINAL TERM MOST IMPORTANT QUESTIONS

PREPARED BY: SIR MUGHAL (ADMIN, BS BIOLOGICAL SCIENCES)

GROUP LINK: https://chat.whatsapp.com/BjC6W0jZx8o9yaC2ISh1Gu

1. Name and explain types of enzymes shortly.

There are the main types of enzymes, each classified based on the type of reaction they catalyze:

1. ***Oxidoreductases:***
   Catalyze oxidation-reduction reactions (transfer of electrons). *Example:* Dehydrogenase.
2. ***Transferases:***
   Transfer functional groups (e.g., methyl, phosphate) from one molecule to another. *Example:* Kinase.
3. ***Hydrolases:***
   Break bonds using water (hydrolysis). *Example:* Amylase, which breaks down starch.
4. ***Lyases:***
   Add or remove groups to form double bonds without hydrolysis or oxidation. *Example:* Decarboxylase.
5. ***Isomerases:***
   Rearrange atoms within a molecule to form isomers. *Example:* Isomerase.
6. ***Ligases:***
   Join two molecules using energy from ATP. *Example:* DNA ligase.

2. Write 5 Advantages of overlapping layout consensus?

There are five advantages of the Overlap-Layout-Consensus (OLC) approach used in genome assembly:

1. ***Accurate Assembly of Long Reads:***
   OLC works well with long-read sequencing data (like PacBio or Oxford Nanopore), improving accuracy in assembling complex regions.
2. ***Better Handling of Repeats:***
   Long reads allow OLC to distinguish between repetitive sequences more effectively than short-read assemblers.
3. ***High Consensus Quality:***
   The consensus step helps produce a more accurate final sequence by combining overlapping regions.

4. ***Fewer Gaps in Assembly:***
   Due to long-read support and overlap matching, OLC can generate more continuous genome assemblies with fewer gaps.
5. ***Good for Small Genomes:***
   OLC is especially efficient and effective for assembling small genomes, such as bacterial or viral genomes, where high precision is needed.

3. Differentiate between Sanger's and Maxam Gilbert method?

Sanger's method and the **Maxam-Gilbert method** are both classical DNA sequencing techniques, but they differ significantly in their approach. Sanger's method, also known as the chain termination method, relies on the use of dideoxynucleotides (ddNTPs) that terminate DNA synthesis during replication. It requires a single-stranded DNA template and a DNA polymerase enzyme. The incorporation of ddNTPs results in DNA fragments of varying lengths, which are then separated by electrophoresis to determine the sequence. This method is relatively simple, safe, and capable of producing long read lengths (up to 1000 base pairs), making it widely used, especially in small-scale sequencing.

In contrast, the **Maxam-Gilbert method** is based on chemical cleavage of DNA at specific bases. It involves treating radioactively labeled double-stranded DNA with various chemicals that break the DNA at particular nucleotides. While it was one of the first sequencing methods developed, it is more complex, requires hazardous chemicals, and generally produces shorter read lengths. Due to these limitations, Maxam-Gilbert sequencing has become largely obsolete and has been replaced by safer and more efficient techniques like Sanger sequencing and modern high-throughput methods.

4. Write functions of miRNA.

Functions of miRNA (microRNA):

1. ***Gene Regulation:*** miRNAs regulate gene expression by binding to messenger RNA (mRNA) and either degrading it or inhibiting its translation, preventing the production of specific proteins.
2. ***Cell Differentiation and Development:*** They play a crucial role in controlling cell fate, tissue development, and organ formation by fine-tuning gene activity.
3. ***Apoptosis and Cell Cycle Control:*** miRNAs can promote or inhibit programmed cell death (apoptosis) and regulate cell division, helping maintain cellular balance.
4. ***Immune System Modulation:*** They influence immune responses by regulating cytokine production and immune cell development.
5. ***Disease Association:*** Dysregulation of miRNAs is linked to diseases such as cancer, cardiovascular disorders, and neurological conditions, making them potential biomarkers and therapeutic targets.

