DNA LIBRARY/ GENE BANK

A **DNA library** is a collection of **DNA** fragments that have been cloned into suitable vectors usually a plasmid or bacteriophage or viral vector, from where they can be isolated for study.

The first DNA-based genome ever fully sequenced was achieved by Frederick Sanger, in 1977. Sanger and his team created a library of the bacteriophage, phi X 174, for use in DNA sequencing

DNA library technology is a mainstay of current molecular biology, genetic engineering, and protein engineering, and the applications of these libraries depends on the source of the original DNA fragments. There are differences in the cloning vectors and techniques used in library preparation, but in general each DNA fragment is uniquely inserted into a cloning vector and the pool of recombinant DNA molecules is then transferred into a population of bacteria

Types of gene library-There are different types of gene library. They are- Genomic library , cDNA library, Mutant library etc. Among these Genomic library and cDNA library are the two main types.

Genomic library:- A genomic library is a collection of the total genomic DNA from a single organism. The DNA is stored in a population of identical vectors, each containing a different insert of DNA. In order to construct a genomic library, the organism's DNA is extracted from cells and then digested with a restriction enzyme to cut the DNA into fragments of a specific size. The fragments are then inserted into the vector using DNA ligase.[1] Next, the vector DNA can be taken up by a host organism - commonly a population of Escherichia coli or yeast - with each cell containing only one vector molecule. Using a host cell to carry the vector allows for easy amplification and retrieval of specific clones from the library for analysis

Applications of genomic libraries include:

- Determining the complete genome sequence of a given organism (see genome project)
- Serving as a source of genomic sequence for generation of transgenic animals through genetic engineering
- Study of the function of regulatory sequences *in vitro*
- Study of genetic mutations in cancer tissues

Genomic library construction-

Construction of a genomic library involves creating many recombinant DNA molecules. An organism's genomic DNA is extracted and then digested with a restriction enzyme. For organisms with very small genomes *(~10 kb)*, the digested fragments can be separated by gel electrophoresis. The separated fragments can then be excised and cloned into the vector separately. However, when a large genome is digested with a restriction enzyme, there are far too many fragments to excise individually. The entire set of fragments must be cloned together with the vector, and separation of clones can occur after. In either case, the

fragments are ligated into a vector that has been digested with the same restriction enzyme. The vector containing the inserted fragments of genomic DNA can then be introduced into a host organism.

Below are the steps for creating a genomic library from a large genome.

1. Extract and purify DNA.
2. Digest the DNA with a restriction enzyme. This creates fragments that are similar in size, each containing one or more genes.
3. Insert the fragments of DNA into vectors that were cut with the same restriction enzyme. Use the enzyme DNA ligase to seal the DNA fragments into the vector. This creates a large pool of recombinant molecules.
4. These recombinant molecules are taken up by a host bacterium by transformation, creating a DNA library.

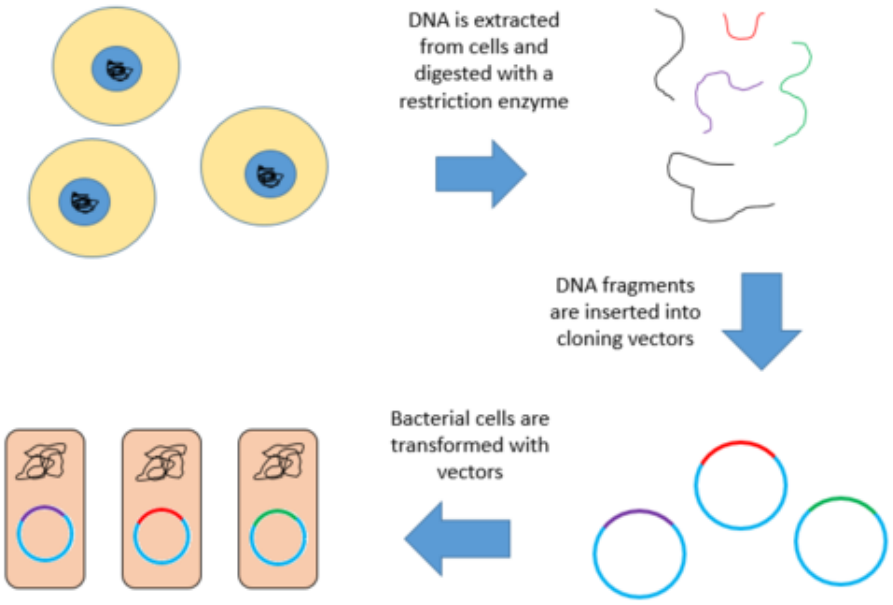

Fig- Genomic library construction

cDNA libraries:- cDNA library represents a sample of the mRNA purified from a particular source (either a collection of cells, a particular tissue, or an entire organism), which has been converted back to a DNA template by the use of the enzyme reverse transcriptase. It thus represents the genes that were being actively transcribed in that particular source under the physiological, developmental, or environmental conditions that existed when the mRNA was purified. cDNA libraries can be generated using techniques that promote "full-length" clones or under conditions that generate shorter fragments used for the identification of "expressed sequence tags".

cDNA libraries are useful in reverse genetics, but they only represent a very small (less than 1%) portion of the overall genome in a given organism.

