Unit-IV: Genetic Engineering:

"Biotechnology is a branch of biology which deals with the techniques of using live organisms, enzymes or biological processes to produce products & provide services for human welfare". Biotechnology is concerned with exploitation of biological components for production of useful products. Term biotechnology- Karl Ereky (1917).

Making curds or bread which involves microorganisms can be considered as the oldest form of biotechnology. Fermentation in the prod: of wine & other alcoholic beverages is also an application of biotechnology. The biotechnological processes gradually became sophisticated. They are being used to produce valuable materials like vitamins and antibiotics produced by microbes.

Now, biotechnology involves DNA manipulations (recombinant DNA technology), tissue culture, protoplast fusion, protein engineering etc. Many techniques included under biotechnology have wide applications in the field of agriculture, medicine, chemical industry, pharmaceutical industry & environment.

* Recombinant DNA technology /- Genetic Engineering :-

Genetic Engineering is defined as "the manipulation of genes by man." To engineer means to design, construct, & manipulate according to a set plan.

Application of genetic engineering ranges from cloning genes to cloning organisms including

Transgenic ~~mat~~ microbes, agriculturally important crops & farm animals. Thus, by the process of genetic engineering, we can get genetically modified food products, human gene products, pharmacologically & therapeutically useful products.

* Tools and techniques of rDNA technology :-

Genetic engineering involves cutting and pasting fragments of DNA. The DNA may be cut & pasted in the vector, whose DNA then becomes the recombinant DNA.

(A) Tools of genetic Engineering :-

The various tools required for genetic engineering are – i) Desired gene, ii) Host, iii) Enzymes & iv) Vectors/Cloning vectors.

(1) Desired gene (DNA) :-

It is the gene which is transferred from one organism to another by combining it with vehicle DNA. It is obtained from donor organism. eg.: Nitrogen fixing gene (NiF gene), insulin gene, etc. It is called passenger DNA.

(2) Host :-

It is the ~~cell~~ organism, where recombinant DNA (rDNA) is allowed to multiply to produce several copies. Bacteria & yeasts are used as host. The bacterium E. coli is commonly used as host.

(3) Enzymes :-

Many kind of specific enzymes are used in genetic engineering.

(a) Lysing enzymes :- These are used to open the cell. eg.: Lysozymes.

(b) Cleaving enzymes :- These are used to break DNA molecule. DNA segment can be excised by "molecular scissors" or "chemical scalpels/knives", which are known as "restriction enzymes" in biotechnological language. They are called 'restriction enzymes' because they restrict infection of bacteria by certain bacteriophages, by degrading viral DNA without affecting the bacterial DNA. Thus, they are helpful in breaking DNA molecule at a specific point. ~~For example~~ For example, a restriction enzyme EcoRI will cut DNA only if sequence GAATC / CTTAG is present in the double-stranded DNA. Hundreds of such restriction enzymes have been isolated from several bacteria & are commercially available. They are named with reference to the organism from which they are isolated. EcoRI is from Escherichia coli & Hae III is from Haemophilus aegypticus.

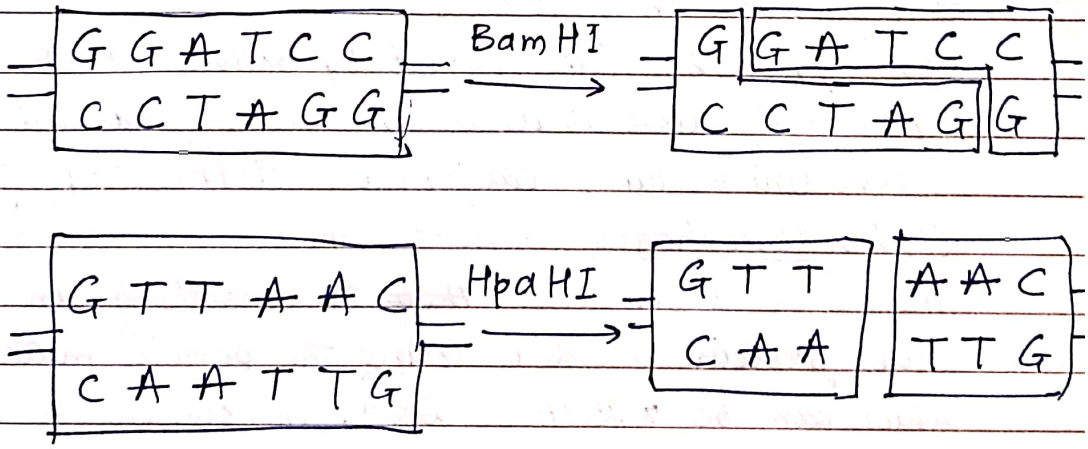


Fig : DNA splicing by restriction endonuclease
BamHI = Bacillus amyloliquefaciens.
HpaHI = Haemophilus parainfluenzae

(c) Joining enzymes :- These enzymes join the DNA fragments by forming phosphodiester bonds. Such enzymes are called DNA ligases.

(d) Synthesizing enzymes :- These enzymes synthesize DNA strands on suitable templates. This DNA is called complementary DNA (cDNA).

eg.: DNA polymerase synthesizes DNA complementary to existing DNA. Reverse transcriptase synthesizes DNA complementary to RNA template.

