PITHAPUR RAJAH'S GOVERNMENT COLLEGE

Autonomous and NAAC Re-accredited with 'A' Grade (3.17/4.00 CGPA)

Affiliated to Adikavi Nannaya University

Opp. Mc Laurin School, Raja Ram Mohan Roy Road, Kakinada 533001, Andhra Pradesh, India

E-mail: kakinada.jkc@gmail.com, Tel: 0884-2379480

N. Jessica M.sc APSET

LECTURER IN MICROBIOLOGY

DEPARTMENT OF MICROBIOLOGY

P.R.G.C (A),KAKINADA

rDNA technology, bioinformatics and biostatistics

Unit - I : Recombinant DNA Technology

Basic Principle of genetic engineering Steps in gene cloning

Gene cloning:

* The production of exact copies of a particular gene or DNA sequence using genetic engineering techniques is called gene cloning
* The term "gene cloning", "molecular cloning" and "recombiant DNA technology"
* When DNA is extracted from an organism, all its genes are obtained. In gene (DNA) cloning a particular gene is copied forming "clones".

→ DNA cloning can be achieved by two different methods:

(1) Cell based DNA cloning

(2) Cell free DNA cloning (PCR)

Requirements for Gene cloning (cell based):

* DNA fragment containing the desired genes to be cloned.
* Restriction enzymes and ligases enzymes
* vectors - to carry, maintain and replicate cloned gene in host cell.
* Host cell - in which recombinant DNA can replicate.

Steps in Gene cloning:

The basic 8 steps involved in gene cloning they are:

(1) Isolate the gene (or) DNA fragment:

* The target DNA or gene to be cloned must be first isolated a gene of interest is a fragment of gene whose product a protein or enzyme or hormone.
* The desired gene may be isolated by using restriction endonuclease (RE) enzyme, which cut DNA at specific recognition nucleotide sequences known as restriction sites towards the inner region producing blunt or sticky ends.

(2) Selection of suitable cloning vector

* The vector is carrier molecule which can carry the gene of interest into the host, replicate there along with it making its multiple copies
* cloning vectors are limited to the size of insert they can carry. depending on the size.
* the different types of vector for cloning are plasmids, bacteriophages, BACs, YACs and MACs.

(3) Essential characteristics of cloning vectors.

* All cloning vectors are carrier DNA molecules it must be self-replicating inside host cell.
* It must possess a unique restriction site for RE enzymes. introduction of donar DNA fragment must not interfere with replication.

(4) Formation of Recombinant DNA

* The plasmid vector is cut open by the same RE enzymes used for isolation of donar DNA fragment.
* In the presence of DNA ligase, base pairing of donar DNA fragment and plasmid of donar

DNA fragment and plasmid vector occurs.

(5) Transformation of recombinant vector into suitable host

* the recombinant vector is transformed into suitable host cell mostly a bacterial cell.

(6) Isolation of Recombinant cells

* the transformation process generates a mixed population of transformed and non-transformed host cells.
* For isolation of recombinant cell from non-recombinant cell, marker gene of plasmid vector is employed.

(7) Multiplication of selected host cells

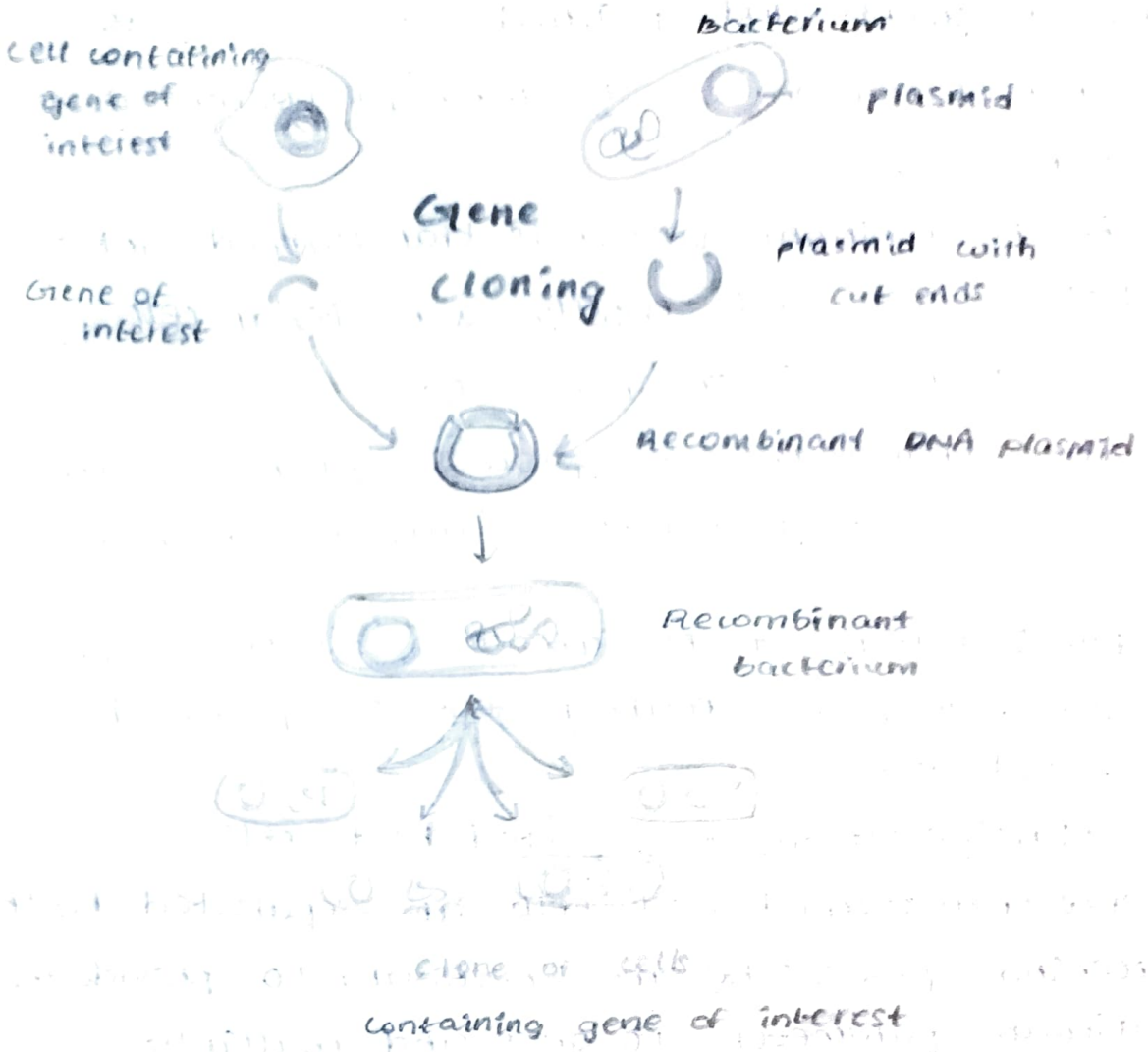
* Once transformed host cells are separated by the screening process: becomes necessary to provide them optimum parameters to grow and multiply.
* the transformed host cells are introduced into fresh culture, obtaining numerous copies of gene simply replication of the host cell.

