GOVERNMENT AUTONOMOUS COLLEGE(A), RAJAMAHENDRAVARAM
Department of Microbiology
Question bank, Semester - 4, II YEAR B.Sc. Microbiology Honours
Major **C 11** rDNA Technology, Bioinformatics and Biostatistics

UNIT- 1 Recombinant DNA Technology

Essays (8 Marks each)

1. What is Gene cloning? What are the steps involved in Gene Cloning

ANSWER : **GENE CLONING**

Gene cloning is a process of recombinant DNA technology, that combines DNA from different sources (Organisms)

STEPS INVOLVED IN GENE CLONING

1. **Isolation of Genetic material (DNA):** Extract the DNA from the organism using physical or chemical methods.
   Example: Extraction of the human insulin gene and plasmid DNA from E. coli.

2. **Cutting of DNA with Restriction Enzymes**: Use restriction enzymes to cut DNA at specific sequences called restriction sites. Restriction enzymes are used to cut both the vector plasmid and the DNA fragment to create complementary sticky or blunt ends.
   Example: EcoRI cuts DNA at 5'-GAATTC-3'.

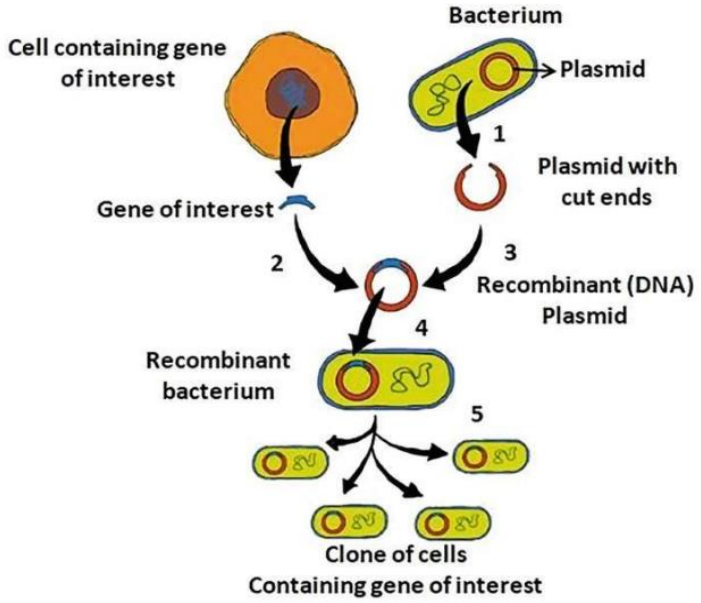
3. **Amplification of DNA:** The desired DNA fragment is amplified using Polymerase Chain Reaction (PCR).
   Example: Amplification of the insulin gene for diabetes treatment.

4. **Ligation of DNA into Vectors:** The Gene of interest (DNA) fragment is inserted into the plasmid vector and the gaps are sealed using the enzyme DNA ligase, creating recombinant DNA.
   Example: Ligating the human insulin gene into the pBR322 vector.

5. **Introduction of Recombinant DNA into Host Cells**: The recombinant DNA is transferred into a host organism for propagation.
   Example: Introducing rDNA into E. coli using heat shock.

6. **Selection of Transformed Cells:** Transformed cells are identified by using selectable markers like antibiotic resistance genes.
   Example: Antibiotic resistance markers (ampicillin resistance) used to select transformed cells. Plating transformed E. coli on agar containing ampicillin; only cells with the plasmid grow.

7. **Expression of the Inserted Gene:** The host cells are cultured to produce the protein by the inserted gene.
   Example: E. coli produces recombinant human insulin (Protein)



Applications of Gene Cloning:

1. A particular gene can be isolated and its nucleotide sequence determined.
2. Control sequences of DNA can be identified & analyzed.
3. Protein/enzyme/RNA function can be investigated.
4. Mutations can be identified, e.g. gene defects related to specific diseases Organisms can be 'engineered' for specific purposes, e.g. insulin production, insect resistance, etc.

---

2. Explain different Gene transfer methods

ANSWER: METHODS OF GENE TRANSFER

Gene transfer methods are techniques used to introduce foreign DNA into host cells. These methods can be broadly classified into physical, chemical, and biological methods.

PHYSICAL METHODS

1. **Microinjection:** A fine needle injects DNA directly into a cell.

   Example: DNA injected into an animal egg for research.

2. **Gene Gun**: Shoots tiny particles coated with DNA into plant cells.

   Example: Creating genetically modified (GM) crops like golden rice.

3. **Electroporation:** An electric field creates pores in the cell membrane to let DNA enter.

   Example: Introducing genes into bacteria or animal cells. Non-toxic and non-immunogenic.

CHEMICAL METHODS

Chemical methods use chemical agents to enhance the uptake of DNA by cells. These agents modify the cell membrane for DNA entry.

1. **Calcium Phosphate Precipitation**: DNA is mixed with calcium chloride and phosphate buffer to form a fine precipitate that adheres to the cell surface and is internalized by endocytosis.

2. **Polyethylene Glycol (PEG)-Mediated Transfer :** PEG promotes the fusion of the cell membrane with DNA-containing vesicles or liposomes.
3. **DEAE-Dextran Method** : Positively charged DEAE-dextran binds negatively charged DNA and facilitates its uptake by cells.
4. **Cationic Lipid-Mediated Transfer (Lipofection)** : Cationic lipids form complexes with DNA and facilitate its fusion with the cell membrane for internalization.

BIOLOGICAL METHODS

Biological methods utilize natural biological systems such as viruses, bacteria, or other cellular mechanisms to transfer DNA into target cells.

1. **Plasmids:** Small, circular, double-stranded DNA molecules found in bacteria and some eukaryotes. Widely used as vectors to carry foreign DNA into host cells.

   Examples: pBR322, pUC series.
2. **Bacteriophages** : Viruses that infect bacteria. Vectors for introducing DNA into bacterial cells.

   Examples: λ-phage and M13 phage.
3. **Cosmids** : Hybrid vectors combining features of plasmids and bacteriophage λ DNA. • Designed to carry larger DNA fragments (up to 45 kb).
4. **Bacterial Artificial Chromosomes (BACs)** : Large plasmids modified to carry up to 300 kb of foreign DNA. Ideal for cloning large DNA fragments in bacterial hosts.
5. **Yeast Artificial Chromosomes (YACs)** : Synthetic chromosomes that replicate in yeast cells. Capable of carrying very large DNA fragments (up to 1 Mb). Suitable for eukaryotic DNA studies.
6. **Agrobacterium-Mediated Transfer** : Agrobacterium tumefaciens transfers T-DNA from its Ti plasmid into plant cells.

Example: Used in creating genetically modified plants like Bt cotton.

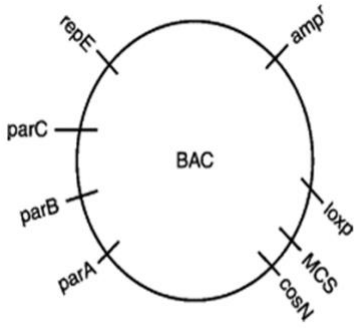
---

3. Discuss about BAC and YAC and explain their role in Genetic Engineering

ANSWER: BACTERIAL ARTIFICIAL CHROMOSOMES (BACs)

BACs are engineered vectors derived from the F-factor of bacteria (Escherichia coli), capable of cloning large DNA fragments up to 100-300 kb in size.

- These vectors maintain a supercoiled circular form, promoting stability and reducing insert chimerism during propagation.



Structure of BAC

1. Origin of replication (ori) for replication in bacteria.
2. Antibiotic resistance gene for selection.
3. Par genes for stable inheritance during cell division.

Role of BAC in Genetic Engineering

Cloning Large DNA Fragments: BACs are used to clone large pieces of DNA, such as genes or entire genomic regions.

Example: Human genome fragments were cloned in BACs during the Human Genome Project.

1. **Gene Mapping:** BAC libraries are used for physical mapping of genomes to determine gene locations.
2. **Transgenics:** BACs can carry regulatory elements and large genes for studying their function in model organisms.
   Example: Studying human genes in mice.
3. **Therapeutic Development:** Used to study genes related to diseases and potential treatments.

Advantages of BACs

1. High stability in host cells.
2. Can replicate and maintain large DNA fragments.
3. Easier to manipulate than YACs.

YEAST ARTIFICICAL CHROMOSOMES (YACs)

- YACs are eukaryotic vectors that can clone extensive DNA segments, up to 500 kb-2000kb, within yeast cells such as Saccharomyces cerevisiae.
- They are particularly useful for cloning large genomic sequences and studying gene function in a eukaryotic context.
- It can carry larger DNA fragments (up to 1,000 kilobases or 1 megabase).

