Lesson Prepared Under MHRD project " National Mission on Education Through ICT"

Discipline: Botany

Paper: Plant Biotechnology

National Coordinator: Prof. S.C. Bhatla

Lesson: Construction of genomic and cDNA libraries

Lesson Developer:Dr. Vibha G. Checker

Department/College:Kirori Mal College, University of Delhi

Lesson Reviewer: Dr Manisha Agarwal

Department/College: Genetics, UDSC

Language Editor: Dr Namrata Dhaka

Department/College:Genetics, UDSC

Lesson Editor: Dr Rama Sisodia, Fellow in Botany ILLL

Table of Contents

Chapter: Construction of genomic and cDNA libraries

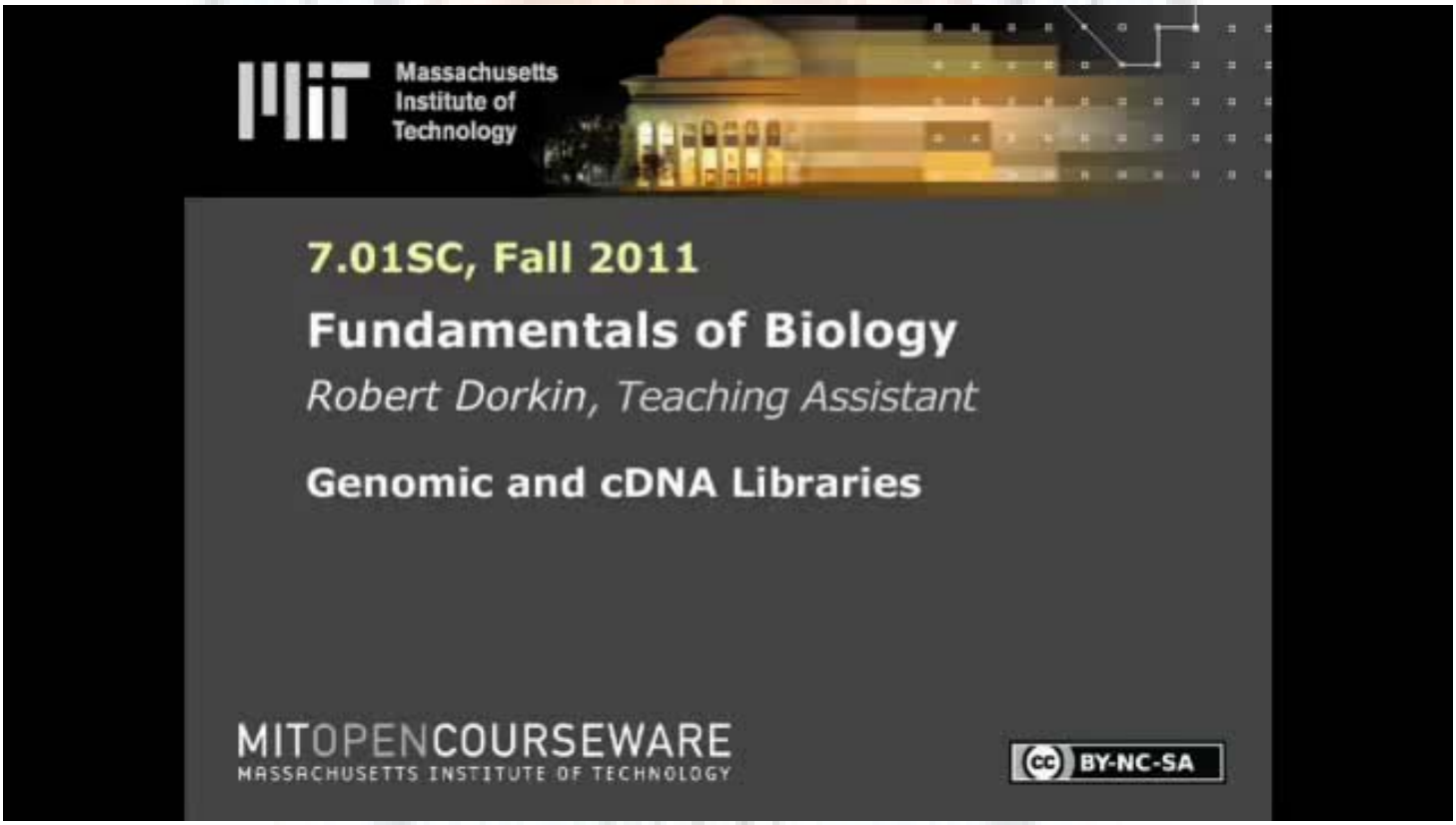
Introduction

In this chapter, we shall discuss about the ways in which whole genome of an organism can be fragmented into small segments and cloned into vectors, for various purposes.