(8) Isolation and purification of the product

* the next step involves isolation of multiplied gene attached with the vector or protein.
* this is followed by purification of the isolated gene copy / protein.

Applications:

* A particular gene can be isolated and its nucleotide sequence
* control sequences of DNA can be identified / analyzed
* Protein / enzyme / RNA function can be investigated



Basic Principles of genetic engineering:

* Genetic engineering, also known as genetic modification, is the process of altering an organism's DNA to introduce desirable traits.
* The main principle of organism's genetic material to produce improved or new organisms.

(1) Recombinant DNA technology: which involves transferring a desired gene from one organism to another

(2) cloning: the process of using a cell's ability to divide and differentiate into a complete organism.

(3) Isolation of DNA fragments: the process of

extracting DNA from an organism to be used in genetic engineering.

(4) cloning vectors: the vector used to transfer genes, such as plasmid.

(5) Insertion of recombinant DNA: the process of introduction the desired gene into the host.

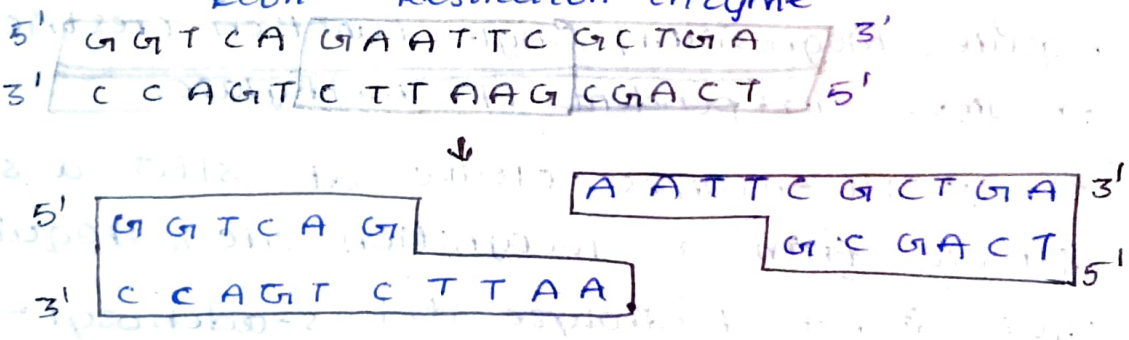
(6) Maintenance of introduced DNA: the process of ensuring that the introduced DNA is passed on the next generation.

Restriction Endonucleases

* A restriction endonuclease is a bacterial enzyme that cuts DNA at specific sites. they are also known as restriction enzymes, RE ases, E Nases, or restrictases.

* RE cut the DNA double helix in very precise, it cleaves DNA into fragments at near specific sites is known as restriction sites.

* they have the capacity recognize specific base sequences on DNA and then to cut each strand at a given place. Hence they are called as "molecular scissors".



Sources of Restriction Enzymes

* The natural source of restriction endonucleases are bacterial cells.

Recognition sites:

* The DNA sequences are recognized by restriction enzymes are called palindromes.
* Palindromes are the same on the two strands but in opposite directions.

If the sequence on one strand is GAATTC read in 5' → 3' direction.

the sequence on the opposite strand is CTTAAG read in 3' → 5' direction.

Types of Restriction Enzymes:

Traditionally, four types of restriction enzymes are recognized, designated I, II, III and IV, which differ primarily in structure, cleavage sites, specificity and cofactors.

(1) Type I enzymes: cleave at sites remote from a recognition site, require both ATP and s-adenosyl-L-methionine to function; multifunctional protein with both restriction and methylase activities.

(2) Type II enzymes: cleave within or at short specific distances from a recognition site; most require magnesium; single function (restriction) enzymes independent of methylase.

(3) Type III enzymes: cleave at sites a short distance from a recognition site; require ATP (but do not hydrolyze it); s-adenosyl-L-methionine stimulates the reaction but is not required; it exists as part of a complex with a modification methylase.

(4) Type IV enzymes: target modified DNA, eg: methylated, hydroxymethylated and glucosyl-hydroxymethylated DNA.

Applications of type II restriction endonucleases

Restriction enzymes are used in molecular biology for DNA analysis, cloning and fragmentation. they are also used to study DNA protein interactions and phosphodiester hydrolysis.

DNA analysis:

It is used to analyze DNA for genetic or hereditary diseases or to detect mutated genes.

Gene cloning

It is used to create recombinant DNA molecules by cutting DNA at specific position.

DNA fragmentation

It is used to create reproducible endonucleases are used to create reproducible DNA fragments that can be separated by length using gel electrophoresis.

Protein - nucleic acid interactions

It is used to aremodel system to study how proteins and nucleic acids interact.

DNA polymerase and ligase

DNA polymerase and DNA ligase are enzymes that work together to replicate and repair DNA.

DNA polymerase

* Synthesis new DNA strands by adding nucleotides to a growing chain.

Vectors

(a) Cosmid :

* Cosmid vectors are hybrid plasmids that contain the cos sequence from the bacteriophage lambda (λ).
* They are used to clone large DNA fragments such as genomic DNA, in genetic engineering.
* Cosmids are constructed by inserting the cos region of the λ phage into a plasmid.
* The cosmid vector replicates like a plasmid within the host cell.
* The cosmid vector can also be packaged into phage particles. 5.4 kB
* The packaged cosmid DNA can be transferred into bacterial cells.
* The transformed cells can be identified by plating on a medium containing an antibiotic.

Uses :

* Cosmids are used to clone large DNA fragments that cannot fit into standard plasmid vectors.
* Cosmid are used to build genomic libraries
* Cosmids are used in homologus recombination between two different plasmids in the same cell.