(4) Cloning Vectors :-

"Cloning vectors are the vehicles which are used to transfer the foreign DNA from one cell to another."

A vector should have following properties :-

(i) An origin of replication,
(ii) Suitable restriction sites,
(iii) Markers for DNA insertion,
(iv) Suitable size, &
(v) High copy number.

The common cloning vectors are as follows :-

(I) Plasmids :-

Plasmid is an extra-chromosomal, circular, double-stranded DNA present in bacterial cells. They have independent origin of replication. Most of the plasmids are not required for the survival of the organism. They are naturally found in bacteria & yeasts.

On the basis of copies per cell, they are of two types —

(a) Relaxed plasmids :- They have high copy number.
(b) Stringent plasmid :- They have low copy number.

The commonly-used plasmids as vectors in genetic engineering are as follows :-

(i) pBR 322

p - denotes it is a plasmid
BR - Boliver & Rodrigues who discovered this plasmid.
322 - It is a number to distinguish this plasmid from other plasmids developed in the same laboratory.

(ii) pBR-327 (iii) pBR 328, (iv) pBR 345.

pUC Vectors :- Series of plasmids constructed in

in University of California & discovered by Messing & Viera. Ex :- pUC 118, pUC 18, pUC 19.

(II) Bacteriophages :-

Bacteriophages are the viruses which infect bacteria. They consist of an outer protein capsid enclosing genetic material, either DNA or RNA. The cloning of a single gene is best carried out by using plasmid. However, for cloning the large pieces of DNA, bacteriophages are used. Larger DNA molecules can be infected in host bacterial cell by bacteriophages. Insertion of viral DNA into host cell is called transfection.

Ex: Commonly used bacteriophages as cloning vectors are M13 and Lambda phage which infect E. coli.

(III) Plant and animal viruses :-

A number of plant & animal viruses are also used as cloning vectors.

eg: Plant viruses :- Cauliflower Mosaic virus (CMV), Tobacco Mosaic Virus (TMV), etc.

Animal Viruses :- Adenovirus, Reterovirus, etc

(IV) Transposons / Jumping genes :- Transposons of higher plants are also used as cloning vectors.

eg: Ds Ac of Maize, Transposons of Drosophila, etc.

(B) ★ Technique of genetic engineering :—

(Gene Cloning)