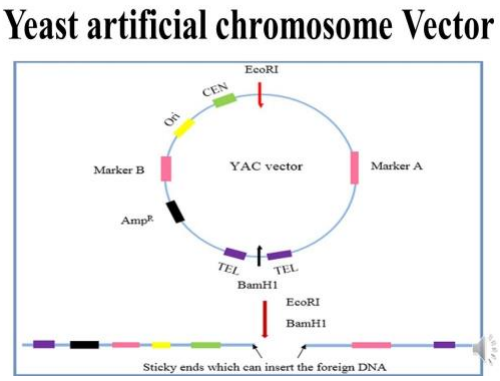
Structure of YAC •

1. Telomeres (for chromosome stability).
2. Centromere (for proper segregation during cell division).
3. Autonomously replicating sequence (ARS) for replication.
4. Selectable marker genes for yeast growth.

Role of YAC in Genetic Engineering

1. **Cloning Very Large DNA Fragments:** YACs are used to clone large genes, entire chromosomes, or other large DNA fragments.
2. **Genome Studies:** YAC libraries allow for the study of complete genomes, including eukaryotic systems. **Example:** Used in the Human Genome Project.
3. **Studying Complex Genes**: YACs can carry genes along with their regulatory elements, enabling the study of gene expression and regulation.
4. **Modeling Genetic Diseases**: YACs are used to study diseases.
   Example: Studying Huntington's disease by introducing the large huntingtin gene into yeast or mice.

5. **Advantages of YACs** Can clone larger DNA fragments than BACs.
    1. Useful for studying eukaryotic genes with complex regulatory regions. Limitations of YACs
    2. Less stable than BACs due to recombination events.
    3. Difficult to manipulate compared to BACs.



Short Answers

1. What are restriction enzymes?

Ans: Restriction enzymes, are enzymes initially isolated from bacteria that cut DNA at sequence-specific sites, producing known DNA fragments.

- They do not discriminate between the DNA of bacteria, fungi, mice, or humans. If they recognize their target site, they cut.
- These enzymes are widely found in bacteria and archaea, providing them with defense against invading bacteriophages by cutting Viral DNA. Werner Arber, Hamilton O. Smith, and Daniel Nathans discovered and characterized them in the late 1960s and early 1970s.

2. What are cosmids?

Ans: Cosmids are hybrid vectors that combine features of plasmids and bacteriophage λ, enabling the incorporation of larger DNA fragments, typically up to 42 kb.

- These vectors are constructed by integrating the cos region of a phage vector into a plasmid backbone, allowing for efficient replication within host cells.
- Cosmids are particularly advantageous for cloning larger DNA sequences that exceed the capacity of standard plasmids, making them valuable in genomic library construction.

3. What are BACs

Ans: BACs are engineered vectors derived from the F-factor of bacteria (Escherichia coli), capable of cloning large DNA fragments up to 100-300 kb in size.

- These vectors maintain a supercoiled circular form, promoting stability and reducing insert chimerism during propagation.

4. What is Genetic Engineering ?

Ans: Genetic Engineering refers to the techniques used to manipulate genetic material to alter, add, repair, enhance or suppress the form and functioning of that particular molecule.

- It is the process of altering the genetic material of an organism by introducing, removing, or modifying specific genes to achieve desired traits.
- This involves combining DNA from different organisms to produced genetically modified organisms (GMO).
- Gregor Mendel is the “father of Genetics”

5. What is DNA polymerase?

Ans: DNA polymerases are enzymes that synthesize new copies of DNA (the double-stranded genetic material) of our cell. Just before a cell divides during cell division, DNA polymerases copy the double-stranded parent strand into two identical new DNA molecules by a process known as DNA replication. These enzymes add nucleotides to the growing chain one by one, complementing the opposite strand.

**********UNIT 1 END**********

UNIT – 2 Applications of rDNA technology

Essays (8 Marks)

1. Explain about various applications of Genetic engineering

Ans: APPLICATIONS OF GENETIC ENGINEERING IN INDUSTRY, MEDICINE & AGRICULTURE

Genetic Engineering in industry

1. Protein Engineering:

Protein engineering involves insertion of chemically synthesised DNA into desired organisms to produce modified proteins. These techniques made it easy to alter one or few amino acids in a protein and thus alter its structure and behaviour. In this way, enzymes and bioactive peptides used in different industries can be created with different characteristics.

2. Biodegradable Plastic Industry:

- Biodegradable plastics like polyhydroxylbutyrate (PHB) can be obtained commercially by fermentation with the bacterium Alcaligenes eutrophus.But its production cost is very high.

3. Recently model plant Arabidopsis is produced for plastic production(PHB globules) in their chloroplasts without affecting plant's growth and development. Industry has already started to explore the production of biodegradable plastics from other transgenic tree plants.

4. **Oil Industry:** Plants store oil in their seeds (e.g., ground nut, mustard, rapeseed, sunflower, sesamum, soya ben etc.) or in fruits (e.g., olive, avocado, oil palm etc.). Such vegetable oils are used either as food or in industrial purposes. According to various requirements the fatty acid quality and yield can be improved by using genetic engineering technology.

5. Fuel Industry:

- In recent years, ethanol has found its use as an important chemical feedstock and as a fuel supplement. Ethanol is generally produced by fermentation of some sugar (starch, cellulose) rich products with the help of yeast.
- At present E. coli and Klebsiella planticola carrying genes from Z. mobilis have been developed which could utilize glucose and xylose as the substrate to give maximum yield of ethanol.

Genetic Engineering in Agriculture

1. **Golden Rice:** It is variety of rice genetically modified produce beta carotene, a precursor to vitamin A, to combat vitamin A deficiency in developing countries.
2. **Bt Crops (e.g., Bt Cotton, Bt Corn)**

These crops are modified to express a protein from Bacillus thuringiensis (Bt) that acts as a pesticide, protecting plants from pests like the bollworm.

6. **Flavr Savr Tomato**

It is a genetically modified tomato designed to delay ripening and enhance shelf life

Genetic Engineering in Medicine

1. **Insulin Production**

Human insulin is produced using genetically engineered bacteria (e.g., Escherichia coli). It is used in diabetes treatment.

2. **CRISPR-based Gene Editing**

This technology allows precise editing of DNA, helping in treating genetic disorders like sickle cell anemia and cystic fibrosis.

3. **Gene Therapy for Genetic Disorders –** In this faulty genes in patients are corrected or replaced. Example severe combined immunodeficiency (SCID) or spinal muscular atrophy (SMA).

---

2. Explain PCR technique and steps involved in PCR

Ans: Polymerase Chain Reaction (PCR) is a nucleic acid amplification technique used to amplify the DNA or RNA in vitro enzymatically.

- PCR is developed in mid-1980s by Kary Mullis and his associates. He was awarded the Nobel Prize in chemistry with Michael Smith in 1993. Since then, it has been the most important tool in molecular biology and genetics as a basic tool for DNA and RNA analysis.

PRINCIPLE OF PCR

PCR combines the principle of nucleic acid hybridization with the principle of nucleic acid replication with the temperature variations of cyclic heating and cooling throughout the process.

- It is based on the fact that double-stranded DNA can be thermally broken into single-stranded DNA segments.
- In these ssDNA templates, primers can anneal to their complementary sequences based on the nucleic acid hybridization principle.
- DNA polymerase then elongates the primer by sequentially adding the nucleotides to the 3' end and generates a dsDNA following the principle of DNA replication.
- By continuously regulating the temperature, these 3 steps; denaturation, annealing, and elongation, can be permitted to occur in a cyclic manner leading to millions of copies of a single targeted nucleic acid sequence.

Components needed in PCR technique

1. **Nucleic Acid Template (Template DNA)**
2. **DNA Polymerase:** *Taq DNA polymerase*

3. **Primers**
4. **Nucleotide triphosphates**
5. **PCR** buffers
6. **Thermocycler**

Steps in PCR

Step I: Amplification :

- It is the main reaction process occurring in PCR. The double stranded DNA is taken as sample for amplification. The amplification step includes denaturation, annealing, and elongation occurring orderly in a cyclic manner one after another for a certain number of cycles pre-programmed by the user.