Examples :

* c2RB is a classic example of a cosmid vector.
* pJB8, pHC79 and pLER5 are examples of commonly used cosmids.

production ; uses the F plasmid, a type of plasmid found in bacteria.

* stability : more stable than YACs making them better for functional analysis in mammalian cells.

YAC : Yeast artificial chromosome :

* YAC is a vector that clones DNA fragments in yeast cells.
* structure : linear, resembling natural chromosomes.
* cloning capacity : can clone DNA fragments of upto 1 megabase (Mb)
* Production : produced in a region specific manner.
* stability : less stable than BACs.

Transformation of DNA by chemical method :

Cell membrane is a sheet like assembly of amphipathic molecules (having both hydrophilic and hydrophobic parts) that separate cells from their environment.

the commonly used methods of chemical transformation :

(1) calcium chloride mediated transformation
(2) Polyethylene glycol (PEG) mediated transformation.

(1) Calcium chloride induced transformation

calcium chloride induced transformation involves making the E.coli cells competent to take up the plasmid DNA.

Unit-II: Applications of r-DNA Technology

Genomic and C-DNA Libraries

Genomic DNA library

(1) Genomic library is a collection of the clones bearing the total genomic DNA of an organism.

(2) Genomic library consists of the entire genomic DNA including noncoding (introns and regulatory) DNA.

(3) Genomic library is large

(4) The starting material is DNA.

(5) Reverse transcription does not happen

(6) It include all possible fragments of DNA from a given cell or organism.

(7) Vector used genomic library include plasmid, cosmid, lambda phage, BAC and YAC in order to accommodate large fragments.

(8) Restriction endonucleases and ligases are important for its construction.

(9) They carry introns also.

(10) They may represent the DNA of both eukaryotic and prokaryotic organisms

CDNA Library

(1) CDNA library is a collection of the clones bearing the complementary DNA to the mRNA of an organism

(2) CDNA library contains only the coding sequences; it does not contain introns.

(3) cDNA library is small.

(4) Starting material is mRNA

(5) Reverse transcription happens in the first cDNA strand synthesis.

(6) cDNA library carries only expressed gene sequences.

(7) Vectors used cDNA library include plasmid, phagemids, lambda phage etc to accommodate small fragments as cDNA has no introns.

(8) Reverse transcriptase enzyme plays an important role in its construction.

(9) They lack introns.

(10) They represent the DNA of only eukaryotic organisms.

Polymerase chain reaction (PCR):

It is a technique used in molecular biology to amplify a single copy or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.

It is in vitro (Test tube) amplification to a specific sequence.

Principle:

PCR is to amplify DNA by heating it to separate the double stranded DNA into single strands, which are then copied using DNA polymerase.

Components of PCR:

DNA template: DNA is to be copied, this can come from a variety of sources, including viruses, bacteria.

DNA polymerase: An enzyme that synthesizes new DNA strands. Taq polymerase is a well-known DNA polymerase.

Primers: short stretches of single-stranded DNA that bind to the template

Oligonucleotide:

Deoxynucleotide triphosphates (dNTPs): the building blocks of new DNA strands. dNTPs are single units of bases that provide energy for polymeriza-tion.

Types of PCR:

Real-time PCR: uses a fluorescent reporter to detect DNA amplification in real time.

Reverse transcriptase PCR (RT-PCR): Measures RNA expression levels.

Quantitative PCR (qPCR): Uses DNA amplification linearity to quantify a known sequence in a sample.

Steps involved in PCR:

(1) Denaturation:

Denaturation occurs reaction mixture hea-ted 94°C for about 0.5 to 5 minutes this breaks the hydrogen bonds between the two strands of DNA converts into the single stranded DNA.

the single stranded template is the pro-duction of new strands of DNA.

(2) Annealing:

This reaction temperature is lowered 54-68°c for around 20-40 seconds here the primers bind to their complementary sequences on template DNA they serve as the starting point of the synthesis of DNA.

(3) Elongation:

At this step the temperature is raised to 72-80°c the bases are added to the 3' end of the primers by the taq polymerase enzyme.

Applications:

Medicine:

Testing of genetic disease mutations

Forensic science

Identifying the DNA samples in the crime scene.

detecting degarding microorganisms.

degarding microorganisms in toxic waste and pollutants.

~~RFLP~~ RFLP: Restriction Fragment Length polymorphism:

Restriction fragment length polymorphism (RFLP) is a technique used in microbiology to analyze DNA differences between organisms

* It is a technique that exploits variations in homologous DNA sequences.

Working of RFLP:

(1) Isolate DNA from the organism
(2) Digest the DNA with restriction enzymes
(3) separate the digested DNA by size using gel electrophoresis
(4) compare the patterns of DNA fragments to identify differences

Uses of RFLP:

* Genetic disease diagnosis: RFLP can help diagnose genetic diseases in families with known disorders.
* Paternity testing: RFLP can help determine paternity
* Forensics: can help identify criminals based on their DNA.
* Genome mapping: RFLP can help map genes.
* Microbiological diagnosis: can help identify and specify bacterial species that cause infections.

Random amplified polymorphic DNA (RAPD):

Random amplified polymorphic DNA (RAPD) is a DNA based technique used to identify genetic variation and type microorganisms, especially during clinical out breaks.

It is also known as arbitrarily primed -polymerase chain reaction (APPCR).

Working of RAPD:

* RAPD uses a single primer of arbitrary

nucleotide sequence to amplify random segments of genomic DNA.

* The amplified DNA segments are separated on an agarose gel to produce a strain-specific profile.
* The pattern of amplified DNA segments on the gel is unique to each organism.
* A single short primer with a random nucleotide sequence is used to amplify DNA segments.

Applications:

Taxon identification: RAPD can be used to identify microbes.

Gene mapping: RAPD can be used to map genes in microbes.

Population genetics: can be used to study the genetic structure of microbial populations.

Molecular evolutionary genetics: can be used to study the evolution of microbes.

Applications of genetic engineering:

Industry; Agriculture; medicine:

Genetic engineering is used in many industries, including agriculture, medicine and manufacturing.