5. What is output of Genescan?

The output of GeneScan is a detailed prediction of gene structure within a DNA sequence. It provides:

1. ***Predicted Gene Locations*** – identifies the positions of genes within the input DNA.
2. ***Exon and Intron Boundaries*** – shows where exons (coding regions) and introns (non-coding regions) start and end.

3. ***Start and Stop Codons*** – indicates the locations where gene translation begins and ends.
4. ***Coding Potential*** – scores or probabilities showing how likely each predicted region is to be part of a functional gene.

6. Write functions of Microarray?

Functions of Microarray:

1. ***Gene Expression Analysis:***
   Microarrays can measure the expression levels of thousands of genes simultaneously, helping identify which genes are active or inactive in different cells or conditions.
2. ***Genotyping and SNP Detection:***
   They are used to detect genetic variations such as single nucleotide polymorphisms (SNPs), which are important in studying genetic diseases and individual responses to drugs.
3. ***Disease Diagnosis and Classification:***
   Microarrays help in identifying gene expression patterns associated with specific diseases, especially cancers, allowing for more accurate diagnosis and classification.
4. ***Drug Discovery and Toxicology:***
   Used to study how genes respond to drug treatments, helping identify potential drug targets and assess toxicity.
5. ***Comparative Genomic Hybridization (CGH):***
   Microarrays can detect chromosomal gains or losses in different samples, which is useful in identifying genetic abnormalities.

7. Write factors of using gene prediction.

Factors to consider when using gene prediction include:

1. ***Sequence Quality:*** High-quality, error-free DNA sequences are essential for accurate prediction.
2. ***Organism Type:*** Prediction methods vary between prokaryotes and eukaryotes due to differences in gene structure (e.g., presence of introns in eukaryotes).
3. ***Gene Structure Features:*** Elements like promoters, start/stop codons, splice sites, and coding regions must be accurately identified.
4. ***Availability of Training Data:*** Gene prediction tools rely on known gene data to train models; better training data improves accuracy.
5. ***Comparative Genomics:*** Using conserved gene sequences from related organisms helps improve predictions.
6. ***Annotation Tools and Algorithms:*** The choice of software (e.g., GeneMark, AUGUSTUS, GENSCAN) affects prediction quality based on their specific algorithms.
7. ***Regulatory Elements:*** Inclusion of upstream and downstream regulatory regions can enhance prediction accuracy.

8. Write down the aims of proteomics.

Aims of Proteomics:

1. ***Identify and Catalog Proteins:***
   To identify all the proteins expressed in a cell, tissue, or organism at a given time.
2. ***Determine Protein Functions:***
   To understand the biological roles of proteins and how they contribute to cellular processes.
3. ***Study Protein-Protein Interactions:***
   To explore how proteins interact with each other in pathways and networks.
4. ***Analyze Post-Translational Modifications:***
   To detect chemical modifications (like phosphorylation or glycosylation) that regulate protein activity.
5. ***Compare Protein Expression Patterns:***
   To study how protein levels change under different conditions, diseases, or treatments.
6. ***Support Drug Discovery and Biomarker Identification:***
   To find proteins that can serve as disease markers or therapeutic targets.

9. Write properties of gene super families.

Properties of Gene Superfamilies:

1. ***Shared Ancestry:*** Gene superfamilies consist of genes that evolved from a common ancestral gene through duplication and divergence.
2. ***Structural Similarity:*** Members often have conserved structural domains or motifs, even if their functions differ.
3. ***Functional Diversity:*** Despite structural similarities, superfamily members may perform a wide range of biological functions.
4. ***Sequence Homology:*** They exhibit significant sequence similarity, especially in key functional regions.
5. ***Widespread Occurrence:*** Gene superfamilies are often found across many species, indicating their evolutionary importance.
6. ***Evolutionary Adaptability:*** Superfamilies expand and evolve through gene duplication, allowing organisms to adapt to new functions or environments.