Genome and Genomics

Dr. Tom Roderick, a geneticist at the Jackson laboratory, China, coined the term "genomics" which refers to the study of entire genome of organisms to gain insights about its structure and function. The term "genome" originated from the German word *'Genom'*, and was first used by Hans Winkler. Genes present in all thechromosomes of a particular organism constitute its genome and the study of this total genetic material of an organism or species is called genomics.The study of genomics involves the use of various techniques of creating recombinant DNA and DNA sequencing in order to elucidate structure and function of genomes.

DNA libraries are collection of cloned DNA fragments



Source:https://www.youtube.com/watch?v=SvjeCxVu2dI

DNA cloning is a technique to construct recombinant DNA molecules. Desired DNA fragmentis inserted with a vector molecule which serves as a means to propagate the cloned DNA within the cell. DNA library is a large collection of all the sequences from a specific cell/ tissue/ organism cloned into a suitable vector. Just like a book library is a collection of books where specific books for varioussubjects can be searched, researchers

use a DNA library to isolate specific genes. DNA libraries are important source of DNA for genome sequencing, gene identificationand functional analysis.Broadly two different types of DNA libraries can be constructed depending upon the source of starting DNA:

(A) Genomic library

(B) cDNA library

When the source of starting DNA is complete genomic DNA of a particular organism, the DNA library is called a genomic library. It therefore representsat least one copy of every gene in an organism's genome. Theoretically,the genomic DNA is digested with restriction enzymes and all the generated fragments are cloned into suitable vectors and these vectors are transferred to host cells by transformation. These transformed cells represent genomic library,i.e. collection of atleast one copy of all the sequences present in the genome.

This type of library can be used for many purposes:

- To generate DNA for sequencing a genome,
- To investigate the structure of a given chromosome,
- To study the untranslated regions of a gene including promoter or introns and
- To clone smaller genes.

The DNA sequences in this type of library contain coding and non-coding sequences; therefore these genomic libraries are more suited to represent the genetic constitution of simpler organisms like bacteria or yeast. The genes of higher eukaryotes contain intron sequences (usually larger than exons), which makes genes of higher eukaryotes too large to be inserted intact into vectors. Consequently, the complete DNA sequence of a given gene is represented in more than one clone. Many a times, DNA inserts in cloning vectors do not contain sequence of coding region of the DNA but contain non coding region of the gene,as a result it becomes difficult to identify coding regions of a gene that encodes protein sequences.