Agriculture:

(1) Crop protection: Genetic engineering can help protect crops from disease and pests. For example, genetic engineering was used to create papayas that are resistant to the

C-DNA Library:

* A CDNA (complementary DNA) library is a collection of CDNA molecules derived from mRNA.
* Unlike genomic libraries, CDNA libraries represent only the expressed genes of an organism, excluding non expressed genomic regions such as introns and other non-coding sequences.
* C-DNA libraries are useful for studying gene expression, protein functions, and producing recombinant proteins.
* CDNA libraries contain smaller DNA fragments.
* It contains only coding regions or expressed genes.
* CDNA libraries generally use plasmid vectors, which are suitable for smaller CDNA fragments

Preparation of CDNA libraries:-

(1) Isolation of mRNA

* construction of CDNA library starts isolation of mRNA from eukaryotic cells.
* mRNA is purified using column purification. oligomeric dT nucleotide coated resins that bind only mRNA with a poly-A-tail.
* they contains a poly-A tail at the 3'end. cell lysate is passed through the poly-T

column. poly A-tails of mRNA molecules bind the oligo-dT sequences.

* Finally the molecules are separated from oligo dT sequences using eluting buffer.

(2) Synthesis of cDNA:

* A synthesis of cDNA molecules from the isolated mRNA.
* a short oligo-dT primer is annealed to the 3' poly-A tails of the mRNA molecule, which initiates the synthesis of the first DNA.
* RNase H enzyme degrades the mRNA strand in the mRNA-DNA hybrid, leaving small RNA fragments that are used as primers.
* DNA ligase enzyme seals the nicks between the newly synthesized DNA fragments of a double-stranded cDNA copy of the mRNA.

(3) cDNA cloning:

* ligation of cDNA molecules into suitable vectors for cloning.
* restriction site linkers or adaptors need to be added to the ends of the cDNA molecules with the vector DNA.
* the cDNA molecules are joined with the vector DNA which creates recombinant DNA molecules.

* (4) Transformation:

* The recombinant DNA molecules are transformed into host cells that can be cultured to produce colonies contain cloned cDNA insert
* the transformed plates cultured on agar plate contains selective medium
* the selectable marker gene on the vector ensures that only cells containing the recombinant DNA colonies from cDNA library.

Applications:

* Gene cloning - can be used to clone genes produce proteins and diagnostic reagents
* Gene expression analysis: can be used to analyze gene expression patterns.
* Gene therapy: can identify functional genes.

HYBRIDOMA TECHNOLOGY

Hybridoma technology is a well established method to produce monoclonal antibodies (mAbs) specific to antigens of interest. hybridoma cell lines are formed via fusion between a short-lived antibody-producing B cell and an immortal myeloma cell.

Principle:

Normal B lymphocyte produced antibody of a single specificity, but can't grow indefinitely so it necessary to immortalize B cells that produce a specific antibody. This is achived by cell fusion of somatic cell hybridization

between a normal antibody producing B cell and a myeloma cell. such fusion derived immortalized antibody producing cell line, are called hybridomas and the antibodies they produced are monoclonal antibodies.

Steps:

(1) Fusion between myeloma cells, which are immortal, do not produced hypoxanthine - guanine phosphosibosyl transferase (HGPRT) and spleen cell, which are $HGPRT^+$, mortal and secretes antibody.

(2) Fusion is achieved by incubating a suspension of two cell types with fusogenic agent like inactivated sendai virus or polyethelyne glycol (PEG)

(3) In the process of cell fusion, first the cell membrane and cytoplasm of the two cells become confluent keeping the two nuclei intact known as heterokaryon some heterokaryons become hybrid upon fusion of nuclei.

(4) Only a small number of cells actually fused. so in the culture mixture of unfused cells (B cell, Myeloma cells), fused with like cell (B cell - B cell, myeloma - myeloma) and B cell - myeloma fusion product obtained.

(5) The cells are transferred into HAT

Unit-III : Techniques in genetic engineering and IPR

(1) Blotting techniques:

Blotting techniques are very widely used analytical tools for the specific identification of desired DNA or RNA fragments from thousands of molecules.

Blotting refers to the processes of immobilization of sample nucleic acids on solid support (nitrocellulose or nylon membrane). the blotted nucleic acids are then used as targets in the hybridization experiments for their specific detection.

The mostly commonly used blotting technique

(1) Southern blotting (for DNA)
(2) Northern blotting (for RNA)
(3) Western blotting (for Proteins)
(4) Dot blotting (DNA / RNA)

(1) Southern blotting: Sir Edwin Southern (1975)

Southern blot is the process of transfer of DNA fragments that are separated by electrophoresis onto a membrane for immobilization and Identification.

Steps involved in southern blotting:

(1) DNA Extraction and purification: Extract DNA from cells or tissues using a suitable method such as phenol-chloroform extraction or a commercial DNA extraction. purify the DNA to remove contaminants and ensure it is suitable for southern blotting.

(2) Restriction Enzyme Digestion: Digest the purified DNA with a specific restriction enzyme that cuts the DNA at a known sequence. this step generates DNA fragments of varying sizes.

(3) Agarose gel electrophoresis: separate the DNA fragments by size using agarose gel electrophoresis. the DNA fragments migrate through the gel, with smaller fragments moving faster than larger ones.

(4) Denaturation and Neutralization: Denature the DNA by treating the gel with an alkaline solution (eg: sodium hydroxide). this step converts the double-stranded DNA into single-stranded DNA. then, neutralize the gel with a neutralizing solution, (eg) Tris-HCl, to restore the pH.

(5) Transfer to membrane: transfer the separated DNA fragments from the agarose gel to membrane (nylon or nitrocellulose) using a process is blotting

(6) Hybridization: Hybridize the transfered DNA fragments with a labeled probe that is complementary to the target DNA sequence.

(7) Detection: Detect the hybridized probe using a suitable method, such as autoradiography

Applications:

* Genetic engineering
* Forensic analysis
* Gene mapping
* Diagnosis of genetic disorders
* Study of gene expression

(2) Northern blotting:

It is a laboratory technique used to detect and analyze RNA molecules in a sample. It involves separating RNA molecules by size using electrophoresis, transferring to a membrane, and then detecting the target RNA sequence using a probe.

Steps involved

(1) **RNA extraction and purification:** Extract RNA from cells or tissues using a suitable method, such as phenol-chloroform extraction or a commercial RNA extraction. purify RNA to remove contaminants and ensure it is suitable.

(2) **Electrophoresis:** separate the RNA molecules by size using agarose gel electrophoresis or ~~PAGE~~ PAGE

(3) **Transer for Membrane:** transfer the separated RNA molecules to a membrane (nylon / nitrocellulose) process called blotting.