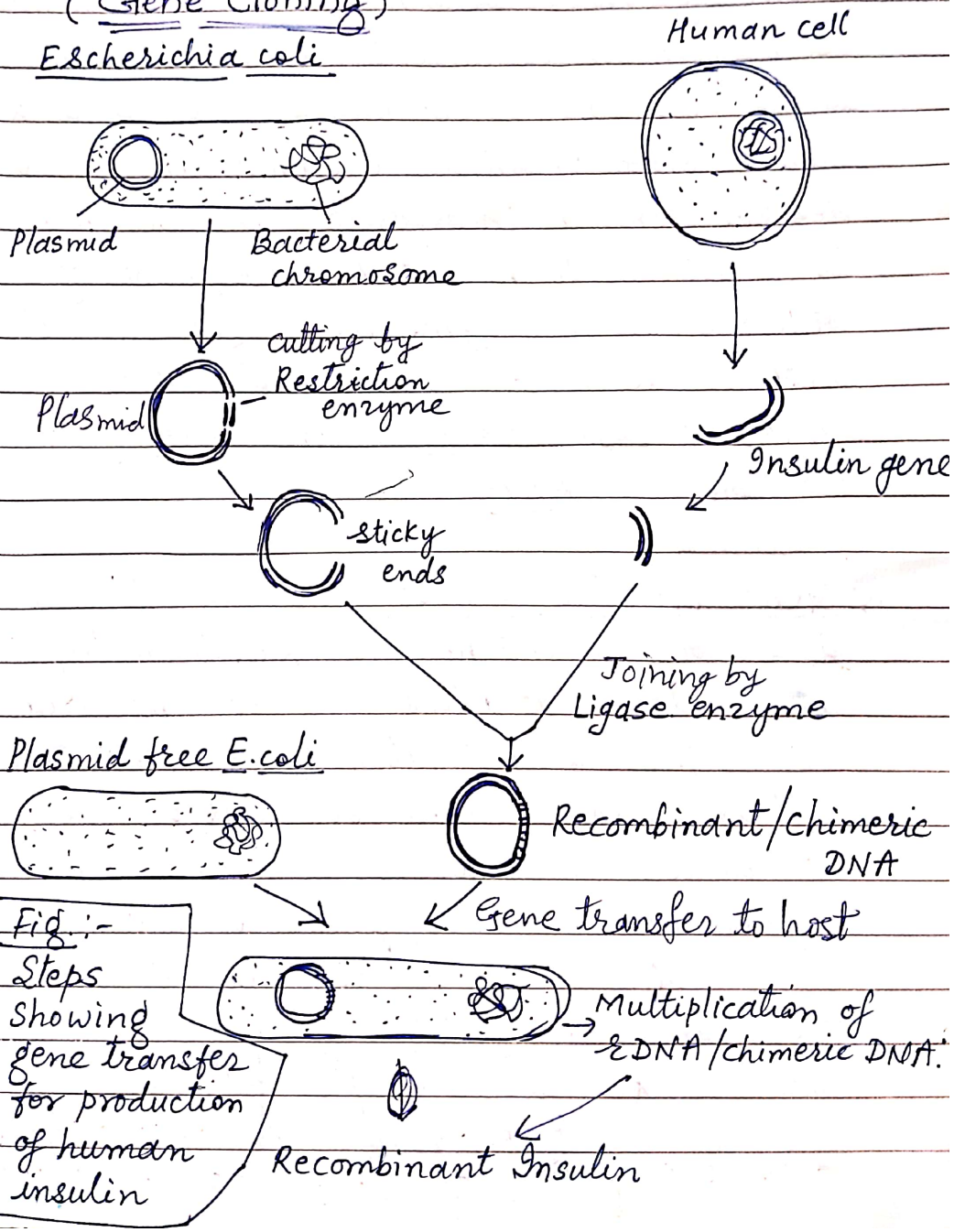


Fig :- Steps showing gene transfer for production of human insulin

The sequence of steps in rDNA technique are as follows :—

1) <u>Isolation of desired gene</u> :- The donor individual having desired gene is selected. The cells are removed and cultured. The cell is broken & DNA is isolated by breaking the nucleus. The desired gene is obtained by cutting the DNA with the help of restriction endonuclease. The DNA containing desired gene is called as <u>passenger DNA</u>.

2) <u>Formation of recombinant DNA</u> :- In this, vector DNA is isolated. The commonly used vector is the plasmid. The plasmid is cleaved with the help of ~~endo~~ restriction endonuclease. Now, the desired gene is incorporated into plasmid and it is joined to the cut DNA segment of plasmid with the help of enzyme DNA ligase. The plasmid DNA containing the desired gene (passenger DNA) is called <u>recombinant DNA (rDNA)</u> or <u>chimeric DNA</u>.

3) <u>Gene Transfer to host</u> :- The recombinant DNA is introduced into host cell. Rapidly multiplying cell is selected as a host cell, eg.: bacterial cell (<u>E. coli</u>). The incorporation of rDNA into host cell is commonly done by <u>electroporation</u>, in which bacterial cell is made permeable by creating temporary pores in cell-membrane.

When the bacterial cell divides rapidly, the plasmid alongwith the desired gene multiplies. This is called gene cloning. The several copies of desired gene are produced. ~~[illegible]~~

Genomic & cDNA Library :-

In Molecular Biology, a library is a collection of DNA fragments from a particular species. It is stored and propagated in a population of micro-organisms through cloning. There are two types of libraries :-
i) Genomic library & ii) cDNA library

(1) Genomic Library :-

It is also called as gene bank. "Genomic library is a collection of the clones of all the DNA fragments, that represent the complete genome of an organism."

Procedure :-

(i) **Isolation of DNA :-** This is the first step in genomic library. The total DNA is isolated from an organism. It is broken down into smaller fragments by using restriction enzymes. The resulting fragments contain all the genes of the organism. If such fragments are cloned, the resulting group of clones is called genomic library, because it contains all genes sequences of the organism.

(ii) **Insertion of DNA into cloning vectors :-** The commonly used plasmid is pUC 19. It is cleaved by suitable restriction enzyme. As a result, circular plasmid is converted into linear DNA molecule containing cohesive ends. The plasmid and DNA fragments are then mixed together, so that the cohesive ends of plasmid & DNA fragments can hybridize to one-another. As a result, there is formation of recombinant plasmid molecules containing inserted DNA segments.

(iii) **Replication of recombinant cloning vectors :-** The recombinant vectors (plasmids) are then transferred to suitable organisms, such as bacteria or yeast, one in each host cells. These organisms are cultured to

produce their clones & are stored..

When genomic DNA is fragmented, it is not known which fragment has the desired gene. Therefore, all the fragments are cloned & copies of each are stored & separately. Screening of desired gene can be done through complementatio or using probes.

Collection of all these recombinant DNA transferred bacterial ~~cell~~ population will constitute the genomic library. Library, thus constructed, can be maintained on agar plates for longer duration & can be used for studies in due course of time.