Applications of cDNA libraries include:

- Discovery of novel genes
- Cloning of full-length cDNA molecules for *in vitro* study of gene function
- Study of the repertoire of mRNAs expressed in different cells or tissues
- Study of alternative splicing in different cells or tissues

Construction of cDNA Library:

cDNA is created from a mature mRNA from a eukaryotic cell with the use of reverse transcriptase. In eukaryotes, a poly-(A) tail distinguishes mRNA from tRNA and rRNA and can therefore be used as a primer site for reverse transcription.

Firstly, the mRNA is obtained and purified from the rest of the RNAs. Several methods exist for purifying RNA such as trizol extraction and column purification. Column purification is done by using oligomeric dT nucleotide. The mRNA is eluted by using eluting buffer and some heat to separate the mRNA strands from oligo-dT.

Once mRNA is purified, *oligo-dT* (a short sequence of deoxy-thymidine nucleotides) is tagged as a complementary primer which binds to the poly-A tail providing a free 3'-OH end that can be extended by reverse transcriptase to create the complementary DNA strand. Now, the mRNA is removed by using a RNAse enzyme leaving a single stranded cDNA (sscDNA). This sscDNA is converted into a double stranded DNA with the help of DNA polymerase. However, for DNA polymerase to synthesize a complementary strand a free 3'-OH end is needed. This is provided by the sscDNA itself by generating a *hairpin loop* at the 3' end by coiling on itself. The polymerase extends the 3'-OH end and later the loop at 3' end is opened by the scissoring action of *S_1 nuclease*. Restriction endonucleases and DNA ligase are then used to clone the sequences into bacterial plasmids.

Vector and Types of vectors for DNA library

Vectors play an important role in the construction of DNA library, as they rear the DNA fragments. Genome size varies among different organisms and the cloning vector must be selected accordingly. For a large genome, a vector with a large capacity should be chosen so that a relatively small number of clones are sufficient for coverage of the entire genome.

Example of some vectors commonly used for genomic libraries and the insert size that each generally holds is showing in the following Table-

Vector type	Insert size (thousands of bases)
Plasmids	up to 10
Phage lambda (λ)	up to 25
Cosmids	up to 45
Bacteriophage P1	70 to 100

Bacterial artificial chromosomes (BACs)	120 to 300
Yeast artificial chromosomes (YACs)	250 to 2000

Plasmids

A plasmid is a double stranded circular DNA molecule commonly used for molecular cloning. Plasmids are generally 2 to 4 kilobase-pairs (kb) in length and are capable of carrying inserts up to 15kb. Plasmids contain an origin of replication allowing them to replicate inside a bacterium independently of the host chromosome. Plasmids commonly carry a gene for antibiotic resistance that allows for the selection of bacterial cells containing the plasmid. Many plasmids also carry a reporter gene that allows researchers to distinguish clones containing an insert from those that do not.

Phage lambda (λ)

Phage λ is a double-stranded DNA virus that infects *E. Coli* The λ chromosome is 48.5kb long and can carry inserts up to 25kb. These inserts replace non-essential viral sequences in the λ chromosome, while the genes required for formation of viral particles and infection remain intact. The insert DNA is replicated with the viral DNA; thus, together they are packaged into viral particles. These particles are very efficient at infection and multiplication leading to a higher production of the recombinant λ chromosomes. However, due to the smaller insert size, libraries made with λ phage may require many clones for full genome coverage.

Cosmids

Cosmid vectors are plasmids that contain a small region of bacteriophage λ DNA called the cos sequence. This sequence allows the cosmid to be packaged into bacteriophage λ particles. These particles- containing a linearized cosmid- are introduced into the host cell by transduction. Once inside the host, the cosmids circularize with the aid of the host's DNA ligase and then function as plasmids. Cosmids are capable of carrying inserts up to 40kb in size.