1. **Denaturation :** It is the 1st step of the amplification reaction where the double-stranded DNA is thermally denatured into two single-stranded DNA templates. Temperature is raised to about 94°C (90 to 95°C) for about 30 to 90 seconds.
    - dsDNA is converted into 2 ssDNA templates
2. **Annealing**

The primer anneals the ssDNA templates at their complementary sites. The forward primer anneals at the complementary site of the antisense strand, and the reverse primer anneals at the complementary site of the sense strand of the template DNA. For annealing to occur, the temperature is reduced to 55°C-70°C. About 30 to 60 seconds are enough for annealing in most of the PCR processes.

- ssDNA + Forward and reverse primers → ssDNA with annealed primers

3. **Elongation**

It is the final step in the amplification reaction where the temperature is raised to 72°C so that the Taq DNA polymerase enzyme begins synthesizing new DNA strands in the 5' to 3' direction. The DNA polymerase enzyme adds nucleotides from the reaction mixture to the 3' OH- end of the annealed primer forming a new complementary strand.

- At the end of elongation, two new dsDNA will be formed from a single dsDNA template
- 2 ssDNA with annealed primers + dNTPs → 2 new dsDNAs

STEP II: Product Analysis Phase

It is the phase after completion of the PCR where the reaction mixture subjected to PCR is analyzed to confirm that desired amplification is achieved. For this, mostly agarose gel electrophoresis is employed in order to check for amplified DNAs or RNAs. However, no additional step is required in some types of PCR, like real-time PCR.

Applications of PCR

1. **Identification and Classification of Organism**

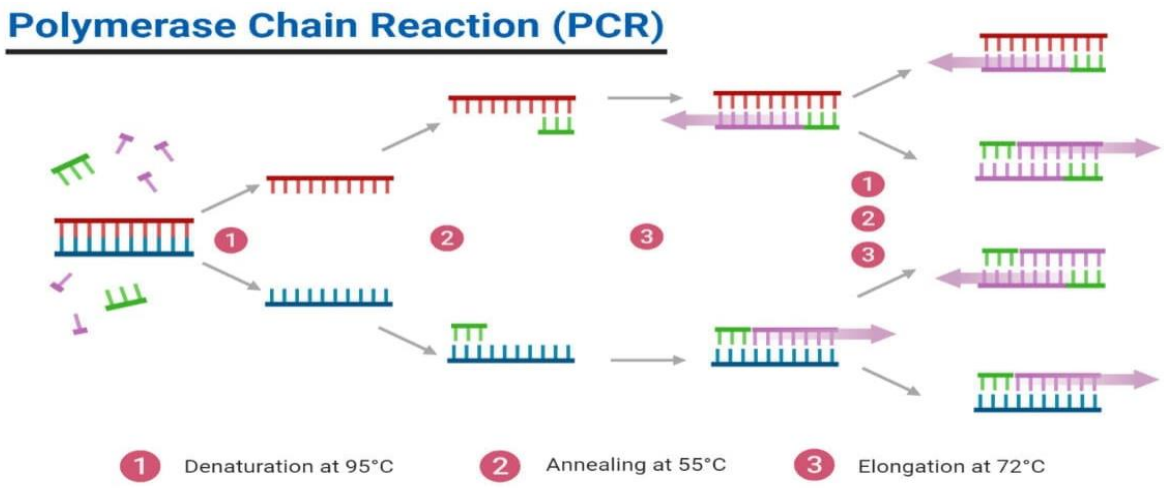
PCR is widely used in identifying microorganisms up to the level of subspecies and strains. This has reduced the time required for microbial identification from days to a few hours. Additionally, larger animals can also be identified and systematically classified using PCR.

2. **Infectious Disease Diagnosis**

The use of PCR in the identification of pathogens has led to the quick and accurate diagnosis of infections. Identification of Antimicrobial resistant genes in the pathogen is also possible

3. **Detection of Gene Mutation and Genetic Disorders**

Mutation in any segment of a gene can be detected using PCR. Knowing this mutation, we can confirm a genetic disorder. Detection of cancerous cells is another very important application of PCR in medicine



.

4. **DNA Fingerprinting**
In forensics, PCR is used for DNA fingerprinting. DNA fingerprinting is used for the identification of criminals or individuals and for confirming parents.

5. **Gene Sequencing**
For gene sequencing, a gene must be amplified into a large number using techniques like PCR.

6. **DNA and RNA Quantification**
PCR can also be used for the quantification of sample DNA and RNA. Quantitative Real-Time PCR (RT – qPCR) is one common type of PCR used for the quantification of sample DNA.

7. **As a Tool in Genetic Engineering**
PCR is used in genetic engineering for analyzing modified DNAs and amplifying target or vector DNA. Desired genes are amplified using PCR and applied in the required process.

8. **Gene Expression Analysis and Genetic Imprinting**
PCR of RNA (Reverse Transcription PCR) is used in gene expression analysis, study genetic imprinting, etc. It is also used in drug and vaccine discovery, human genome projects, paleontology, and evolutionary biology.

---

3. Explain Hybridoma technique and steps involved in hybridoma technique.

Ans: HYBRIDOMA TECHNOLOGY

Hybridoma technology involves fusing short-lived antibody-producing B cells with immortal myeloma cells, creating cell lines that produce a never-ending supply of a specific monoclonal antibody The technique was created in 1975 by Nobel prize-winning scientists Georges Kohler and Cesar Milstein.

Preparation of monoclonal antibodies using hybridoma technology

STEP-1 : Immunization

- In the first step the laboratory animals like rabbits or mice are injected with a selected antigen. So that antibodies are raised against antigen.
- The injection of antigen to mice/rabbit in a series of injections over a period of several weeks is given to stimulate B cell differentiation into plasma B cells and memory B cells.
- Once a sufficient number of antibodies are created in the animal serum. The spleen is removed from the mice.

STEP II: Isolation of B lymphocytes

- The spleen is removed in aseptic conditions to isolate the activated B-cells. The procedure is performed using density gradient centrifugation.
- The presence of antibodies in the Serum is identified using methods like ELISA or flow cytometry. The serum contains the activated B lymphocytes (that produce antibodies). The activated B lymphocytes contain HGPRT genes but with limited life span.

STEP III: Preparation of myeloma cell lines:

- Myeloma cells are metastatic tumor cells that are produced in a laboratory mice/rabbit.
- Myeloma cells have longer life span but fuctional hypoxanthine-guanine phosphoribosyltransferase (HGPRT) gene is absent.
- Non-functional HGPRT metastatic tumor cells are sensitive to HAT media.

STEP IV: Cell fusion:

- The activated B lymphocytes are fused with HAT-sensitive myeloma cells.
- This step is performed by centrifugation of freshly obtained activated B-cells with HAT-sensitive myeloma cells in a fusion-promoting media.

STEP V: Hybridoma selection and screening of hybridoma cells

- Cells that are fused to form hybridoma cells.
- The cells that are not fused together can be separated by incubating the cell mixture in HAT(Selective media) for 10–14 days. HAT medium contains hypoxanthine-aminopterin-thymidine.
- Due to a limited life span, unfused B cells perish within a few days. Unfused malignant neoplastic cells die as a result of the lack of the (HGPRT) gene.
- Therefore, the remaining viable cells left in the media are the hybrid cells.
- After the separation and isolation of different hybridomas, screening is performed for selecting hybridomas that produce the desired antibodies targeting specific epitopes for an antigen

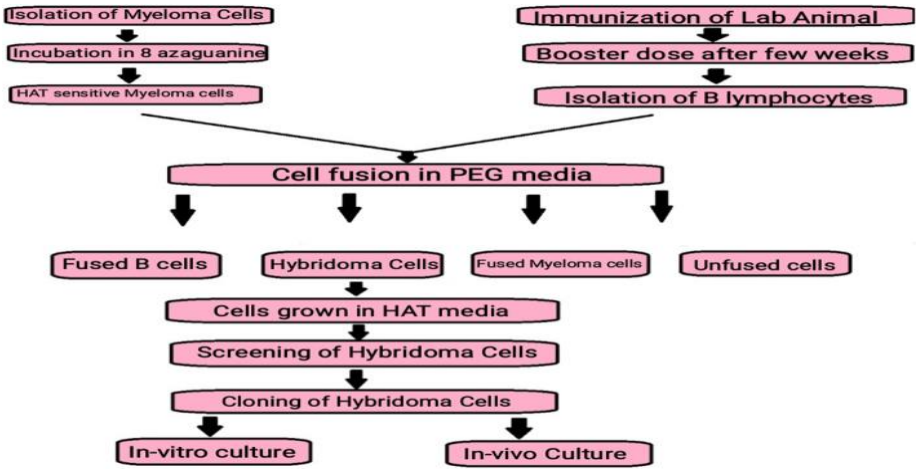
STEP VI : Cloning and Propagation of hybridoma cells

- Hybridomas producing desired antibodies are selected and are then transferred into large culture vessels or flasks.
- The hybridoma cell lines are cultured using in vivo or in vitro methods.