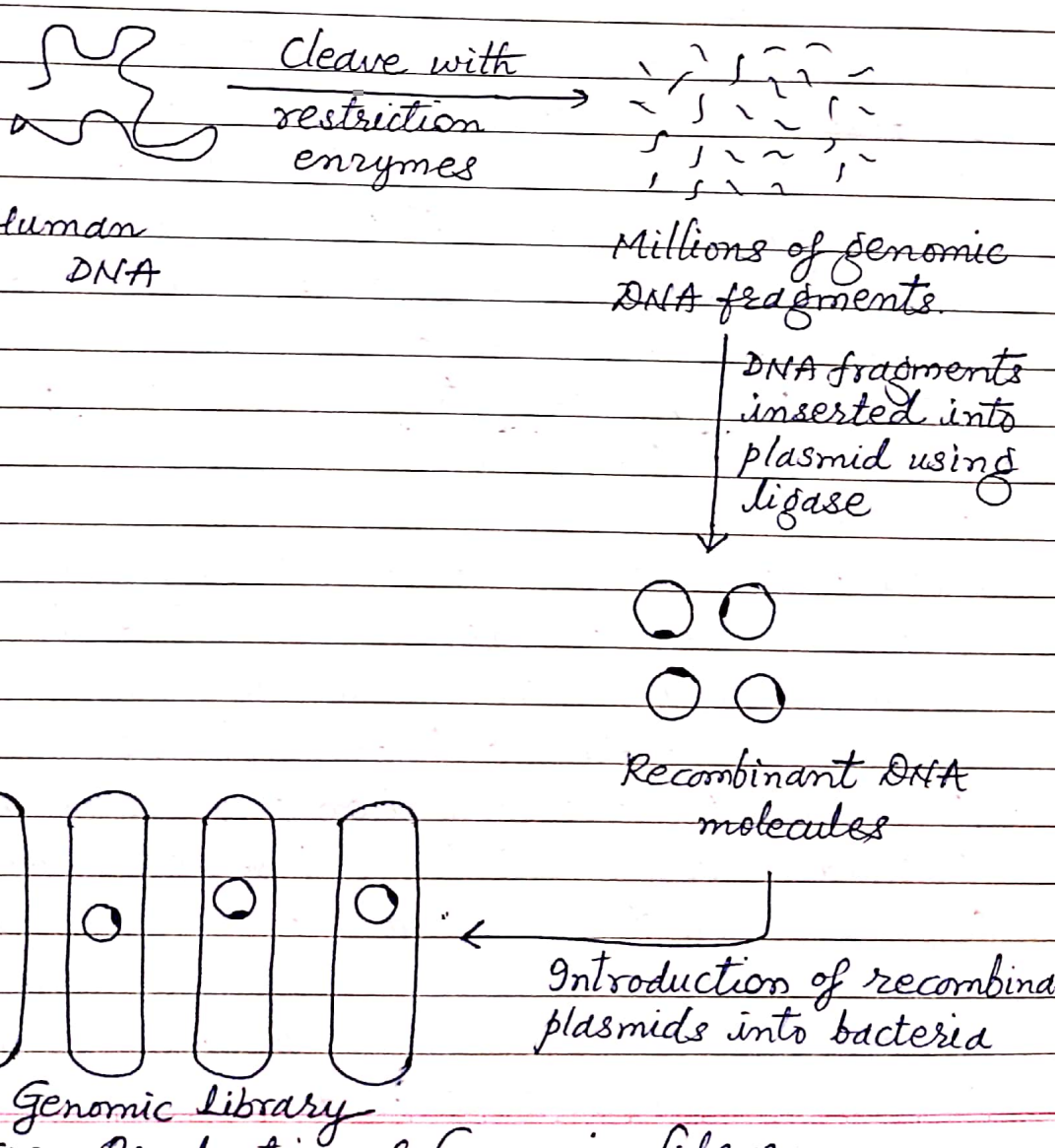


Fig.: Production of Genomic Library

cDNA Library :-

- cDNA is the complementary DNA. It is produced using mRNA by process called as "reverse transcription". This process is also called as "Terminism" named after its discoverers Temin & Baltimore. Mostly for eukaryotic organisms, cDNA library is constructed.
- The enzymatic action of reverse transcriptase on mRNA produces a population of DNA molecules referred to as complementary DNA (cDNA), because they are complementary in sequence to the mRNA employed as template.
- The mRNAs can not be directly cloned because they are unstable. Hence, the entire ~~pop~~ mRNA population of a cell is isolated and copied into cDNA, for cloning. The resulting group of clones is called cDNA library. The advantage of cDNA library is that, it contains the sequences expressed in the form of mRNA, rather than reflecting the entire DNA content of the cell.
- A cDNA library having cDNA for each and every type of protein can be constructed by inserting it into a suitable vector and then cloning in a proper host such as E. coli.
- Production of human proteins, such as interferons, insulin & blood-clotting factor III can be done using bacterial cultures having cDNA.

P.T.O.

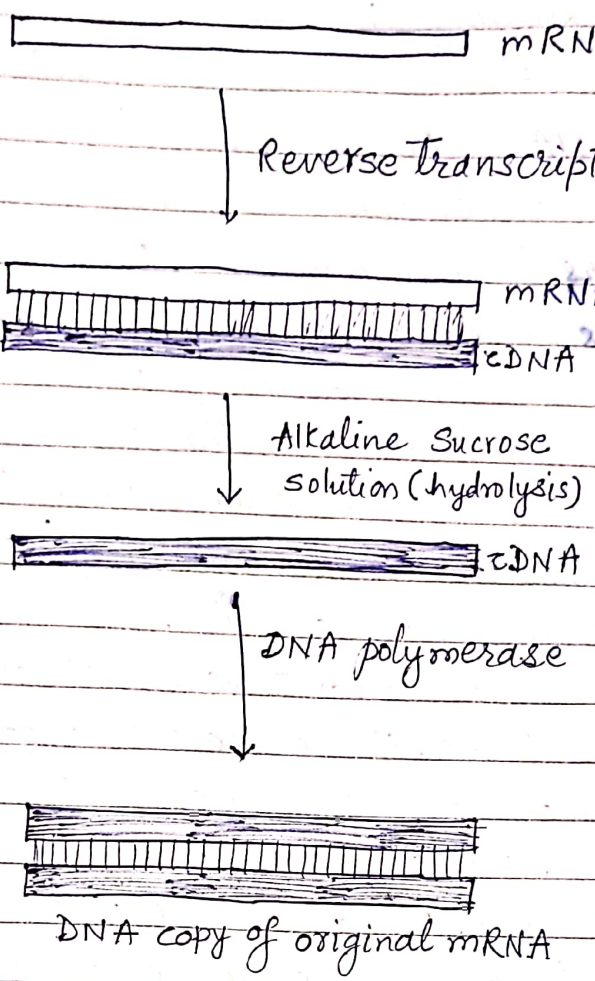


Fig: Synthesis of cDNA from mRNA.