10. Name types of symmetry of proteins.

Types of symmetry in proteins include:

1. ***Cyclic Symmetry (Cn):***
   Proteins have rotational symmetry around a single axis, with n identical subunits arranged in a circle.
2. ***Dihedral Symmetry (Dn):***
   Combines a cyclic symmetry axis with a perpendicular twofold axis, forming a more complex symmetrical arrangement.
3. ***Polyhedral Symmetry:***
   Includes symmetry like tetrahedral, octahedral, and icosahedral, often seen in viral capsids.
4. ***Helical Symmetry:***
   Subunits are arranged in a spiral or helix, common in filamentous proteins.

5. ***Asymmetric (No Symmetry):***
   Some proteins do not exhibit any symmetry in their structure.

11. What are factors of genome assembly?

Factors affecting genome assembly include:

1. ***Read Length:*** Longer sequencing reads make it easier to assemble genomes accurately, especially across repetitive regions.
2. ***Sequence Coverage (Depth):*** Higher coverage provides more data to ensure the genome is well-represented and reduces gaps.
3. ***Genome Complexity:*** Presence of repetitive sequences, heterozygosity, and polyploidy can complicate assembly.
4. ***Sequencing Errors:*** Errors in raw data can lead to misassemblies or gaps.
5. ***Assembly Algorithms:*** Different software and algorithms handle data differently, affecting assembly quality.
6. ***Computational Resources:*** Large genomes require significant memory and processing power for assembly.
7. ***Quality of Input Data:*** High-quality, contamination-free sequences improve assembly accuracy.

12. Define Alpha helix structure of protein.

The alpha helix is a common secondary structure of proteins characterized by a right-handed coiled or spiral shape. In this structure, the polypeptide backbone forms a helix stabilized by hydrogen bonds between the carbonyl oxygen of one amino acid and the amide hydrogen of an amino acid four residues earlier. This regular pattern of hydrogen bonding gives the alpha helix a stable and rigid structure. The side chains of the amino acids project outward from the helix, allowing interactions with other molecules or parts of the protein. The alpha helix plays a crucial role in the overall folding and function of many proteins.

13. Explain biological processes of proteins.

Biological processes of proteins encompass a wide range of essential functions that sustain life. Proteins act as enzymes to catalyze biochemical reactions, speeding up processes like digestion and metabolism. They serve as structural components, providing support and shape to cells and tissues—for example, collagen in connective tissues. Proteins also play key roles in transporting molecules across cell membranes and within the bloodstream, such as hemoglobin carrying oxygen. They function in cell signaling and communication by acting as receptors and messengers, helping cells respond to their environment. Additionally, proteins regulate gene expression and control the cell cycle, ensuring proper growth and division. Immune response is another critical area where proteins, like antibodies, identify and neutralize pathogens. Overall, proteins are fundamental to virtually all biological processes, from energy production to defense mechanisms.

14. Write algorithm of splice junction.

Algorithm for Splice Junction Prediction (Simplified Overview):

1. ***Input:***
   Obtain the DNA sequence containing potential splice sites.
2. ***Identify Candidate Sites:***
   Scan the sequence to find consensus splice site motifs, typically:
   - ***Donor site (5' splice site):*** Usually contains the "GT" dinucleotide at the exon-intron boundary.
   - ***Acceptor site (3' splice site):*** Usually contains the "AG" dinucleotide at the intron-exon boundary.
3. ***Score Splice Sites:***
   Use position-specific scoring matrices (PSSMs) or machine learning models to assign scores to candidate donor and acceptor sites based on known splice site patterns.
4. ***Filter Sites:***
   Remove low-scoring or unlikely sites to reduce false positives.
5. ***Predict Exon-Intron Boundaries:***
   Pair donor and acceptor sites to predict possible introns and exons.
6. ***Validate Predictions:***
   Check predictions against known biological features (e.g., branch point sequences, polypyrimidine tract) and splice site strength.
7. ***Output:***
   Provide predicted splice junction positions within the DNA sequence.