1. **In vivo**
    - The in vivo method uses mice for the production of monoclonal antibodies.
    - Mice are injected intra peritoneally with 105 to 110 viable hybridoma cells.

- After a few weeks, the ascites fluid is collected from an anesthetized mouse to check the presence of monoclonal antibodies.
- These monoclonal antibodies are to be purified.

2. **In vitro**
   - This is another method in which hybridoma cells are cultured in laboratory conditions.
   - It involves growing the hybrid cells in a culture media followed by isolation of monoclonal antibodies from the media.
   - This method is more suitable for the culturing of hybrid cells as it reduces the risk of contamination.
   - In vitro antibody production leads to the production of highly pure antibodies.



Applications of Hybridoma cells

1. **Diagnostic testing:** Monoclonal antibodies are commonly employed in the diagnosis of a variety of disorders. It is used to check the presence of any foreign antigen such as toxins, drugs, hormones, or internal and surface proteins of bacteria or viruses
2. **Testing of pregnancy:** Monoclonal antibodies are used to identify the presence of human chorionic gonadotropin [hCG] as a mark for recognition of pregnancy.
3. **Radioimmunodetection (RID) of cancer:** Monoclonal antibodies are also used to detect the presence of specific tumor-type in the body.
4. **Malaria herpes virus testing:** mAbs are used in the diagnosis of various diseases caused by viruses such as malaria herpes viruses.
5. **Identification of different strains of pathogens:** Monoclonal antibodies can be used to differentiate between different strains of a single pathogen, for example, *Neisseria gonorrhoeae.*

Short answers

1. **Write about Genomic library**

Ans: A genomic library is a collection of DNA fragments that together make up the total genomic DNA of a single organism.

- A single organism's DNA is stored in a population of identical vectors, each vector containing a different DNA fragments of that particular organism.
- The first DNA-based genome ever fully sequenced was achieved by two-time Nobel Prize winner, Frederick Sanger, in 1977. Sanger and his team of scientists created a library of the bacteriophage, phi X 174, for use in DNA sequencing.

2. Write about PCR

Ans: Polymerase Chain Reaction (PCR) is a nucleic acid amplification technique used to amplify the DNA or RNA in vitro enzymatically.

- PCR is developed in mid-1980s by Kary Mullis and his associates. He was awarded the Nobel Prize in chemistry with Michael Smith in 1993. Since then, it has been the most important tool in molecular biology and genetics as a basic tool for DNA and RNA analysis.

3. What is Taq DNA polymerase?

Taq DNA polymerase :

- The DNA polymerase enzyme extracted from the bacterium *Thermus aquaticus*, is the most widely and the best–known DNA polymerase used in PCR since its establishment.
- *Taq* DNA polymerase is thermally stable and continues its activity after the repeated heating and cooling cycle. It is stable up to 95°C and shows the most effective reaction at around 72°C to 78°C incorporating about 60 bases per second.

4. Write about Hybridoma Technology?

Ans: Hybridoma technology involves fusing short-lived antibody-producing B cells with immortal myeloma cells, creating cell lines that produce a never-ending supply of a specific monoclonal antibody The technique was created in 1975 by Nobel prize-winning scientists Georges Kohler and Cesar Milstein.

**********UNIT -2 END**********

UNIT -3 Techniques in genetic engineering and IPR

Essays (8 Marks)

1. E**xplain in detail about DNA sequencing- Sanger's method**

ANSWER: Sanger sequencing is also known as dideoxy sequencing or chain termination method. It is a method that identifies the order of nucleotide bases in DNA strand based on chain termination by modified nucleotides called dideoxynucleotide triphosphates (ddNTPs).

This method was the first DNA sequencing method developed in 1977 by Frederick Sanger and his colleagues.

PRINCIPLE OF SANGER SEQUENCE

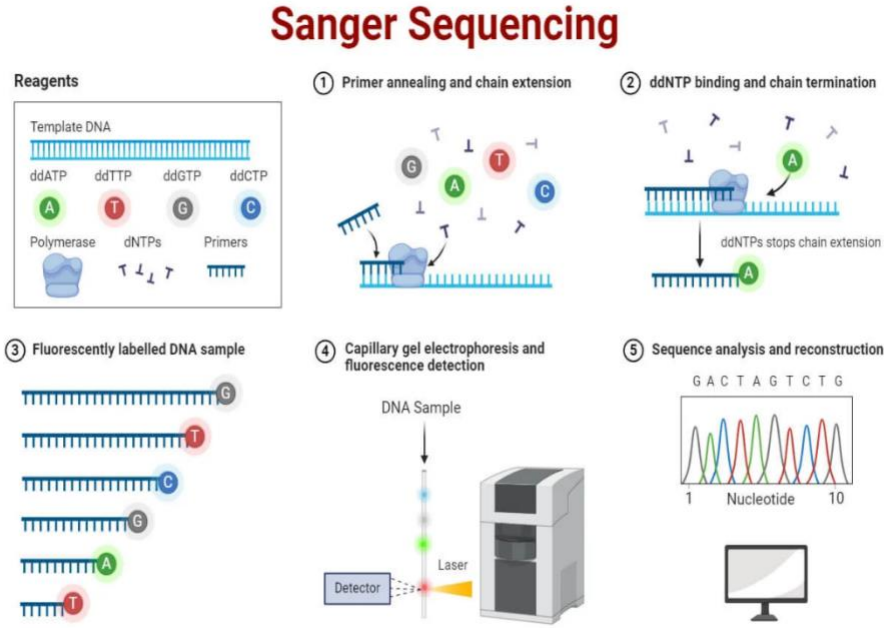
The principle of Sanger sequencing is based on the termination of DNA strand elongation by ddNTPs. These modified molecules are chemical analogs of DNA nucleotides that lack the 3' hydroxyl group necessary for the formation of a phosphodiester bond that elongates the DNA

strand. The addition of ddNTPs in the Polymerase Chain Reaction (PCR) reaction terminates DNA elongation.

During the sequencing process, labeled ddNTPs, dNTPs, and template DNA are mixed in a PCR reaction. When ddNTPs are added, they terminate the DNA chain, producing fragments of different lengths. These fragments are separated by electrophoresis. The fluorescent labels on the ddNTPs indicate which base terminated each fragment and are used to determine the DNA sequence.

STEPS IN SANGER SEQUENCE

- **DNA Template Preparation :** The first step in Sanger sequencing is DNA template preparation. At first, the DNA of interest should be extracted from the source.
- The target DNA is then amplified using PCR which starts with initial denaturation followed by multiple cycles of denaturation, annealing, and extension and ends with a final hold at 4°C.
- The PCR reaction includes template DNA, primers, all four dNTPs, and DNA polymerase. In the traditional sequencing method, the PCR reaction is performed in four separate tubes.



- Along with the dNTPs, each tube is added with one type of labeled ddNTP. Each ddNTP is labeled with a distinct fluorescent marker.
- ddNTPs lack the 3’ hydroxyl group needed for further elongation causing chain termination.
- ddNTPs are present in smaller amounts compared to dNTPs so chain termination occurs at various lengths. This results in a collection of newly synthesized DNA strands of different lengths, each terminating at a ddNTP.
- In automated Sanger sequencing, all four ddNTPs are included in a single reaction, each labeled with a unique fluorescent marker.
- **Separation of DNA Fragments:** Electrophoresis is performed to separate the DNA fragments according to their fragment lengths.

- The DNA fragments are separated either in polyacrylamide gel or capillary gel system. Fragments from each PCR reaction tube are run in separate column to identify the corresponding ddNTP.
- **Detection and Analysis:** After separation, the DNA fragments are passed through a fluorescence detector.
- The detector identifies the fluorescent label attached to the ddNTP at the end of each DNA fragment. Each fluorescent signal represents the nucleotide at the end of the terminated fragment.
- The emitted signals from the nucleotides are captured and a chromatograph is generated, showing the fluorescent peaks of each labeled fragment corresponding to the DNA sequence.

ADVANTAGES OF SANGER SEQUENCE

1. Sanger sequencing is considered the gold standard method for many research and clinical applications as it provides highly accurate results.
2. Sanger sequencing technology is well-established with a straightforward process. This leads to reliable and reproducible outcomes.
3. Sanger sequencing is suitable for small-scale projects or those including short DNA regions.
4. Sanger sequencing produces relatively long read lengths.
5. Data analysis from Sanger sequencing does not require complex bioinformatics tools and expertise
6. The quality of the data obtained from Sanger sequencing is high with clear and easily interpretable chromatograms.