Procedure

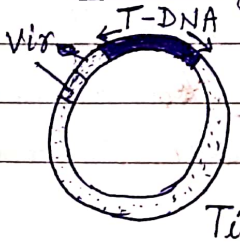
i) Production of cDNA molecules by reverse transcrip.
ii) Insertion of DNA into cloning vectors.
iii) Replication of recombinant cloning vectors.

* Agrobacterium - mediated gene transfer

Agrobacterium tumefaciens is a plant-pathogenic bacterium which can transfer part of its plasmid DNA and infect host plants. Species of Agrobacterium produces crown galls or plant tumors in several dicot plants, such as tomato, potato, tobacco, etc.

These bacteria contain large Ti-plasmids (ie. tumor-inducing plasmids) ~~Agrobacterium tumefaciens~~ which pass on their tumor-causing gene into the genome of host plant. Thus, galls are formed on the host plant & therefore, these bacteria are known as natural genetic engineers of plants. Ti-plasmids can be used as vectors for delivering useful foreign gene into target plant cells & tissues. Ti-plasmid vectors can be used for genetic transformation in many important dicots (eg.: potato, tomato, cotton, sunflower, etc.) and monocot (eg.: wheat, rice, etc.) plants.

The process of infection by Agrobacterium tumefaciens results in transfer of a small part of pTi (tumor-inducing plasmid) into the plant cell genome. This DNA sequence is called T-DNA (transferred DNA). The plasmid naturally transfers the T-DNA into the host plants genome, which makes Agrobacterium a natural genetic engineer.



Ti Plasmid

P.T.O.

Agrobacterium-mediated gene transfer technique :-

Gene transfer thro' Agrobacterium is achieved in the following two ways —
(1) Co-culture with tissue explants, &
(2) In planta transformation.

(1) Coculture with tissue explants :-

(Diagram Page No. 16)

The appropriate gene construct is inserted within the T-region of a disarmed Ti-plasmid. The recombinant DNA is placed in Agrobacterium, which is then cocultured with the plant cells or tissues to be transformed for about 2 days. In case of many plant species, small (a few mm diameter) leaf discs are excised from surface-sterilized leaves & used for coculturing. eg.: tomato, tobacco, petunia, etc.

During leaf disc-Agrobacterium coculture, acetosyringone released by plant cells induces the 'vir' genes, which brings about the transfer of recombinant T-DNA into many of the plant cells. The T-DNA would become integrated into the plant genome and the transgene would be expressed. As a result, the transformed plant cells would become resistant to kanamycin. After 2 days, the leaf discs are transferred on to a regeneration medium containing appropriate concentrations of kanamycin and carbenicillin. Kanamycin allows only transformed plant cells to divide

and regenerate shoots in about 3-4 weeks, while carbenicillin kills Agrobacterium cells. The shoots are separated, rooted & finally transferred into soil.

Agrobacterium infects some monocot plants species & form crown galls, eg. Asparagus or induces swellings, eg: Dioscorea bulbifera, chlorophytum & Allium cepa. Opine is produced by gall or swelling tissues. Integration of T-DNA into the genomes of D. bulbifera and Oryza sativa has been demonstrated, but the efficiency is rather low. Efficient transformation of monocot cells can be obtained by providing acetosyringone during coculture of plant cells with Agrobacterium.

P.T.O. for diagram (P.No. 16)

(2) In Planta Tranformation :—

Imbibition of Arabidopsis seeds ~~with~~ in fresh cultures of Agrobacterium leads to stable integration of T-DNA in the Arabidopsis genome. It appears that, Agrobacterium cells enter the seedlings during germination, are retained within the plant & when flowers develop, they transform either the zygotes or the cells that give rise to zygotes.

Alternatively, Arabidopsis plants about to flower are immersed in a fresh culture of Agrobacterium & partial vacuum is created to facilitate the entry of bacterial cells into the plants. The plants are grown, selfed & the

progeny so obtained are screened for the identification of transformants.