(4) **Hybridization:** Hybridize the transferred RNA molecules with a labeled probe that is complementary to the target RNA sequence.

(5) **Detection:** Detect the hybridized probe using a suitable method, such as autoradiography.

(6) **Interpretation:** Interpret the results by analyzing the image obtained by previous step the presence of a band or signal indicate the presence of the target RNA sequence.

Applications:

* Gene expression Analysis
* RNA localization
* RNA stability
* Diagnosis of RNA-related diseases
* Forensic sciences.

(3) Western Blotting:

It is a technique used to detect and analyze specific proteins in a complex mixture. It involves separating proteins by size using electrophoresis, transferring them to membrane and then detecting the target protein using antibodies.

steps involved:

(1) sample preparation: prepare a protein sample from cells, tissues or biological fluids.

(2) SDS-PAGE: separate the proteins by size using sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)

(3) Transfer to membrane: transfer the separated proteins to a membrane, such as nitrocellulose or PVDF, using electroblotting.

(4) Blocking: Block non-specific binding sites on the membrane with a blocking solution.

(5) Primary Antibody incubation: Incubate the membrane with a ~~secondary~~ primary antibody ~~that~~ specific ~~recognizes~~ to the ~~primary antibody~~ target protein.

(6) Secondary antibody incubation: Incubate the membrane with a secondary antibody that recognizes the primary antibody.

(7) Detection: Detect the target protein using a detection method such as chemiluminescence.

(8) Analysis: Analyze the results by comparing the detected protein bands to molecular weight ladder.

Applications:

* Protein expression Analysis
* Protein-protein interaction
* Disease diagnosis
* Vaccine development

(4) Dot Blotting:

It is a laboratory technique used to detect and analyze specific nucleic acids (DNA or RNA) or proteins in a sample, it involves applying a small volume of sample onto a membrane followed by hybridization or antibody binding to detect the target molecule.

Steps involved:

(1) Sample preparation: prepare a sample containing the target nucleic acid or protein.

(2) Membrane preparation: cut a membrane (nylon or nitrocellulose) to the desired size.

(3) Sample Application: Apply a small volume (usually 1-5ul) of sample onto the membrane, creating a 'dot or spot'.

(4) Fixation: Fix the sample onto the membrane using UV light, heat or chemical treatment.

(5) Hybridization: Hybridize the membrane with a labeled probe complementary to the target protein. Antibody binding

(6) Detection: Detect the hybridized probe or bound antibody using a detection method.

(7) Analysis: Analyze the results by comparing the intensity of the dot to a control.

Applications:

* Gene expression Analysis
* DNA or RNA Detection
* protein Expression Analysis
* Diagnostic Testing
* Environmental monitoring
* Food safety Testing

Labeling of DNA:

Labeling of DNA refers to the process of attaching a detectable marker or label to a DNA molecule, allowing it to be visualized, detected or tracked.

Methods of DNA labeling:

(1) Nick Translation: Use DNA polymerase to incorporate labeled nucleotides into a DNA molecule.

(2) Random Priming: Use random primers to initiate DNA synthesis, incorporating labeled nucleotides.

(3) PCR Labeling: Use PCR to amplify a DNA sequences, incorporating labeled nucleotides.

(4) End labeling: Use terminal transferase to add labeled nucleotides to the 3' end of a DNA molecule.

Types:

Radioactive labels: Use radiactive isotopes to label DNA

Fluorescent labels: Use fluorescent dyes, such as FITC or Cy3, to label DNA.

Chemiluminescent labels: Use chemiluminescent tags such as biotin or digoxigenin

Enzymatic labels: Use enzymes such as alkaline phosphate.

Applications:

* Southern blotting - detect specific DNA sequence
* In situ hybridization - Visualize specific DNA sequence in cells or tissues.
* DNA microarrays - Analyze gene expression levels
* Next generation sequencing prepare dna libraries for sequencing.

DNA sequencing - sanger's method:

→ Sanger sequencing is a method that identifies the order of nucleotide bases in DNA based on chain termination by modified nucleotides called dideoxynucleotide triphosphates (ddNTPs).

→ It is also known as dideoxy sequencing or chain termination method.

→ This method developed in 1977 by Frederick sanger.

Principle:

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 terminate DNA elongation.

Steps involved:

1) DNA Template preparation:

* to prepare DNA Template, the DNA of interest should be extracted from source.
* there are several methods of extract of DNA, including - chemical extraction, column-based extraction and magnetic bead based extraction.
* they are identical, single stranded molecules

2) chain Termination PCR:

* Target DNA is amplified using PCR, starts with initial denaturation. they includes template DNA, primers, all four dNTPs, DNA polymerase
* they add all 4 ddNTPs in small amount, each ddNTP is labeled with a fluorescent marker. they lack 3' hydroxyl group for further elongation, causing chain termination.
* the collection of newly synthesized DNA strands of different lengths.
* this reaction is occur in four separate tubes each containing one type of ddNTP.
* sequencing clean-up is done before electrophoresis which removes unincorporated ddNTPs and other contaminants.

3) Separation of DNA Fragments:

* In electrophoresis to separate the DNA fragments to their lengths, either polyacrylamide gel or capillary gel system.
* Fragments from each PCR reaction are run in separate lanes to identify the corresponding ddNTP.

4) Detection and Analysis:

* the fragments are passed through a fluorescence detector, identifies the fluorescent label attached to ddNTP at the end of each DNA fragment
* ddNTPs - ddATP, ddTTP, ddGTP, ddCTP

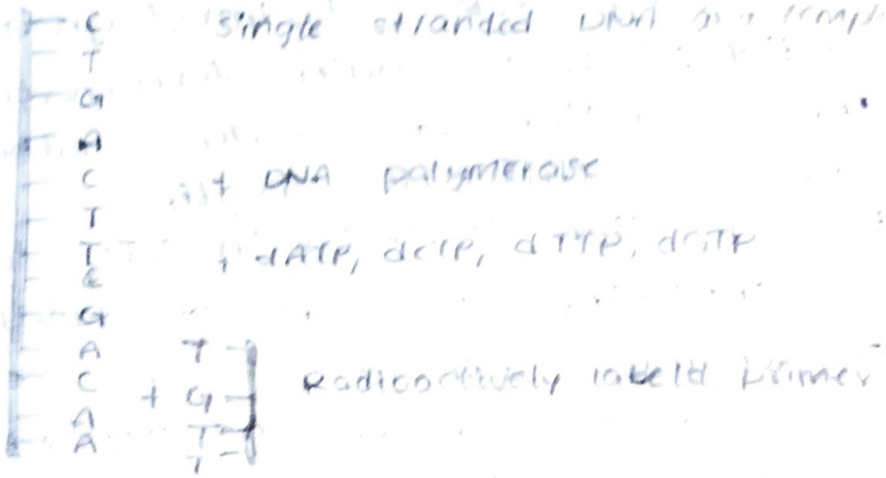
Applications:

Genetic disease diagnosis: Is used to identify genetic mutations with inherited diseases

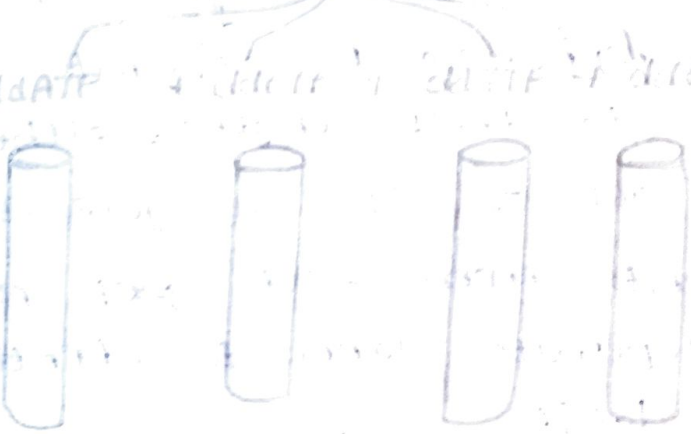
cancer research: to identify mutation in cancer cells.

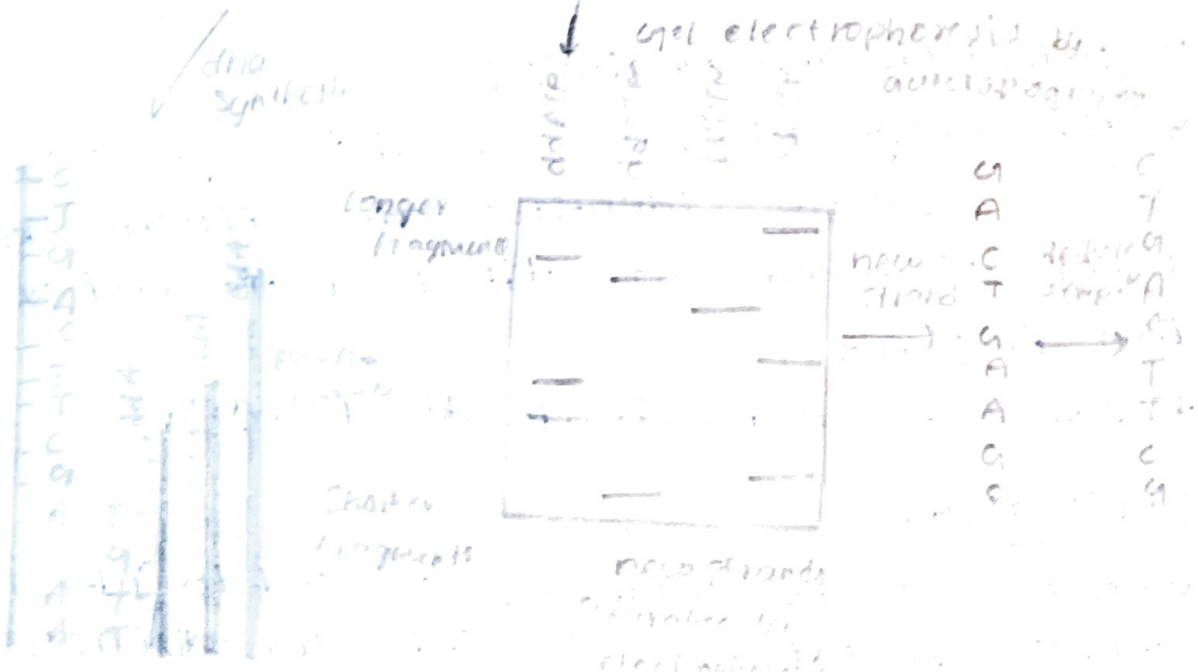
Paternity testing: by analyzing DNA samples from parents and children

Gene discovery: to identify and characterize new genes.



Prepare four reaction mixtures





Outlines of intellectual property rights : (Patents, trademark , copyright)

Intellectual property rights (IPR) are legal rights that protect the use of creations and inventions. they encourage innovation and creativity and protect the interests of creators.

Patents:

* A patent is an exclusive right granted to an inventor for a new and useful invension
* Give the patent owner the right to decide how others can use the invention
* protect inventions, such as products, designs or processes.
* In exchange the patent owner shares technical information about the invention.
* types of patents : utility patents, design patents and plant patents.

Trademark:

* A trademark is a symbol, word, or phrase that identifies a business or product and distinguishes it from others.
* Prevent others from using similar marks that could confuse consumers.
* protect words, phrases, symbols or designs that identify a product or service.
* Types of ~~patents~~ trademarks : word marks, design marks and composite marks.

Copyrights:

* A copyright is an exclusive right granted

Introduction

- Bioinformatics is the science concerned with the development and application of computer hardware and software to the acquisition, storge,analysis, and visualization of biological information.
- It has the following three component.

  - The development of new algorithms and statistics for assessing the relationship among large sets of biological data.

  e.g DNA Sequence data.

  - Application of these tools for the analysis and interpretation of the various biological data. e.g nucleotide sequences, amino acid sequences.

  - The development of database of database for an efficient storage, access and management of various biological informations.
- The 'bioinformatics' is a combination of 'biology' and informatics.

Definition

- Bioinformatics derives knowledge from computer analysis of biological data.
- These can consist of the information stored in the genetic code, but also experimental results from various sources, patient statistics, and scientific literature.
- Research in bioinformatics includes method development for storage, retrieval, and analysis of the data.
- Bioinformatics is a rapidly developing branch of biology and is highly interdisciplinary, using techniques and concepts from informatics, statistics, mathematics, chemistry, biochemistry, and physics.
- It has many practical applications in different areas of biology and medicine.