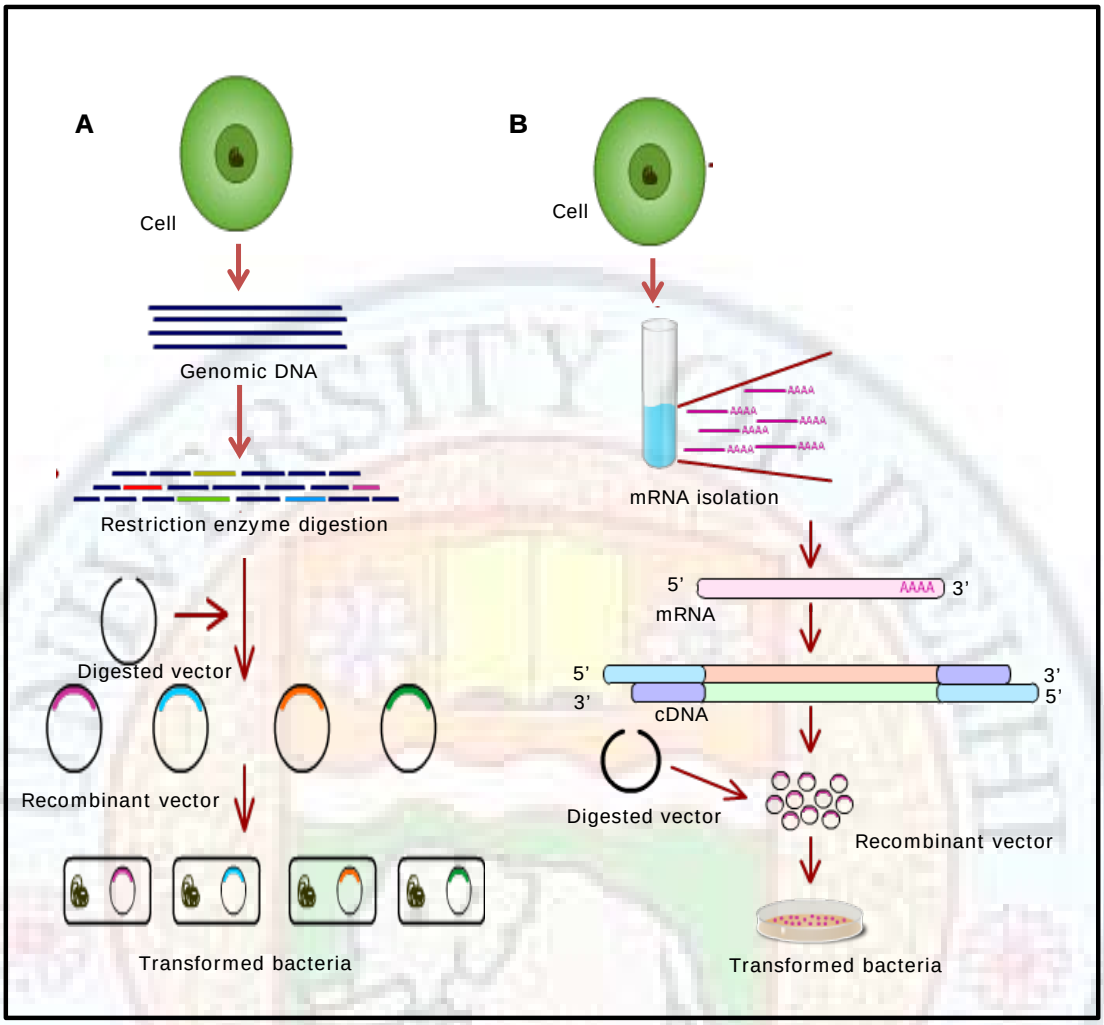


Figure: An overview of DNA libraries (A) A genomic library (B) A cDNA library
Source: Author

To clone only the coding sequences in the library,cDNA library, is constructed. For this type of library, the starting material is complementary DNA (cDNA is mRNA converted to complementary DNA sequence). This large collection of cDNA clones representsonly those genes that are expressed (transcribed into mRNA) in a cell type. A cDNA library is small and includes only the genes that were active/expressed in the cells used as a starting material for construction of the library. The search for a specific gene is made easier by using a cDNA library synthesized from the mRNAs of a cell known to express that gene. cDNA libraries are also useful to identify the genes that are active in one cell type and not another, and therefore reveal genes responsible for cell differentiation and specialization. For example, when a plant is subjected to any biotic or abiotic stress, expression of mRNA specific to that

stress takes place in that cells. These stress specific genes may be absent in the cells which has not been exposed to stress.

This type of library is also helpful to study the relative gene expression in normal and diseased cells. Another advantage of cDNA library is that cDNA clones can be used to express eukaryotic genes in bacteria. The genomic sequences typically contain introns as well as other non-coding sequences and sinceprokaryotes do not have introns, they do not possess enzymatic machinery to process eukaryotic genes correctly. However, cDNAs do not have introns and therefore bacteria can transcribe and translate them to produce eukaryotic proteins.

Construction of Genomic library

A genomic library is synthesized from randomly generated DNA fragments from complete genome of a particular organism.

- The first step in construction of genomic DNA library is isolating genomic DNA from the organism.
- Genomic DNA is then partially cleaved into thousands of fragments (of 5-100 kb) by restriction endonucleases to get inserts of desired size range, compatible with the cloning vector (such as plasmid, phage lambda, cosmid, bacteriophage P1,Bacterial Artificial Chromosome [BAC] or Yeast Artificial Chromosome [YAC]) used for library construction. Genomic libraries are usually constructed in vectors which can accommodate insert of large size for eg. lambda phage, BAC orYAC.