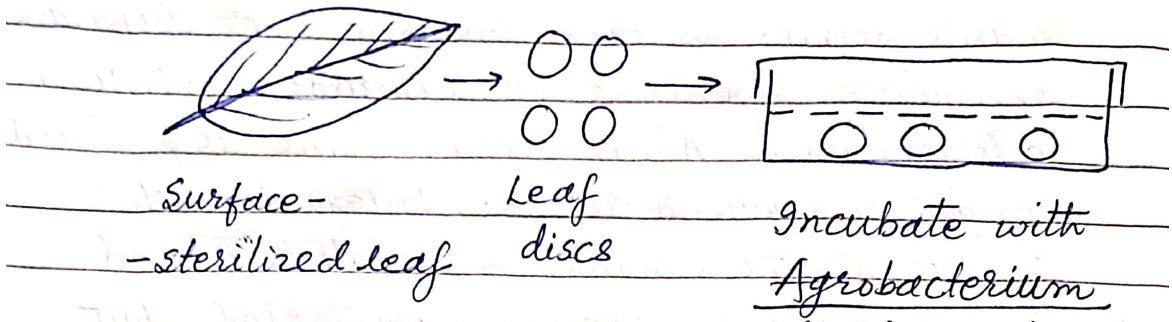


Fig. : Procedure for Agrobacterium-mediated gene transfer. (Coculture with tissue explants)

* Transgenic Plants :-

"The plants in which one or more foreign genes are incorporated through recombinant DNA technology, are called as transgenic plants or genetically modified plants or GM plants."

The foreign genes which are incorporated in GM plants are called transgenes and the process of incorporation is called transgenesis or transformation.

So far more than 60 transgenic dicot plants have been produced. These transgenic plants contain selected traits, such as pathogen-resistance, insect-resistance, herbicide-resistance, ~~etc~~ seed-storage protein, modified ripening, etc.

Examples of transgenic plants :-

See page No. 201

1) Bt Cotton - Resistant to boll worm.
2) Flavr Savr tomato - Fruits have more shelf life.
3) Golden rice - High vitamin 'A' content.
4) Transgenic potato - Protein-rich potato.
5) Tobacco 6) Brinjal 7) Sunflower ~~8) Cotton~~ 8) Beet 9) Pea 10) Maize 11) Wheat 12) Soybean.

(a) Insect-resistant transgenic plants :-

The Bt gene (cry gene) of a bacterium Bacillus thuringiensis has been found to encode the toxin called "endotoxin", which is toxic to certain insect pests. When specific insects (species of Lepidoptera, Diptera, Coleoptera, etc.) ingest the toxin, they are killed.

The Bt gene is isolated & introduced into Ti-plasmid of Agrobacterium tumifaciens. The genetically modified Agrobacterium is allowed to infect the desired plant. Thus, Ti-plasmid mediated transformation of some

plants has been done, eg: tobacco, cotton, tomato, corn, etc. Some examples of Bt-Crops are brinjal, cauliflower, cabbage, canola, corn, potato, tobacco, tomato, rice, soybean, etc.

The second insect resistant gene is cowpea trypsin inhibitor (CpTI) gene, due to which cowpea is resistant to the attack of major storage pest of seeds called 'bruchid beetle'. CpTI has been found toxic in nature to many insect pests & hence it is incorporated into Agrobacterium. The genetically engineered Agrobacterium is inserted into tobacco.

(b) Herbicide-resistant transgenic plants :-

Herbicides are used in agriculture for killing the weeds (the unwanted plants). However, the herbicides disturb the metabolic activity of photosynthesis & synthesis of amino acids. In addition, due to their overuse, environmental pollution occurs. Therefore, biodegradable new herbicides are being developed, which will be ecofriendly & environmentally safe.

eg ①. Herbicide-resistance is a genetic trait. The herbicide tolerance gene expressed enzyme which detoxifies the herbicide & tolerate the effects. ② Transgenic plant Roundup Ready has been produced, which is tolerant to the herbicide Roundup.

eg: ② A gene resistant to PPT (L-phosphino-thricin), an active ingredient of herbicide 'Basta' was isolated from Medicago sativa &

incorporated into tobacco, as a result, transgenic tobacco was produced which was resistant to PPT.

A number of microorganisms have ability to degrade herbicides. This degradation is due to genes encoding specific enzymes. By using these genes, several transgenic crops are produced, such as transgenic potato, sugarbeet, tomato, corn, cotton, soybean, etc. which are herbicide-resistant.

(C) Pathogen-resistant transgenic plants :-

Genetic engineering has also been used in the development of such crop varieties which are resistant to certain diseases caused by fungi, bacteria, viruses & nematodes.

The most successful approach for the production of virus-resistant plants is the transfer of virus protein coat gene into the plants.

The gene that encodes coat protein is isolated from the genome of the virus that causes concerned disease. Now this gene is transferred & expressed in the host. The coat protein produces resistance in the host to the virus. This approach has been used in producing a virus-resistant variety of squash. Such varieties are successful in reducing the yield losses due to various crop diseases and thus enhance agricultural production.

Transgenic pathogen-resistant plants include tomato, potato, squash, papaya, etc.

Continued from page No. 17 :----

Technique of developing transgenic plants :-

(i) Isolation of desired gene :- The organism containing desired gene is selected. The desired gene is isolated from the selected organism.

(ii) Incorporation of desired gene in the vector :- Vector is the vehicle which transfers the desired gene ~~into~~ the plant genome. Commonly used vector is the Ti-plasmid (Tumor inducing plasmid) present in the soil bacterium Agrobacterium tumefaciens. This bacterium produces galls or tumors in many dicot plants. The Ti-plasmid is isolated and cut open at the T-region with the help of ~~restriction~~ restriction enzyme & isolated desired gene is joined in this region with the help of ligase enzyme. Thus the Ti-plasmid gets integrated with desired gene & called recombinant DNA.