---

2. What are Intellectual Property Rights(IPR) and explain it types
ANSWER:

Intellectual Property Rights or IPR refers to the legal rights that creators or owners have over their Intellectual properties.

Intellectual Property is reffered to the creations of the human mind, exclusive to a single person. It includes inventions, literature and artistic masterpieces, designs, emblems, titles, Music and pictures e.t.c., utilized in commerce and business. Such creations are prone to plagiarism or copying. So, Intellectual Property Rights protect them from unauthorized production, distribution, and display.

Examples of intellectual property rights include:

1. Patents
2. Domain names
3. Industrial design
4. Confidential information
5. Inventions
6. Trademarks
7. Design rights
8. Business or trade names
9. Commercial secrets
10. Computer software, e.t.c.

Types of Intellectual Property

There are four main types of intellectual property rights
Includes

1. **Patents**
2. **Trademarks**
3. **Copyrights**
4. **Trade secrets.**

- Owners of intellectual property frequently use more than one of these types of intellectual property law to protect the same intangible assets.
- **Example:** Trademark law protects a product's name, whereas copyright law covers its tagline.

PATENTS

The U.S. Patent and Trademark Office grants property rights to original inventions, from processes to machines. Patent law protects inventions from use by others and gives exclusive rights to one or more inventors. The three types of patents consist of:

- **Design patents:** Protection for style, shape or design of a device or invention.
  Example : Coca-Cola bottle), emojis, fonts, or any other distinct visual traits.
- **Plant patents:** Safeguards for new varieties of plants. But inventors may also want a design patient if the tree has unique visual properties.
  Example: pest-free versions of fruit trees.
- **Utility patents:** Protection for a product that serves a practical purpose and is useful.
  Example: vehicle safety systems, software, and pharmaceuticals.

TRADE MARKS

- Trademarks protect logos, sounds, words, colors, or symbols used by a company to distinguish its service or product.
  Example: Twitter logo, McDonald's golden arches
- Although patents protect one product, trademarks may cover a group of products. The Lanham Act, also called the Trademark Act of 1946, governs trademarks, infringement, and service marks.

COPY RIGHT

- Copyright law protects the rights of the original creator of original works of intellectual property.
- Unlike patents, copyrights is also important.
  Example: You can't copyright an idea. But you can write down an original speech, poem, or song and get a copyright.
- Once someone creates an original work of authorship (OWA), the author automatically owns the copyright. But, registering with the U.S. Copyright Office gives owners a head-start in the legal system.

TRADE SECRET

- Trade secrets are a company's intellectual property that isn't public, has economic value, and carries information, it may be a formula, recipe, or process used to gain a competitive advantage.
- To qualify as a trade secret, companies must work to protect proprietary information actively.
- Once the information is public knowledge, then it's no longer protected under trade secrets laws.
  Example: The recipe for Coca-Cola.
- Although a patent is public, trade secrets remain unavailable to anyone but the owner.

****

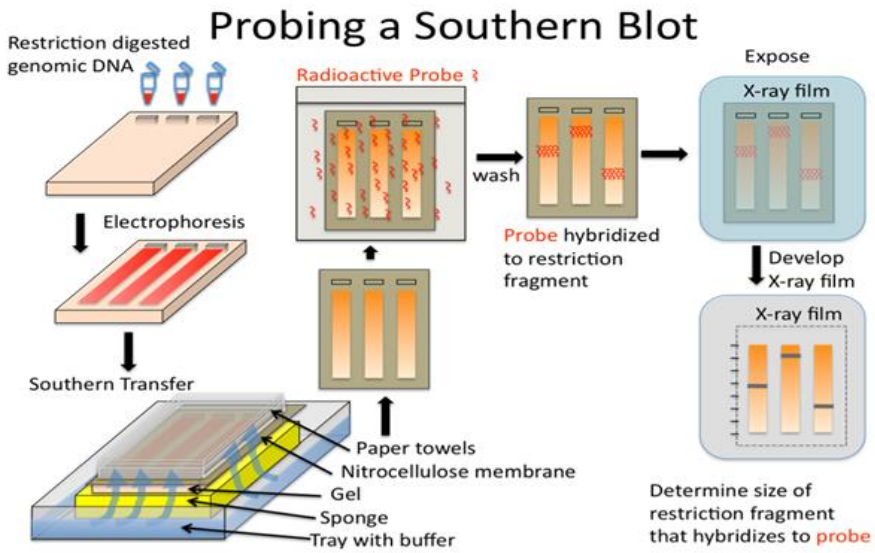
3. Explain about blotting techniques and state their applications

Answer: **BLOTTING TECHNIQUE**

- In molecular biology and genetics, blotting is an analysis technique used for the detection of specific biomolecules (proteins, DNA or RNA) in samples of complex composition by transferring them onto a carrier such as a nitrocellulose, polyvinylidene fluoride or nylon membrane.
- The blotting method was first discovered by Edwin M. Southern in 1975 for DNA.

PRINCIPLE:

- In his method, the DNA restriction fragments which are electrophoretically separated in an agarose gel are transferred to a solid support (nitrocellulose) and recognized as discrete bands following hybridization to a complementary DNA probe.
- After transferring, the proteins, DNA or RNA molecules are visualized by using colorant staining such as radio labelled molecules or specific labeling of some proteins or nucleic acids or a chemiluminescent reaction which is recorded by photographic film.



TYPES OF BLOTTING

There are different types of blotting techniques, among them, three are important such as, Southern blotting, western blotting, and Northern blotting.

1. SOUTHERN BLOTTING

This method was named after **Edward M. Southern** and used for the analysis of DNA sequences only.

STEPS

- At First, the Large DNA is fragmented into small pieces by using Restriction endonucleases.
- This fragmented DNA are separated according to their size by using gel electrophoresis.
- After electrophoresis, the separated fragments are transferred onto a nitrocellulose sheet/membrane.

- Then the membrane is exposed to ultraviolet radiation as result the fragments is visualized on X-ray film with the help of autoradiography.

APPLICATIONS OF SOUTHERN BLOTTING

1. **Gene mapping:** Southern blotting can be used to map the location of specific genes within a genome. By comparing the hybridization patterns of DNA from different individuals or species, researchers can identify similarities and differences in gene organization and function.
2. **DNA fingerprinting:** Southern blotting can be used to generate DNA fingerprints. This technique has important applications in forensic science and paternity testing.
3. **Detecting gene mutations:** Southern blotting can be used to detect mutations in specific genes that may be associated with disease or genetic disorders. By comparing the hybridization patterns. Researchers can identify the presence of mutations and their potential role in disease.

2. WESTERN BLOTTING

This technique was named after **W. Neal Burnette** and used for the detection and analysis of protein in a given sample.

STEPS

- At first, the desired proteins are isolated from the particular sample.
- After the loading of SDS protein complex in the well, the current is passed across the gel. The proteins which are tightly bounded to the SDS will move towards the positive pole because they are negatively charged. The migration of protein is inversely proportional to its size.
- Next, the gel is set upon a nylon membrane, and current is given across the gel so that all the proteins are transported onto the nylon membrane.
- The primary antibodies are added they will recognize a specific amino acid sequence.
- Now enzyme-labeled secondary antibodies are added which will recognize the primary antibody.
- The chemiluminescent substrates are used for identification. The light is being emitted once the substrate has been added and can be detected with film imager.

APPLICATIONS OF WESTERN BLOTTING

Western blotting has several applications in molecular biology, including:

1. **Protein detection:** It is used to detect specific proteins in a sample. By using antibodies that are specific to the target protein.
2. **Protein quantification:** Western blotting can be used to quantify the amount of a specific protein in a sample. By comparing the signal intensity of the target protein to a known standard, researchers can determine the amount of protein present in the sample.
3. **Protein expression analysis:** It can be used to study the expression of specific proteins in different tissues or under different conditions. By identifying differences in protein expression and gain insights into the regulation of protein expression.
4. **Post-translational modification analysis:** It can be used to study post-translational modifications (PTMs) of proteins, such as phosphorylation, glycosylation, or acetylation.
5. Western blotting is a valuable tool in molecular biology that has a wide range of applications in protein detection, quantification, expression analysis, and PTM analysis.

3. NOTHERN BLOTTING

This technique was given by **Alwine** and used for the detection and analysis of RNA in a given sample.