15. Write steps of genome mapping.

Steps of Genome Mapping:

1. ***Sample Preparation:***
   Isolate high-quality DNA from the organism of interest.
2. ***Marker Selection:***
   Choose genetic markers such as Restriction Fragment Length Polymorphisms (RFLPs), Simple Sequence Repeats (SSRs), or Single Nucleotide Polymorphisms (SNPs) for mapping.
3. ***Genotyping:***
   Analyze DNA samples to detect the presence or absence of selected markers in different individuals or populations.
4. ***Linkage Analysis:***
   Determine the relative positions of markers by studying how often they are inherited together, calculating recombination frequencies.
5. ***Constructing Linkage Maps:***
   Arrange markers on chromosomes based on recombination data to create a genetic map showing marker order and distances.
6. ***Physical Mapping (optional):***
   Use techniques like Fluorescence In Situ Hybridization (FISH) or restriction mapping to assign physical locations to markers.

7. ***Map Validation:***
   Confirm accuracy by comparing with known maps or using additional markers.
8. ***Data Analysis and Interpretation:***
   Use the map to identify gene locations, study genome organization, and assist in breeding or research programs.

16. Name tools to find genes in eukaryotes.

Here are some commonly used tools for gene prediction in eukaryotes:

1. **GENSCAN**
2. **Augustus**
3. **GeneMark-ES**
4. **FGENESH**
5. **GlimmerHMM**
6. **SNAP**
7. **BRAKER**

17. Explain design of siRNA.

Design of siRNA (small interfering RNA):

Designing effective siRNA involves creating a short double-stranded RNA molecule, usually about 21-23 nucleotides long, that can specifically target and silence a gene of interest. The key steps include:

1. ***Target Selection:***
   Choose a unique sequence within the mRNA of the gene to avoid off-target effects. Typically, regions within the coding sequence or untranslated regions are targeted.
2. ***Sequence Criteria:***
   - Select sequences starting with an AA dinucleotide for better processing.
   - Avoid sequences with high GC content (ideal range is 30-50%) to ensure proper stability and binding.
   - Avoid repetitive or homologous regions to prevent unintended gene silencing.
3. ***Strand Design:***
   The siRNA consists of a sense (matching the target mRNA) and an antisense strand (complementary). The antisense strand is incorporated into the RNA-induced silencing complex (RISC) to guide gene silencing.
4. ***Thermodynamic Stability:***
   Design siRNA with lower stability at the 5' end of the antisense strand to facilitate RISC incorporation.
5. ***Chemical Modifications (optional):***
   To improve stability and reduce immune responses, modifications can be added.
6. ***Validation:***
   Test siRNA efficacy experimentally to confirm gene silencing efficiency and specificity.

18. Explain quaternary structure of protein.

The quaternary structure of a protein refers to the arrangement and interaction of multiple polypeptide chains (called subunits) in a multi-subunit protein complex. Unlike primary, secondary, and tertiary structures—which describe the structure of a single polypeptide—quaternary structure involves the association of two or more folded protein chains into a functional unit. These subunits can be identical (homomeric) or different (heteromeric), and are held together by non-covalent interactions such as hydrogen bonds, ionic bonds, hydrophobic interactions, and sometimes disulfide bridges. The quaternary structure is crucial for the protein's biological function, as it often allows for cooperative behavior, regulation, and structural stability. **Example:** Hemoglobin is a classic example; it has four subunits (two alpha and two beta chains) that work together to transport oxygen efficiently in the blood.

19. Why peptide bond is rigid and planer?

The **peptide bond** is **rigid and planar** due to the nature of its chemical structure. It forms between the **carboxyl group** of one amino acid and the amino group of another, creating a partial double bond character between the carbon and nitrogen atoms. This partial double bond arises from resonance, where electrons are delocalized between the carbonyl oxygen and the amide nitrogen. As a result, the bond cannot freely rotate, making it rigid, and the six atoms involved in the bond (C, O, N, H, and the two alpha-carbons) lie in the same plane, making the structure planar. This rigidity and planarity are crucial for maintaining the protein's structure and for forming stable secondary structures like alpha helices and beta sheets.