(iii) Cloning of desired gene :- Cloning is the process of production of several identical copies of desired gene. For this, rDNA is incorporated in Agrobacterium. As the Agrobacterium divides, the plasmid & desired gene are also replicated. ~~Thousands of~~

(iv) Introduction into plant genome :- The bacterium containing rDNA is introduced into plant cell. The desired gene gets incorporated into plant genome.

(v) Culture of plant cells to produce new plant :- The plant cell is cultured by tissue culture method and converted into a new plant. All the cells of this new plant contain the desired gene.

Such new plant produced by introducing foreign gene is called as transgenic plant.

Bioinformatics

* Introduction :-

Bioinformatics is the use of computers to handle biological information. It is the information technology dealing with maintenance and use of data of molecular biology using computers. In short, it is information technology applied to molecular biology. The term Bioinformatics originated from Two epithets - 'Bio' means living & 'Informatics' meaning information science. According to National ~~Board of~~ center for Biotechnology information (NCBI), Bioinformatics is the field of ~~computer~~ science, in which Biology, Computer Science & Information Technology merge into a single discipline.

Bioinformatics is the Science of developing computer databases & algorithms for the purpose of speeding up and enhancing biological research. The computers are used to gather, store, analyse and merge biological data.

Four components construct the framework of Bioinformatics. They are as follows :-

(i) Genomics :- Genomics deals with the study of genomes of organisms. It includes sequencing, assembling & mapping complete genome of organisms. Human genome is completely sequenced under Human Genome Project.

(ii) Transcriptomics :- Transcriptomics deals with the study of expression of genes through mRNA. It helps to understand several molecular mechanisms & pathways.

(iii) Proteomics :- Proteomics deals with the study of structure & functions of proteins. It helps to understand protein-protein interaction and the ~~3~~ 3D structure of proteins.

(iv) Metabolomics :- Metabolomics deals with the study of cellular processes & components involved in it. This field develops the profile of metabolites involved in cellular processes of organisms.

DNA — Genomics

↓

RNA — Transcriptomics

↓

Proteins — Proteomics

↓

Enzymes — Metabolomics

Fig : Components of bioinformatics

Bioinformatics involves :-

(i) Use of different computer software tools to create databases of genomics, proteomics, metabolomics, transcriptomics, etc.

(ii) Use of experimental methods like X-ray crystallography & NMR spectroscopy for determination of three dimensional structural aspects of genes and product of gene (protein molecule).

(iii) Development of algorithms for utilization of & management of various databases in computer modelling.

(iv) Use of this knowledge in Genetic engineering to synthesize desired product.

* Importance of Bioinformatics :

i) Bioinformatics has gained importance through Human Genome Project.

ii) It is a multidisciplinary subject involving Biology, Computer Science, Physics, Chemistry, Mathematics, Statistics & Medicine.

iii) It deals with the practice of storing, searching & distribution of biological data.

iv) It is used to analyse the sequence data of nucleic acids & proteins to study the molecular structure.

v) It develops computational tools, database and methods for biological research.

vi) It provides graphical interface for biological research like drug designing, protein engineering, phylogenetic analysis, etc.

vii) Bioinformatics has greatly decreased the duration of drug discovery. ie. from 15 years to 3-4 years.

viii) Being an young field, it attracts young researchers due to the scope of computers in recent years.

ix) It has now become impossible to perform biological research without the role of computer in it.

x) Bioinformatics has brought out several evolutionary significances based on molecular study of on various organisms.

* Biological Databases :-

A biological database is a collection of data that is organized so that its contents can be easily be assessed, managed & updated. The activity of preparing a database can be divided into -

i) collection of data in a form which can be easily accessed.

ii) Making it available to a multi-user system (always available for the users).