STEPS

- First of all extract and purify mRNA from the cells.
- After that separate those purified mRNA on agarose gel containing the formaldehyde.
- The RNA fragments are transferred onto the carrier membrane such as aminobenzyloxymethyl filter paper.
- The UV or heat is applied to fix the RNA onto the membrane.
- Add DNA labeled probe for hybridization.
- Then, the end mRNA-DNA hybrid are then detected by X-ray film.

APPLICATIONS OF NOTHERN BLOTTING

- **Gene expression analysis:** It can be used to study the expression of specific genes in different tissues or under different conditions. Can identify differences in gene expression and gain insights into the regulation of gene expression.
- **Identifying alternatively spliced transcripts:** It can be used to identify and distinguish between different splice variants of a particular gene. By designing probes that are specific to different splice variants, researchers can determine which variants are present in a particular sample.
- **Detecting gene fusions:** It can be used to detect gene fusions, which occur when two separate genes are fused together. Can identify the presence of the fusion gene and study its potential role in disease.
- **Studying small RNAs:** It can be used to study the expression and processing of small RNAs, such as microRNAs and siRNAs.

Short Answers

1. What is northern blotting ?

Ans: This technique was given by **Alwine** and used for the detection and analysis of RNA in a given sample.

Applications

- **Detecting gene fusions:** It can be used to detect gene fusions, which occur when two separate genes are fused together. Can identify the presence of the fusion gene and study its potential role in disease.
- **Studying small RNAs:** It can be used to study the expression and processing of small RNAs, such as microRNAs and siRNAs.

2. What is IPR

Ans**: Intellectual Property Rights or IPR** refers to the legal rights that creators or owners have over their Intellectual properties. Intellectual Property is refered to the creations of the human mind, exclusive to a single person. It includes inventions, literature and artistic masterpieces, designs, emblems, titles, Music and pictures e.t.c., utilized in commerce and business. Such creations are prone to plagiarism or copying. So, Intellectual Property Rights protect them from unauthorized production, distribution, and display.

3. Write about Patent and Trade Mark

Ans: **PATENT:** A Patent is a form of right granted by the government to an inventor for a period of time, which provides the owner to exclude others from making, using, selling, etc. of his own product.

TRADE MARK: A trademark is a sign, symbol, design or expression of the owner that distinguishes a particular product from other similar products.

4. Write about southern blotting technique

Ans: This method was named after **Edward M. Southern** and used for the analysis of DNA sequences only.

STEPS

- At First, the Large DNA is fragmented into small pieces by using Restriction endonucleases.
- This fragmented DNA are separated according to their size by using gel electrophoresis.
- After electrophoresis, the separated fragments are transferred onto a nitrocellulose sheet/membrane.
- Then the membrane is exposed to ultraviolet radiation as result the fragments is visualized on X-ray film with the help of autoradiography.

**********END OF UNIT -3 **********

UNIT – IV Bioinformatics

ESSAYS (8 Marks)

1. Explain about BLASTA and FASTA. Mention their applications in Bioinformatics

ANSWER : **BLAST (Basic Local Alignment Search Tool)** : It is a widely used bioinformatics program that was first introduced by **Stephen Altschul et al.** in **1990**. One of the most commonly used bioinformatics tools today to study DNA and protein sequences,

- BLAST is a powerful tool for analyzing biological sequence data.
- Since the initial release of BLAST in 1990, it has undergone continuous updates to improve its speed and accuracy.

TYPES OF BLASTA

There are five types (variants) of BLAST that are differentiated based on the type of sequence (DNA or protein) of the query and database sequences.

1. **BLASTN** compares a nucleotide query sequence to a nucleotide sequence database.
2. **BLASTP** compares a protein query sequence to a protein sequence database.
3. **BLASTX** compares a nucleotide query sequence to a protein sequence database
4. **TBLASTN** compares a protein query sequence to a nucleotide sequence database.
5. **TBLASTX** compares a nucleotide query sequence to a nucleotide sequence database.

APPLICATIONS OF BLASTA

- BLAST can be used to identify unknown sequences by comparing them with known sequences in a database which helps in predicting the functions of proteins or genes.
- BLAST can also be used in phylogenetic analysis which is important for understanding the evolutionary relationships between different species.
- BLAST can also be used to identify functionally conserved domains within proteins which is important for predicting the functions of proteins.

FASTA : FASTA is one of the first widely-used database similarity search tool.

- FASTA (or FastA), an abbreviation for 'Fast-All', is a sequence alignment tool that takes nucleotide or protein sequences as input and compares it with existing databases.
- It was first developed by **David J. Lipman** and **William R. Pearson** in **1985** and has since been refined and adapted for various applications.
- The text-based file format for representing nucleotide or protein sequences, which originates from the FASTA program, has now become a standard in bioinformatics.
- FASTA was originally developed for comparing protein sequences.
- The program has been continually updated and improved.

TYPES OF FASTA

FASTA compares a DNA query sequence against a database of DNA sequences or a protein query sequence against a database of protein sequences using the FASTA algorithm.

1. **SSEARCH** performs protein-protein or DNA-DNA comparisons
2. **GGSEARCH/ GLSEARCH** works using a global alignment algorithm (GGSEARCH) or a combination of global and local alignment algorithms (GLSEARCH) to compare protein and nucleotide sequences.
3. **FASTX/ FASTY** compares a DNA sequence and a database of protein sequences by translating the DNA sequence into three frames and allowing gaps and frameshifts.
4. **TFASTX/ TFASTY** compares a protein sequence and a database of DNA sequences. The DNA sequence is translated in six frames – three in the forward direction and three in the reverse direction.
5. **FASTF/ TFASTF** compares mixed peptide sequences against a protein (FASTF) or translated DNA (TFASTF) databases.
6. **FASTS/ TFASTS** compares a set of short peptide fragments against the protein (FASTS) or translated DNA (TFASTS) databases.

APPLICATIONS OF FASTA

- FASTA can be used in the sequence alignment to identify regions of similarity. This is useful for identifying conserved regions in DNA or protein sequences, which can help to identify functional domains or motifs. Identifying these functional domains or motifs can provide insights into the biological function of the sequence.
- FASTA can be used to search large databases of sequences to find matches to a given query sequence. This helps to identify homologous sequences, which can help to predict the function of a newly identified sequence.
- FASTA can construct phylogenetic trees by aligning sequences from different species and identifying evolutionary relationships between them.

---

2. Explain about Sequence alignment in Bioinformatics.

ANSWER: **SEQUENCING ALIGNMENT IN BIOINFORMATICS**

Sequence comparison is a crucial aspect of bioinformatics analysis that involves comparing newly determined biological sequences with previously known sequences stored in databases.

- It arranges two or more nucleotide or amino acid sequences to identify regions of similarity between the sequences.
- These regions of similarity are helpful in understanding the functional, structural, and evolutionary relationships between the sequences.

Two commonly used sequence alignment algorithms are

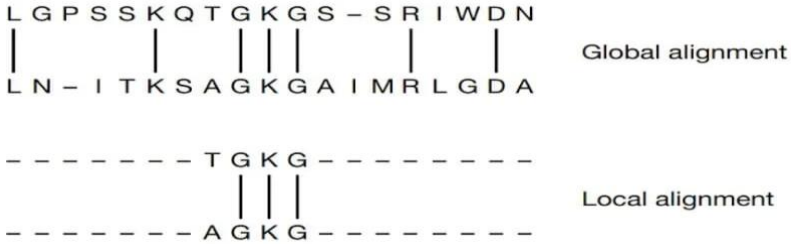
1. Global alignment
2. Local alignment.

Global alignment: Global alignment is a method of comparing two sequences, which aligns the entire length of the sequences by maximizing the overall similarity.

- This method is used when comparing sequences that are of the same length.

Local alignment: In local alignment, instead of attempting to align the entire length of the sequences, only the regions with the highest density of matches are aligned.

- This is useful for identifying short conserved regions in protein or nucleotide sequences.



TYPES OF SEQUENCING ALIGNMENT

1. PAIRWISE SEQUENCE ALIGNMENT

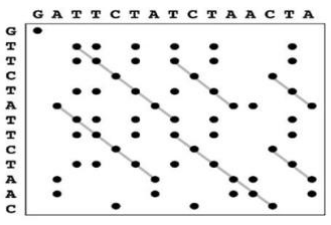
- Pairwise sequence alignment is the type of sequence alignment that involves aligning two sequences to identify the optimal pairing of the sequences.
- It is based on a scoring system that assigns positive scores to matching characters and negative scores to mismatching characters or gaps.
- The main objective of pairwise sequence alignment is to obtain the highest possible score, which indicates the degree of similarity between the two sequences.