20. Write 5 goals of population genomics.

Here are five key goals of population genomics**:**

1. ***Understand Genetic Variation:***
   To study the genetic differences within and between populations to reveal how variation is distributed.
2. ***Trace Evolutionary History:***
   To reconstruct the evolutionary relationships and demographic history (e.g., migration, population size changes) of species.
3. ***Identify Adaptive Traits:***
   To detect genes under natural selection that contribute to adaptation and survival in different environments.
4. ***Associate Genes with Traits or Diseases:***
   To link specific genetic variants with phenotypic traits or susceptibility to diseases.
5. ***Conserve Biodiversity:***
   To support conservation efforts by understanding genetic diversity, inbreeding, and population structure in endangered species.

21. Write 3 ORF finders.

Three commonly used ORF (Open Reading Frame) finders**:**

1. ***NCBI ORF Finder*** – A web-based tool provided by NCBI to identify ORFs in a DNA sequence.
2. ***EMBOSS getorf*** – A command-line tool from the EMBOSS suite for finding and extracting ORFs.
3. ***GeneMark*** – A widely used tool for gene prediction and ORF identification, especially in prokaryotes.

22. How Alpha helix structure of protein is stabilized?

The alpha helix structure of a protein is stabilized primarily by hydrogen bonds. These bonds form between the carbonyl oxygen (C=O) of one amino acid and the amide hydrogen (N-H) of an amino acid located four residues ahead in the sequence.

Additional stabilizing factors:

1. ***Hydrogen Bonding:***
   Regular hydrogen bonds maintain the spiral shape and provide structural stability.
2. ***Compact Packing:***
   The helix allows for tight packing of amino acids, minimizing space and stabilizing the structure.
3. ***Side Chain Orientation:***
   Side chains project outward, reducing steric clashes and allowing favorable interactions with the environment.
4. ***Dipole Moment:***
   The helix has a slight dipole (positive at the N-terminus and negative at the C-terminus), contributing to structural interactions.

23. Write types of transposable elements.

Types of Transposable Elements:

1. ***Class I – Retrotransposons (Copy and Paste):***
   These move via an RNA intermediate. The element is transcribed into RNA, then reverse-transcribed back into DNA and inserted at a new location. Examples are:
   - Long Terminal Repeat (LTR) retrotransposons
   - Long Interspersed Nuclear Elements (LINEs)
   - Short Interspersed Nuclear Elements (SINEs)

2. ***Class II – DNA Transposons (Cut and Paste):***
   These move directly as DNA, using a transposase enzyme to cut the element and insert it elsewhere in the genome. Examples are:
   - Insertion sequences
   - Composite transposons
   - Non-composite transposons

24. Name three categories that NGS assembles based on graphs?

The three main graph-based categories of NGS (Next-Generation Sequencing) assembly are:

1. ***Overlap Graphs:***
   Used in Overlap-Layout-Consensus (OLC) assembly methods, mainly for long-read data. Nodes represent reads, and edges represent overlaps.

2. ***De Bruijn Graphs:***
   Widely used in short-read assemblies. Nodes represent k-mers (subsequences of length k), and edges represent overlaps of k−1 bases.
3. ***String Graphs:***
   A more compact form of overlap graphs, removing redundant overlaps and simplifying the graph structure for better assembly performance.

25. What is directed or undirected graph?

A graph in computational and mathematical terms is a set of nodes (or vertices) connected by edges (or links)**.** Based on the direction of these edges, graphs are classified as:

1. ***Directed Graph (Digraph):***
   In a directed graph, the edges have a direction. Each edge goes from one node to another specific node, indicating a one-way relationship. **Example:** In a social network where A follows B, but B doesn't follow A.
2. ***Undirected Graph:***
   In an undirected graph, the edges have no direction. The connection between nodes is mutual or bidirectional. **Example:** In a road map where roads connect cities in both directions.