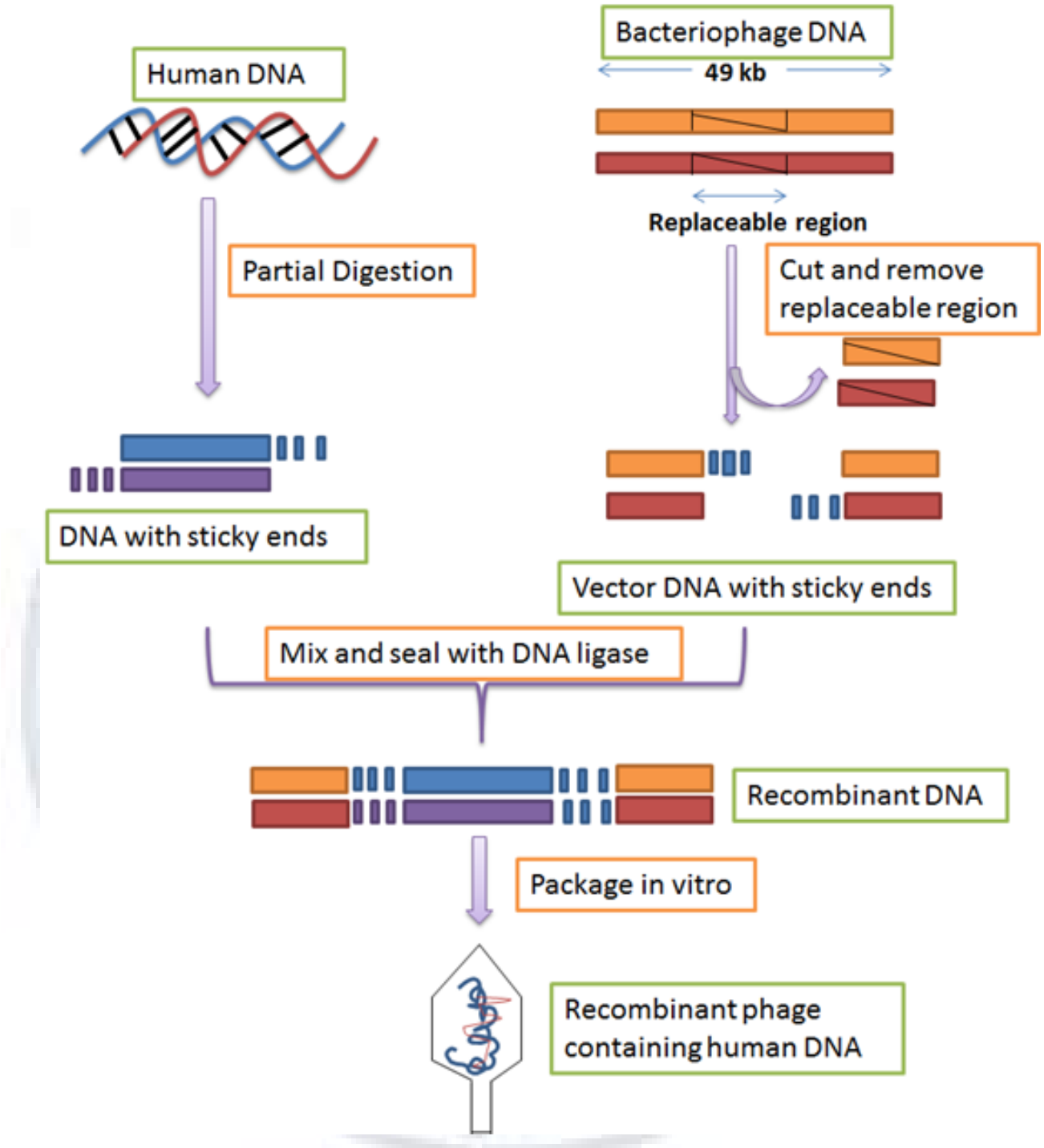


Figure: Generation of a genomic library

Source: http://nptel.ac.in/courses/102103013/20

- All the fragments are ligated into thecloning vector which is cleaved with the same restriction endonuclease and transformed into host organism which is commonly a

population of *Escherichia coli* or *yeast*cells to produce a library with each cellcontaining one vector molecule.

Each cloning vector contains a different fragment of the genome so that all DNA in the genome is represented among the clones in the library. All of the plasmids in a particular host cellharbor the same insert. As the plasmid multiplies, there will be 50-100 copies of the plasmid in each cell. In addition, each transformed bacterium or yeast cellbearing the same cloning vectorwill multiply so there may be over a million cells in one colony on a petri plate. Genomic DNA inserts can be sequenced by isolating individual recombinant cloning vectors from cells. Sequence reads can be aligned using computer programs to compare the sequences of genomic DNA inserts with each other. If the two aligned sequences overlap with each other they represent a contiguous segment of a genome called a contig, which incorporates both sequences. A genomic library contains thousands of long contigs, which can be characterized and assigned to particular chromosomes to form a detailed physical map. The earlier known and characterized sequences within the library can provide clues for genome sequencing projects. By sequencing and analyzing random clones from the library contigs can be made and the wholegenome sequence can be recovered.