Methods of pairwise sequence alignment

There are three main methods for generating pairwise alignments:

a. Dot-matrix method

- Dot matrix method, also known as the dot plot method, is a graphical method of sequence alignment that involves comparing two sequences by plotting them in a two-dimensional matrix



b. Dynamic programming

- Dynamic programming is used to find the optimal alignment between two proteins or nucleic acid sequences by comparing all possible pairs of characters in the sequences.

c. Word or k-tuple method

- Word or k-tuple methods are heuristic methods best known for their use in the database search tools FASTA and BLAST.
- The word method is a fast method for aligning two sequences. It begins by identifying short identical sequences, also known as words or k-tuples, and then uses dynamic programming to align the sequences based on these words.

2. MULTIPLE SEQUENCE ALIGNMENT

- Multiple Sequence Alignment involves aligning multiple (three or more) biological sequences to achieve optimal sequence matching.
- Multiple sequence alignments are used to identify conserved sequence regions and to construct phylogenetic trees, which help us understand the functional and evolutionary relationships between different species or groups of organisms.
- Multiple sequence alignment can be performed using either exhaustive or heuristic approaches.

APPLICATIONS OF SEQUENCE ALLIGNMENT

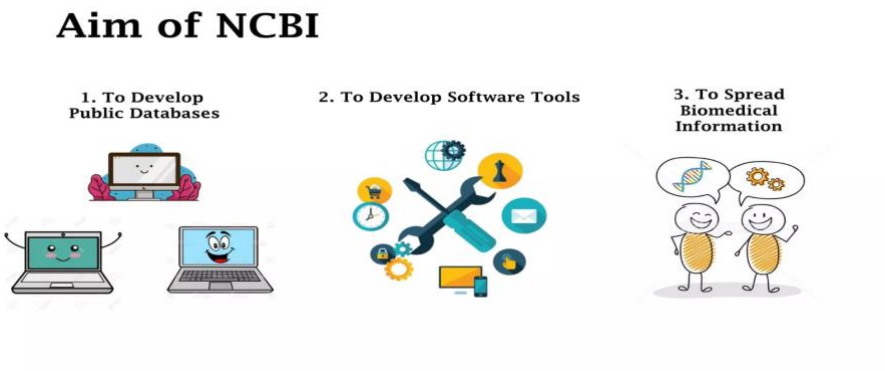
- Sequence alignment can identify unknown sequences by comparing them with already known sequences in databases.
- Sequence alignment is also used to identify conserved sequence patterns and motifs, which helps to characterize the functions of the sequences.
- Sequence alignment can also produce phylogenetic trees and obtain information about the evolutionary relationship between the sequences aligned.
- Sequence alignment can also predict proteins' secondary and tertiary structures. It can also predict gene locations and new members of gene families.
- Sequence alignment can also be used to develop degenerate PCR primers by analyzing multiple related sequences.

---

3. Validate the importance of NCBI and PubMed as Bioinformatic resources

Answer: NCBI – National Center for Biotechnology Information

- The National Center for Biotechnology Information (NCBI) is part of the National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH).
- It is approved and funded by the government of the United States.
- The NCBI is located in **Bethesda, Maryland**, **USA** and was founded in **1988** through legislation sponsored by **US Congressman Claude Pepper**. On November 4, 1998 the President Ronald Signed the Health Omnibus Extension Act to create The National Center for Information as part of National Library of Medicine at NIH.
- The NCBI facilitates the use of databases and software and performs research on advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules including proteins.
- The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services.
- NCBI was directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in bioinformatics.



IMPORTANCE OF NCBI

1. It creates automated systems for knowledge about molecular biology, biochemistry, and genetics.
2. It helps in performing research into advanced methods of analyzing and interpreting molecular biology data.
3. Enable biotechnology researchers and medical care personnel to use the systems and methods developed.
4. It creates public databases, conduct research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information.

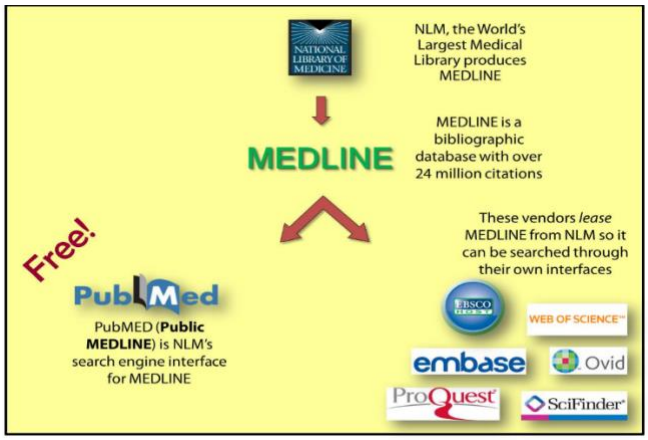
PubMED

- PubMed is a free, government website that allows users to search for biomedical and health related scholarly literature.
- PubMed = search engine for MEDLINE.
- It is produced and maintained by NCBI(National center for Biotechnology Information).
- Anyone with an internet connection can search in PubMed. In order to access full text of articles, you will need to work through the OSU Library subscriptions, or another library's access.
- The majority of the records in PubMed come from the MEDLINE database.
- In addition to MEDLINE records, PubMed also contains a pocket of content that is non-indexed such as
    a. **In-process records** - These are items that will eventually be indexed for MEDLINE, but there has not been enough time for them to go through the full indexing process. PubMed will put up basic or "skeleton" records for these items until they can have the full indexing applied.
    b. **Out-of-scope records** - These are items that are published within a MEDLINE-indexed journal, but the article itself is considered outside of the scope to be added to MEDLINE.
    Example: A general science journal may only have its health science articles indexed, but not those articles talking about non-health science topics like geology or physics.
    c. **PMC records** - These are items that appear in PMC (PubMed Central; see below for more information), but are not published in a MEDLINE-indexed journal. PubMed

will also put up a basic record for these items, but they will never be indexed for MEDLINE.

Importance of PubMed

- It is an interface used to search Medline and additional biomedical and health content.
- It provides free access worldwide via the web.
- PubMed comprises over 22 million citations and abstracts for biomedical literature indexed in NLMs MEDLINE database & from other life science journals and online books.
- PubMed citations and abstracts include the fields of biomedicine and health, & cover portions of the life sciences, behavioural sciences, chemical sciences & bioengineering.
- PubMed uses NCBI's Entrez search and retrieval system.
- The abstract display of PubMed citations may provide links to the full text from other sources, such as directly from a publishers website or PubMed Central(PMC).



Short Answers

1. What is SWISS-PROT
   Ans: SWISS-PROT is a database of protein sequences that includes annotations and links to other databases.

2. What is GenBank ?
   Ans: GenBank is s public data-base of DNA sequences that contain information about more than 300,000 orgnaisms.

3. Define BLAST
   Ans: A tool in bioinformatics that is used for determining the similar biological sequences is known as BLAST. It is a widely used bioinformatics program that was first introduced by **Stephen Altschul et al.** in **1990**.

4. What is PubMed?
   Ans: It is a free database of biomedical and life sciences literature that supports bioinformatics research. PubMed = search engine for MEDLINE. It is produced and maintained by NCBI(National center for Biotechnology Information).

5. What is ExPASy?

Ans: Expasy is an online bioinformatics resource operated by the SIB Swiss Institute of Bioinformatics. It is an extensible and integrative portal which provides access to over 160 databases and software tools and supports a range of life science and clinical research areas, from genomics, proteomics and structural biology, to evolution and phylogeny, systems biology and medical chemistry.

**********UNIT - 4 END**********

UNIT -V Biostatistics

ESSAYS (8 Marks)

1. Define mean, median and mode with suitable examples

Answer:

MEAN

- Mean is the sum of all the values in the data set divided by the number of values in the data set.
- It is also called the Arithmetic Average. Mean is denoted as $\bar{x}$ and is read as x bar.
- The formula to calculate the mean is:

Mean ($\bar{x}$) = Sum of Values / Number of Values

- The symbol used to represent the mean, or arithmetic average, of a dataset is typically the Greek letter **"μ"** (mu) when referring to the population mean, and "$\bar{x}$" (x-bar) when referring to the sample mean.

Example: Find the mean of data sets 10, 30, 40, 20, and 50.