26. Write advantages to study proteomics.

Advantages of Studying Proteomics:

1. ***Understanding Protein Function:***
   Proteomics helps reveal the roles of proteins in cellular processes, beyond what gene expression alone can show.
2. ***Detection of Post-Translational Modifications:***
   It identifies chemical changes (e.g., phosphorylation, glycosylation) that regulate protein activity and function.
3. ***Biomarker Discovery:***
   Proteomics aids in finding disease-related proteins that can be used for early diagnosis, prognosis, or treatment monitoring.
4. ***Drug Target Identification:***
   It helps discover and validate proteins that can be targeted by new drugs.
5. ***Insight into Disease Mechanisms:***
   By comparing protein expression in healthy and diseased states, researchers can understand the molecular basis of diseases.
6. ***Systems Biology Integration:***
   Proteomics contributes to a holistic view of biological systems when combined with genomics, transcriptomics, and metabolomics.

27. Explain Human Mitochondrial genome.

The human mitochondrial genome is a small, circular DNA molecule found in the mitochondria — the energy-producing organelles of cells. Unlike nuclear DNA, which is inherited from both parents, mitochondrial DNA (mtDNA) is maternally inherited**.**

Key Features:

1. ***Size and Structure:***
   The mitochondrial genome is about 16,569 base pairs long and exists as a double-stranded circular DNA molecule.
2. ***Gene Content:***
   It encodes 37 genes**:**
   - 13 protein-coding genes (involved in oxidative phosphorylation, the energy-generating process)
   - 22 tRNA genes (for mitochondrial protein synthesis)
   - 2 rRNA genes **(**12S and 16S rRNA)
3. ***No Introns:***
   Unlike nuclear genes, mitochondrial genes generally lack introns and are tightly packed.
4. ***D-loop Region:***
   A non-coding region involved in the control of replication and transcription of mtDNA.
5. ***High Mutation Rate:***
   mtDNA has a higher mutation rate than nuclear DNA, making it useful in studies of evolution, ancestry, and forensic science.
6. ***Independent Replication:***
   Mitochondria can replicate their DNA independently of the cell cycle.

28. Write principle of pyrosequence.

Principle of Pyrosequencing:

Pyrosequencing is a DNA sequencing method based on the detection of pyrophosphate (PPi) released during DNA synthesis. When a nucleotide is incorporated into the growing DNA strand by DNA polymerase, a molecule of PPi is released. This PPi is then converted into a detectable light signal through a series of enzymatic reactions.

Steps involved:

1. ***DNA Polymerization:***
   DNA polymerase adds a complementary nucleotide to the template strand, releasing pyrophosphate.
2. ***Conversion of Pyrophosphate:***
   The enzyme ATP sulfurylase converts pyrophosphate and adenosine 5' phosphosulfate (APS) into ATP.
3. ***Light Generation:***
   The enzyme luciferase uses the ATP to convert luciferin to oxyluciferin, producing visible light proportional to the amount of ATP.

4. ***Signal Detection:***
   The light is detected by a camera or sensor, and the intensity corresponds to the number of nucleotides incorporated.
5. ***Nucleotide Addition Cycle:***
   Nucleotides are added sequentially, and light emission indicates which nucleotide was incorporated.

29. Name of 3 omic areas and define them.

Here are three major omic areas along with their definitions:

1. ***Genomics:*** The study of the complete set of DNA (genome) in an organism, including gene sequencing, structure, function, and mapping. It focuses on understanding genetic information and variation.

2. ***Proteomics:*** The large-scale study of proteins, including their expression, structure, function, and interactions. Proteomics reveals how proteins carry out biological processes and respond to environmental changes.

3. ***Metabolomics:*** The comprehensive analysis of metabolites (small molecules) within cells, tissues, or organisms. Metabolomics helps understand metabolic pathways and physiological states by measuring the end products of cellular processes.