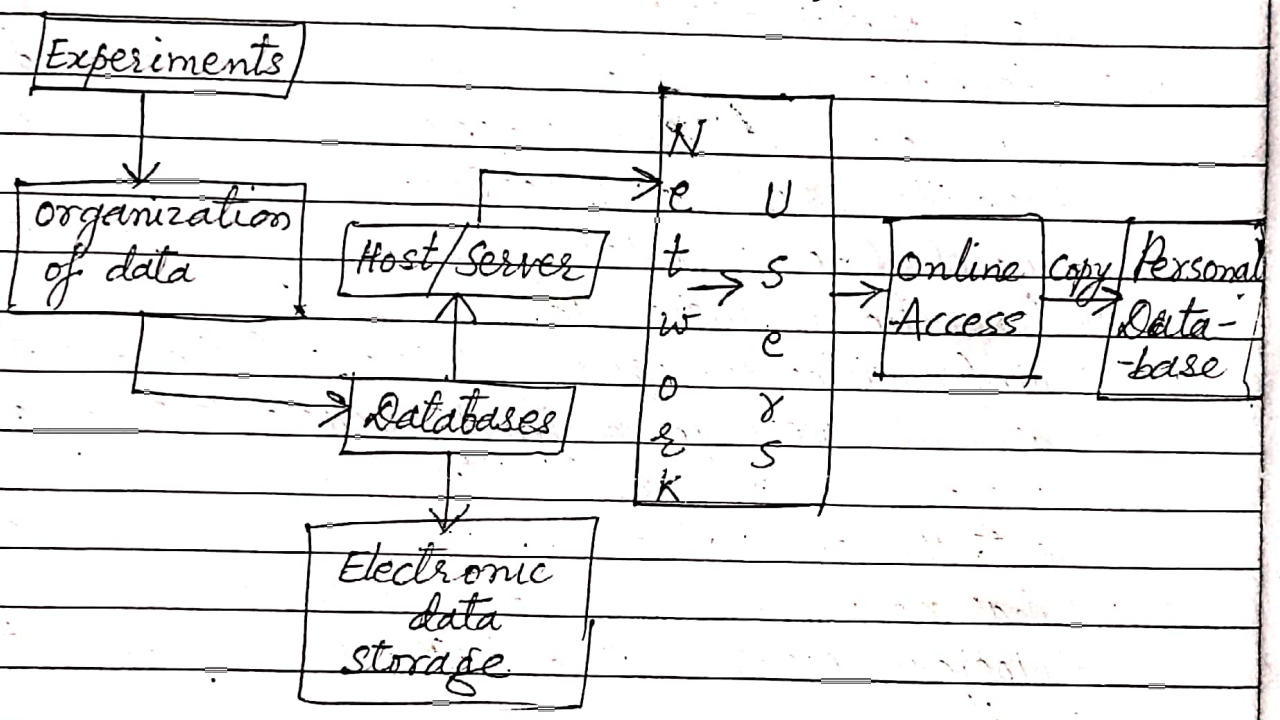


Fig: The network for production, construction & accession of a database.

Database is a repository for collection of computerized data files. It has several advantages over traditional, paper based methods of record keeping :-

i) Compactness, ii) Speed of operation (fast),
iii) Up-to-date information available on demand at any time.
iv) Accuracy & consistency of information.

Biological databases are of two types - sequence databases (for nucleic acids & proteins) & & structural databases (for proteins).

Major databases in Bioinformatics are –
: NCBI and PubMed

(1) NCBI :-

(National Center for Biotechnology Information)

The NCBI was established in 1988 as a division of National Library of Medicine (NLM) & is located on the campus of National Institute of Health (NIH) in Bethesda, Maryland (USA). NCBI maintains GenBank, the NIH-DNA sequence database. It has repository of publicly available genomic and proteomic data.

NCBI houses a series of databases relevant to biotechnology & biomedicine. NCBI is easily available online through entering in search engine. NCBI is directed by David Lipman.

NCBI has responsibility of making available GenBank for DNA sequence database. GenBank coordinates with ~~the~~ individual laboratories & other sequence databases, such as European Molecular Biology Laboratory (EMBL) & the DNA Data Bank of Japan (DDBJ).

NCBI provides facility of books, which are freely downloadable. Online version of selected biomedical books are also available, which includes the topics like molecular biology, biochemistry, cell biology, genetics, microbiology, virology, etc.

NCBI provides –

i) the molecular modelling database (3D prot^{n} structure)
ii) dbSNP (a database of single nucleotide polymorphism)
iii) a reference sequence collection.
iv) a map of human genome.
v) co and coordinates with National Cancer Institute & provides cancer genome anatomy project.

(2) PubMed :-

PubMed stands for publishers on Medicine. It is a biological database. It is search database. It helps to access MEDLINE Database.

PubMed is maintained by National Library of Medicine (NLM) of USA. PubMed offers access to articles on medicine, nursing & health disciplines. Full text articles on Biochemistry, Cell Biology and Microbiology are also obtained from this site. About 14,000 users visit the PubMed to seek information per day.

PubMed tries to answer clinical queries submitted by people all over the world to it. PubMed tool provides clipboard facilities to use PubMed database. PubMed Reader is alternative interface for PubMed. PubMed Central is a digital archieve of biomedical & life science journals. It is available at: http://www.pubmedcentral.com. The opening page of PubMed central shows a box for advanced search. If a name of article is to be searched, it is typed in the box & clicked the Find Article button. The particular article is displayed on the screen. PubMed also facilitates to select a journal & visualise all articles published in that journal.

Uses of PubMed :-

(i) It provides articles on medicine, nursing & health.
(ii) It provides articles on cell Biology, Biochemistry and Microbiology.
(iii) It provides articles free of cost.
(iv) It gives answers to clinical queries.

Page No. : VEDHA
Date : / /

* BLAST :-

(Basic Local Alignment Search Tool)

BLAST is a popular program (software) for searching bio sequences against databases. BLAST was developed and maintained by a group at National Center for Biotechnology Information (NCBI). BLAST was proposed by Altschul et al. (1990). It is ~~popu~~ used for finding high scoring local alignments between two sequences.

BLAST is used to determine whether a particular DNA or protein structure matches any other DNA or protein sequence in one of several public databases. BLAST can be ~~acce~~ accessed through the internet on the website of National Center for Biotechnology Information (NCBI).

BLAST is algorithm for comparing primary biological sequence information, such as amino acid sequences of different proteins or nucleotide sequences of DNA. For example, following the discovery of a previously unknown gene in the mouse, a scientist will typically perform a BLAST search of the human genome to see whether humans carry a similar gene. BLAST will identify sequences in the human genome that resemble the mouse gene based on similarity of the sequence.

Depending on the types of sequences to compare, there are different blast programs, such as BLASTP, BLASTN, BLASTX, TBLASTN, TBLASTX, TBLASTZ, etc.