AD

History of bioinformatics

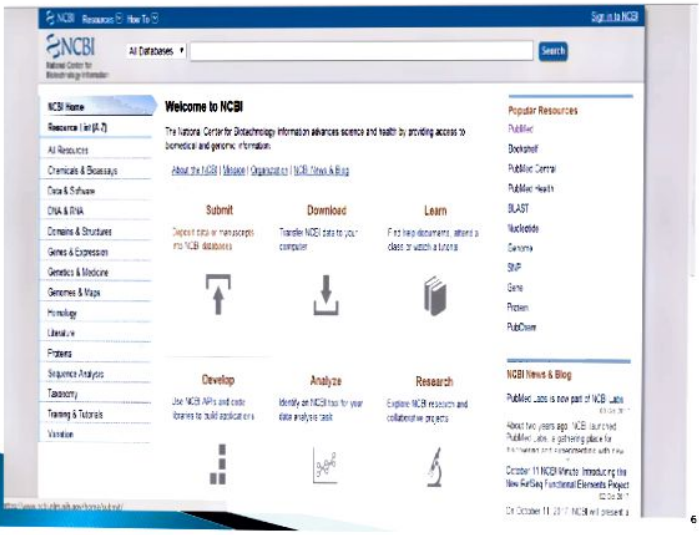
- The collection of amino acids sequences was complied in the 'Atlas of protein sequence and structure' by the National Biomedical Foundation.
- This collection was edited by margaret O.Dayhoff from 1965 to 1978.
- Dayhoff and coworkers contributions to the comparison of amino acid sequences by developing computer software for detecting distantly related sequences.
- The EMBL established their data library in 1980 to collect, organize and distribute nucleotide sequence data and related information.
- NCBI was established in U.S.A. NCBI serves as primary information databank and provider of information.
- The National Biomedical Research Foundation established the PIR in 1984.

DNA Sequences

- The symbols used to represent DNA sequence data.
- The four bases are denoted by single letters A (Adenine), C (cytosine), G (guanine), and T (Thymine)
- But often sequence data contain ambiguities in that it is not clear as to which of the four base present at several positions.
- For example , the sequence data may indicate that the base present at a specific position may be either G or A, it is purine.
- Similarly , if a position may have either C or T, it is pyrimidine.
- The base sequence of the two complementary strands of a DNA molecules are represented by this system of symbols.

NCBI RESOURCES

- PubMed
- OMIM
- Pubchem Compound
- GeneBank



EBI-European Bioinformatics Institute

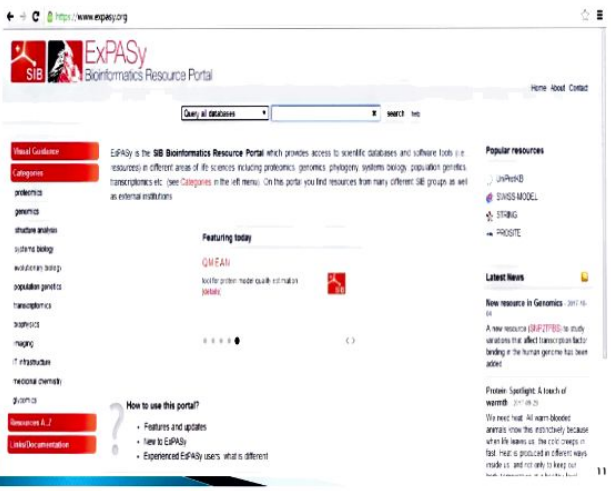
- Centre for research and service in bioinformatics
- Establish in 1980
- Goals – computer database of DNA sequence
- Get the nucleotide sequence for query
- Additional information about query

EMBL-EBI

- Submission tool - webin
- Retrievals system - SRS (sequence retrieval system)
- It provides with the required nucleotide sequence
- Embl database provide user related information

EXPASY

- by the Swiss Institute of Bioinformatics (SIB)
- Establish in 1993
- ExPASy was called ExPASy (Expert Protein Analysis System)
- ExPASy was the first website of the life sciences.



AD

Protein Database

- SWISS-PROT & TrEMBL – Protein sequence database and computer annotated supplement
- UniProt – (Universal Protein Resource) It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.
- PIR – Protein Information Resource

Domain Database

- InterPro – integration of Pfam, PRINTS, PROSITE, SWISS-PROT + TrEMBL
- PROSITE – database of protein families and domains
- Pfam – alignments and hidden Markov models covering many common protein domains

NCBI

- What is NCBI?
- National center for biotechnology information
- Established in 1998
- Part of national library of medicine at national institute of health
- Major aim : public database
- Development of software tools for sequence analysis and disseminate biomedical information

2

2 explain Roles of NCBI

1) Maintenance of biological databases whether primary or secondary. It includes GENEBANK
2) NCBI provides the data retrieval systems such as ENTREZ
3) Provides computational sources for the analysis of the GENEBANK data and other biological data

3

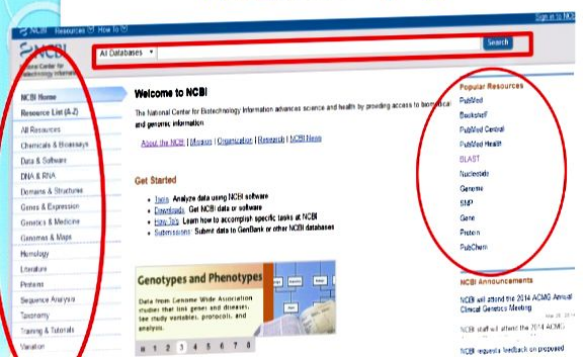
AD

Primary databases	Secondary databases
• Original submission by the experimentalists who have originally searched • Content Is controlled by the submitters • Examples include GENEBANK, SNP and GEO	• Built up from primary data which is retrieved by primary database • Content controlled by third party NCBI • Examples include RefSeq, RefSNP, NCBI Structure, Protein. Etc.

Kinds of databases

4

NCBI homepage



3 Sequence submission to NCBI

- Databases are constantly updated with the newer submissions of the sequences via sequence submission tools such as:
- Bankit
- Sequein

Bank it

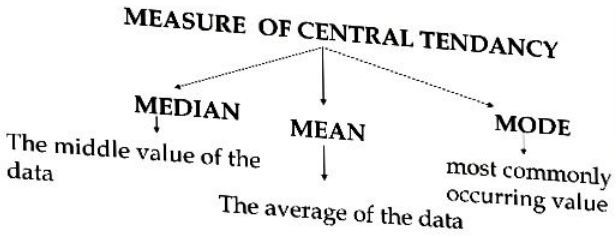
- Web-based sequence submission tool
- Connect to NCBI Home Page
- Connect to GENEBANK side bar at left
- Tool of choice for simple submissions
- Can also be used for updating previously added information

Sequein

- Stand alone sequence submission and updating tool
- Handling multiple sequence submission
- Provides increased capacity for long sequence submissions
- Multiple annotation
- Phylogenetic analysis population

BLAST

- Basic local alignment search tool program
- Sequence similarity searches against a variety of different sequence databases
- Unigene, gene, MMDB, GEO



Mean (Average)

Mean locate the centre of distribution.