30. Define divergence and convergence path.

Divergence Path:

A divergence path refers to a process or pattern where a single entity splits or branches out into multiple different forms, types, or directions. In biology, it often describes how species evolve from a common ancestor and develop distinct traits over time.

Convergence Path:

A convergence path is when different entities independently evolve similar traits or features, often due to adapting to similar environments or functions, despite having different ancestral origins. This is known as convergent evolution in biology.

31. Define Eulerian path and Eulerian circuit?

Eulerian Path:

An Eulerian path is a path in a graph that visits every edge exactly once but does not have to start and end at the same vertex.

Eulerian Circuit:

An Eulerian circuit is an Eulerian path that starts and ends at the same vertex, thus forming a closed loop while visiting every edge exactly once.

32. Write a note on Automated Sequencing?

Automated Sequencing:

Automated sequencing refers to the use of machines and advanced technologies to rapidly and accurately determine the sequence of nucleotides in DNA. This method revolutionized genetic research by speeding up sequencing processes and reducing manual labor and errors.

Key Features:

1. ***Fluorescent Labeling:***
   Instead of radioactive markers, automated sequencing uses fluorescently labeled nucleotides. Each of the four DNA bases is tagged with a different colored dye.
2. ***Capillary Electrophoresis:***
   DNA fragments generated during sequencing are separated by size using capillary electrophoresis, which allows high-resolution and fast separation.
3. ***Laser Detection:***
   As fragments pass a detector, a laser excites the fluorescent labels, and the emitted light is recorded by a sensor, identifying the base at each position.
4. ***Data Analysis:***
   The fluorescent signals are converted into digital sequences by specialized software, producing a readable DNA sequence output.

33. Write problems of 2D SDS-PAGE.

Problems of 2D SDS-PAGE:

1. ***Limited Dynamic Range:*** It can have difficulty detecting very low-abundance proteins when high-abundance proteins dominate the gel.
2. ***Poor Resolution of Hydrophobic and Very Large or Small Proteins:*** Membrane proteins and extremely large or small proteins often do not resolve well.
3. ***Reproducibility Issues:*** Variability between gels can make comparing results across experiments challenging.
4. ***Labor-Intensive and Time-Consuming:*** The process involves multiple steps and can take several days.
5. ***Protein Modification Detection Limitations:*** Post-translational modifications may be missed or not well-resolved.
6. ***Protein Solubility Problems:*** Some proteins may precipitate or be lost during sample preparation.

34. What is Gene cluster?

A gene cluster is a group of two or more genes located close together on a chromosome that often have related functions or are involved in the same biological pathway. These genes are usually co-regulated and can be transcribed together or controlled by shared regulatory elements. Examples:

- The Hox gene cluster, which controls body plan development in animals.
- Operons in bacteria, like the lac operon, where multiple genes are transcribed as a single mRNA.

35. Differentiate between Dicer and Drosha.

Drosha and Dicer are both enzymes involved in the microRNA (miRNA) processing pathway, but they function at different stages and locations within the cell. Drosha is an RNase III enzyme that operates in the nucleus, where it cleaves primary miRNA transcripts (pri-miRNA) into shorter precursor miRNA (pre-miRNA) hairpins. This cropping step is the first crucial phase of miRNA maturation. Afterward, the pre-miRNA is transported to the cytoplasm, where Dicer, another RNase III enzyme, further processes it by cutting the pre-miRNA into mature miRNA or small interfering RNA (siRNA) duplexes. These mature RNA molecules are then incorporated into the RNA-induced silencing complex (RISC) to regulate gene expression. Thus, Drosha initiates miRNA biogenesis in the nucleus, while Dicer completes the process in the cytoplasm.