Bacteriophage P1 vectors

Bacteriophage P1vectors can hold inserts 70 – 100kb in size. They begin as linear DNA molecules packaged into bacteriophage P1 particles. These particles are injected into an E. coli strain expressing Cre recombinase. The linear P1 vector becomes circularized by recombination between two loxP sites in the vector. P1 vectors generally contain a gene for antibiotic resistance and a positive selection marker to distinguish clones containing an insert from those that do not. P1 vectors also contain a P1 plasmid replicon, which ensures only one copy of the vector is present in a cell. However, there is a second P1 replicon- called the P1 lytic replicon- that is controlled by an inducible promoter. This promoter allows the amplification of more than one copy of the vector per cell prior to DNA extraction.

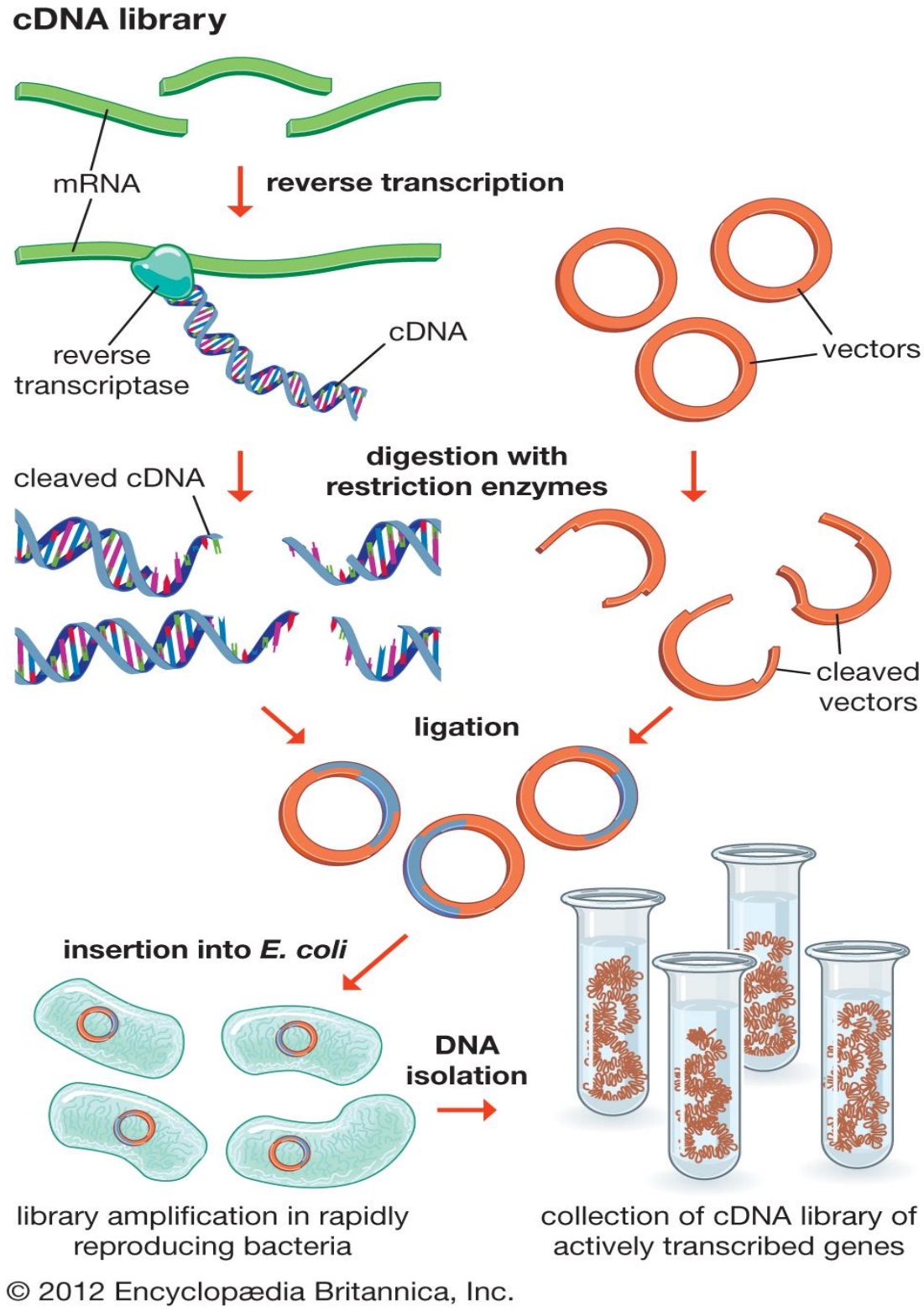

Fig- Steps of cDNA library construction