Also known as arithmetic mean

Most Common Measure

The mean is simply the sum of the values divided by the total number of items in the set.

Affected by Extreme Values

$$X = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Merits:

- It is easy to understand and easy to calculate
- It is based upon all the observations
- It is familiar to common man and rigidly defined
- It is capable of further mathematical treatment.
- It is affected by sampling fluctuations. Hence it is more stable.

Demerits

- It cannot be determined by inspection.
- Arithmetic mean cannot be used if we are dealing with qualitative characteristics, which cannot be measured quantitatively like caste, religion, sex.
- Arithmetic mean cannot be obtained if a single observation is missing or lost
- Arithmetic mean is very much affected by extreme values.

Median

1. Measure of Central Tendency.

2. The median is determined by sorting the data set from lowest to highest values and taking the data point in the middle of the sequence.

3. Middle Value In Ordered Sequence
   - If Odd n, Middle Value of Sequence
   - If Even n, Average of 2 Middle Value

4. Not Affected by Extreme Values

Merits:

- It is rigidly defined
- It is easy to understand and easy to calculate.
- It is not at all affected by extreme values.
- It can be calculated for distributions with open-end classes.
- Median is the only average to be used while dealing with qualitative data.
- Can be determined graphically.

AD

Demerits:

- In case of even number of observations median cannot be determined exactly.
- It is not based on all the observations.
- It is not capable of further mathematical treatment

Mode

1. Measure of Central Tendency
2. The mode is the most frequently occurring value in the data set.
3. May Be No Mode or Several Modes

AD

Merits:

- Mode is readily comprehensible and easy to calculate.
- Mode is not at all affected by extreme values.
- Mode can be conveniently located even if the frequency distribution has class intervals of unequal magnitude
- Open-end classes also do not pose any problem in the location of mode.
- Mode is the average to be used to find the ideal size.

Demerits:

- Mode is ill defined.
- It is not based upon all the observations.
- It is not capable of further mathematical treatment.
- As compared with mean, mode is affected to a great extent by fluctuations of sampling.

Competency & Learning objectives

Competency	Learning Objectives
CM6.4 Enumerate, discuss and demonstrate Common sampling techniques, simple statistical methods, frequency distribution, measures of central tendency and dispersion	The student should be able to • Define measures of dispersion • Calculate different measures of dispersion (SD, Coefficient of variation, Mean Deviation) • Comment on variation of data

What is measure of dispersion?

- Central tendency measures do not reveal the variability present in the data. Dispersion is measure of variation
- Dispersion is the scattered ness of the data series around it average.
- Dispersion is the extent to which values in a distribution differ from the average of the distribution.

Requirement of good measures of dispersion.

1. It should be rigidly defined.
2. It should be easy to understand and easy to calculate.
3. It should be based on all the observations of the data.
4. It should be least affected by the sampling fluctuation.
5. It should not be unduly affected by the extreme values.
6. It should be capable of further mathematical treatment and statistical analysis

Absolute Measure and Relative Measure

Absolute Measure-

Measure of the dispersion in the original unit of the data. Variability distribution can be compared provided they are given in the same unit and have the same average

Relative Measure-

It is the ratio of absolute measures and unit free measurement. This ratio is known as coefficient of absolute dispersion.

POPULATION

- "A population is a complete set of persons or objects that possess some common characteristics that is of interest to the researcher."
- The population for a study usually is described as being composed of two groups-
- Target population

TARGET POPULATION

- The target population which is also called the universe is composed of the entire group of people or objects to which the researcher wishes to generalize the findings of the study.
- The target population consists of people or things that meet the designated set of criteria of interest to the researcher.

ACCESSIBLE POPULATION

- It is aggregate of cases that confirm to designated criteria and are also accessible as subjects for study.
- By identifying the group from which the study sample was chosen, the investigator about the conclusion of the generalize ability of research findings.

SAMPLE

- **Sample** may be defined as representative unit of target population which is to be worked upon by researchers during their study.
- **Quantitative researchers** often select samples that will allow them to achieve stating conclusion validity and generalize their result. They therefore develop a sampling plan that specifies in advance how study participants are to selected and how many to include.
- **Qualitative researchers** make sampling decision during the course of data collection based on information & theoretical need & typically do not develop a formal sampling plan in advance.

DIFFERENT TYPES OF DATA:

- 1. Primary Data
- 2. Secondary Data
- 3. Qualitative Data
- 4. Quantitative Data

Primary and Secondary Data

- Data can be classified as either **Primary** or **Secondary**.
- **Primary Data:**
- Primary data means original data that has been **collected** specially for the purpose in mind. It means when an authorized organization, investigator or an enumerator collects the data for the first time from the original source. Data collected this way is called primary data.
- **For example:** Your own questionnaire, survey, information.

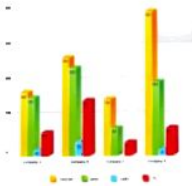
Qualitative Data

- **Qualitative Data** measures a quality or characteristic on each experimental unit. It is a categorical data.

- **Examples:**
  - Hair color (black, brown, blonde, white, grey, mahogany)
  - Make of car (Dodge, Honda, Ford, Toyota)
  - Gender (male, female)
  - Place of birth (Riyadh, Jeddah, Yanbu)

AD

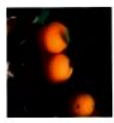
Quantitative Data



- Quantitative data is a numerical measurement expressed in terms of numbers.

- For example: Temperature= "26 degrees"
- Height = "1.8 meters"
- Length = "2.5 feet"
- Age = "9 years"
- Note: Quantitative data always are associated with a scale measure (degree/feet/years).

- **Quantitative Data** measure a numerical quantity on each experimental unit.



Examples

- For each orange tree, the number of oranges is measured.
  - *Quantitative*
- For a particular day, the number of cars entering a college campus is measured.
  - Quantitative
- Time until a light bulb burns out (4 months)
  - Quantitative