36. Write method used for protein extraction.

Method Used for Protein Extraction:

Protein extraction involves isolating proteins from cells or tissues for further analysis. The general steps include:

1. ***Sample Preparation:***
   Collect the biological sample (cells, tissues) and keep it cold to prevent protein degradation.
2. ***Cell Lysis:***
   Break open the cells to release proteins using physical methods (sonication, grinding, freeze-thaw cycles) or chemical methods (detergents, enzymes, or osmotic shock).
3. ***Use of Extraction Buffer:***
   Apply a suitable buffer containing salts, pH stabilizers, protease inhibitors, and detergents to solubilize proteins and protect them from degradation.
4. ***Centrifugation:***
   Spin the lysate at high speed to separate soluble proteins (supernatant) from insoluble debris (pellet).
5. ***Protein Concentration and Purification:***
   Depending on the purpose, proteins can be further purified or concentrated using precipitation, dialysis, or chromatography techniques.

37. Write limitations of OLC.

Limitations of Overlap-Layout-Consensus (OLC) Assembly:

1. ***Computationally Intensive:***
   OLC requires all-against-all comparisons of reads to find overlaps, which becomes very demanding for large datasets, especially with high-throughput sequencing.
2. ***Not Suitable for Short Reads:***
   It performs poorly with short-read data (like Illumina) because short reads have fewer overlaps and more repetitive sequences, making assembly difficult.
3. ***High Memory Usage:***
   Storing and processing overlap graphs for large genomes requires significant memory resources.
4. ***Error Sensitivity:***
   Sequencing errors can cause false overlaps or missed overlaps, complicating the assembly process.

5. ***Inefficient for Highly Repetitive Genomes:***
   OLC struggles to resolve repeats due to ambiguous overlaps, leading to fragmented assemblies.

38. Peptide bond separation opt other than 1D and 2D electrophoresis.

Other than one-dimensional (1D) and two-dimensional (2D) electrophoresis, peptide bond separation and protein analysis can be done using these methods:

1. ***Chromatography Techniques:***
   - ***High-Performance Liquid Chromatography (HPLC):*** Separates peptides based on hydrophobicity or charge.
   - ***Size Exclusion Chromatography (SEC):*** Separates proteins and peptides by size.
   - ***Ion Exchange Chromatography:*** Separates based on charge differences.
2. ***Mass Spectrometry (MS):***
   Peptides are ionized and separated based on their mass-to-charge ratio, allowing precise identification and quantification.
3. ***Capillary Electrophoresis (CE):***
   Uses an electric field to separate peptides based on their charge-to-size ratio inside thin capillaries, providing high-resolution separation.
4. ***Ultrafiltration:***
   Uses membranes with specific molecular weight cut-offs to separate peptides from larger proteins.

39. Write name of greedy graph assembler.

A well-known greedy graph assembler is Greedy Path-Merge (GPM).

40. Write advantages of IEF.

Advantages of Isoelectric Focusing (IEF):

1. ***High Resolution:*** IEF can separate proteins or peptides with very small differences in their isoelectric points (pI), offering excellent resolution.
2. ***Sensitive and Precise:*** It allows precise determination of the pI of proteins, useful for characterizing isoforms.
3. ***Wide pH Range:*** Can be performed over a broad pH gradient, enabling separation of diverse proteins.
4. ***Compatibility:*** IEF is often used as the first dimension in 2D gel electrophoresis, complementing size-based separation.
5. ***Quantitative:*** Provides quantitative data on protein charge variants.
6. ***Non-Denaturing:*** Under certain conditions, IEF can be performed without denaturing proteins, preserving their native state.

41. Name types of Spectrophotometer.

Types of spectrophotometers include:

1. Single-beam spectrophotometer
2. Double-beam spectrophotometer
3. UV-Visible spectrophotometer
4. Infrared (IR) spectrophotometer
5. Fluorescence spectrophotometer
6. Atomic absorption spectrophotometer
7. Fourier Transform Infrared (FTIR) spectrophotometer

Disclaimer:

For best results, it is advised NOT to rely on this file only. Go through video lectures, handouts and other sources as well.

Regards,

Sir Mughal,

BS Zoology, Virtual University of Pakistan

B.Sc (Honors), University of Karachi.