Solution:

Mean of the data 10, 30, 40, 20, 50 is

Mean = (sum of all values) / (number of values)

Mean = (10 + 30 + 40 + 20+ 50) / 5 = 30

MEDIAN

- A Median is a middle value for sorted data.
- The sorting of the data can be done either in ascending order or descending order.
- A median divides the data into two equal halves.
- The formula to calculate the median of the number of terms if the number of terms is even is:

$$\text{Median (n = even number)},\quad \text{Median} = \frac{\left[\left(\frac{n}{2}\right)^{th}\text{term} + \left\{\left(\frac{n}{2}\right)+1\right\}^{th}\text{term}\right]}{2}$$

- The formula to calculate the median of the number of terms if the number of terms is odd is:

$$\text{Median (n = odd number)},\quad \text{Median} = \left[\frac{(n+1)}{2}\right]^{th}\text{term}$$

- The letter "**M**" is commonly used to represent the median of a dataset, whether it's for a population or a sample.

- **To find the median**

Median is the middle term of the data when it is arranged in either ascending or descending order. It is calculated using the formula:

- Median = [(n + 1)/2]th term{**when n is odd**}
- Median = Median = [(n/2)th term + {(n/2) + 1}th term] / 2 {**when n is even**}

Example: Find the median of given data set 30, 40, 10, 20, and 50.

Solution:

Median of the data 30, 40, 10, 20, 50 is,

Step 1: Order the given data in ascending order as:
10, 20, 30, 40, 50

Step 2: Check n (number of terms of data set) is even or odd and find the median of the data with respective 'n' value.

Step 3: Here, n = 5 (**odd**)

Median = [(n + 1)/2]th term

Median = [(5 + 1)/2]th term

= 30

MODE

- The mode or modal value of a data set is the most frequently occurring value.
- It's a measure of central tendency that tells you the most popular choice or most common characteristic of your sample.

Type	Definition	Example Data Set	Modes
Unimodal	When there is only one and only one mode in a dataset.	Set X = {1, 2, 2, 3, 6, 7, 7, 7, 8, 9}	Only 7
Bimodal	When there are two modes in the given data set.	Set A = {1, 1, 1, 3, 4, 4, 6, 6, 6}	1 and 6
Trimodal	When there are three modes in the given data set.	Set A = {2, 2, 2, 3, 4, 4, 6, 6, 6, 7, 9, 9, 9}	2, 6, and 9

Type	Definition	Example Data Set	Modes
Multimodal	When there are four or more modes in the given data set.	Set A = {1, 1, 1, 3, 4, 4, 6, 6, 6, 7, 9, 9, 9, 11, 11, 11}	1, 6, 9, and 11

- The number that comes the most frequently is the mode. Mode can be calculated for both numerical and categorical data. It is symbolised as Z or M_0.
- Depending upon the number of modal solutions, mode is classified into the following categories:
    - Unimodal
    - Bimodal
    - Trimodal
    - Multimodal

Example: In the given set of data: 2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7, the mode of the data set is 4 since it has appeared in the set four times than other numbers.

- Relation between Mean, Median, And Mode
- For any group of data, the relation between the three central tendencies mean, median, and mode is shown in the image below:

Mode = 3 Median – 2 Mean

---

2. Explain about Range, Mean deviation and standard deviation with suitable examples

ANSWER:

RANGE : In a given data set the difference between the largest value and the smallest value of the data set is called the range of data set.

Example: If height(in cm) of 10 students in a class are given in ascending order, 160, 161, 167, 169, 170, 172, 174, 175, 177, and 181 respectively. Then range of data set is (181 – 160) = 21 cm.

Range Formula

The formula to find the Range is:

Range = Highest value – Lowest Value

Example: Find the range of the given data set 12, 19, 6, 2, 15, 4.

Solution:

Given set is {12, 19, 6, 2, 15, 4}
Here,
Lowest Value = 2
Highest Value = 19
Range = 19 − 2 = 17

MEAN DEVIATION

- The mean deviation is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set.

Mean Deviation Formula

The formula to calculate the mean deviation for the given data set is given below.

$$\textbf{Mean Deviation} = [\Sigma |X - \mu|]/N$$

Here,

Σ represents the addition of values

X represents each value in the data set

μ represents the mean of the data set

N represents the number of data values

| | represents the absolute value, which ignores the "-" symbol

EXAMPLE : Determine the mean deviation for the data values 5, 3,7, 8, 4, 9.

Solution:

FIRST STEP Given data values are 5, 3, 7, 8, 4, 9.

We know that the procedure to calculate the mean deviation.

First, find the mean for the given data:

Mean, $\mu = (5+3+7+8+4+9)/6$

$\mu = 36/6$

$\mu = 6$

Therefore, the mean value is 6.

SECOND STEP Now, subtract each mean from the data value, and ignore the minus symbol if any

(Ignore"-")

$5 - 6 = 1$

$3 - 6 = 3$

$7 - 6 = 1$

$8 - 6 = 2$

$4 - 6 = 2$

$9 - 6 = 3$

Now, the obtained data set is 1, 3, 1, 2, 2, 3.

THIRD STEP Finally, find the mean value for the obtained data set

Therefore, the mean deviation is

$= (1+3 + 1+ 2+ 2+3) /6$

$= 12/6$

$= 2$

Hence, the mean deviation for 5, 3,7, 8, 4, 9 is 2.

STANDARD DEVIATION

- Standard deviation is a fundamental concept in statistics that measures the dispersion of data points which defines the extent to which data points in a dataset deviate from the mean, providing a clear sense of the variability or spread within the data.
- Standard Deviation is defined as the degree of dispersion of the data point from the mean value of the data point. It tells us how the value of the data points varies from the mean value of the data point and it tells us about the variation of the data point in the sample of the data.

Standard Deviation of a given sample of data set is also defined as the square root of the variance of the data set. Mean Deviation of the n values (say x1, x2, x3, …, xn) is calculated by taking the sum of the squares of the difference of each value from the mean, i.e.

$$\textbf{Mean Deviation} = 1/n\sum in\ (xi - \bar{x})2$$

Standard Deviation Formula

Standard Deviation Formula for Sample Data

$$s=\sum i=1n(xi-x^{-})2n-1\ s=n-1\sum i=1n(xi-x^{-})2$$

where,

s is Population Standard Deviation

xi is ith observation

$\bar{x}$ is Sample Mean

N is Number of Observations

Standard Deviation Formula of Population Data

$$\sigma=\sum i=1N(xi-\mu)2N\ \sigma=N\sum i=1N(xi-\mu)2$$

where,

σ is Population Standard Deviation

xi is ith Observation

μ is Population Mean

N is Number of Observations

- There are six main steps for finding the standard deviation by hand. We'll use a small data set of 6 scores to walk through the steps.

Step 1: Using the example of Mean deviation

The data set is 1, 3, 1, 2, 2, 3. Square each deviation from the mean

$(1)^2=1X1=1$

$(3)^2=3X3=9$

$(1)^2=1X1=1$

$(2)^2=2X2=4$

$(2)^2=2X2=4$

$(3)^2=3X3=9$

Step 2: Find the sum of squares

- Add up all of the squared deviations. This is called the sum of squares. 1+9+1+4+4+9=28

Step 3: Find the variance

- Divide the sum of the squares by n – 1 (for a sample) or N (for a population) – this is the variance

Since we're working with a sample size of 6, we will use n – 1, where n = 6.

Variance $\frac{28}{6-1}=\frac{28}{5}=5.6$

Step 5: Find the square root of the variance

- To find the standard deviation, we take the square root of the variance.

Standard deviation $\sqrt{5.6} = 2.37$

- From learning that SD =2.37, we can say that each score deviates from the mean by 2.37 points on average.

Short Answers

1. Define mean

   Ans: Mean is the sum of all the values in the data set divided by the number of values in the data set.

   It is also called the Arithmetic Average. Mean is denoted as $\bar{x}$ and is read as x bar.

   The formula to calculate the mean is:

   Mean ($\bar{x}$) = Sum of Values / Number of Values

2. Define Standard Deviation

   Ans**: Standard deviation** is a fundamental concept in statistics that measures the dispersion of data points which defines the extent to which data points in a dataset deviate from the mean, providing a clear sense of the variability or spread within the data. In statistics, the standard deviation is a measure of the amount of variation of the values of a variable about its mean.

3. What is Biostatic software?

   Ans: Biostatic software is used to analyze data and perform statistical analysis of biological and medical data.

4. Define Range and Sample mean

   Ans: **RANGE :** In a given data set the difference between the largest value and the smallest value of the data set is called the range of data set.

   SAMPLE MEAN : A sample is just a small part of a whole. The mean is another word for "average." So the sample mean would be the average amount. The sample mean is useful because it allows you to estimate what the whole population is doing, without surveying everyone.

**************************END OF UNIT -5************************

ALL THE BEST