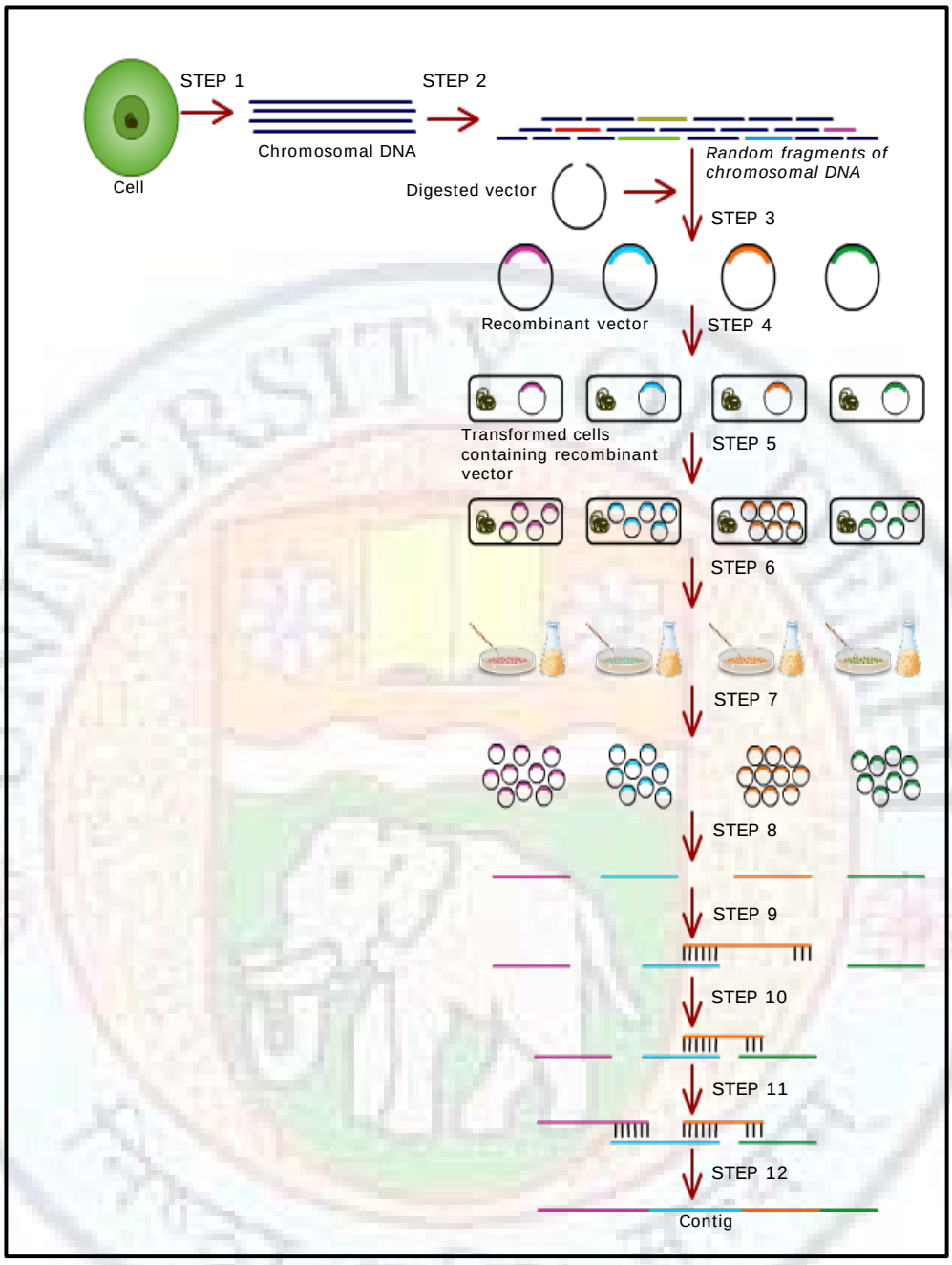


Figure. Steps in the construction of a genomic library. Step 1. Genomic DNA is isolated. Step 2. Genomic DNA is digested with restriction enzymes. Step 3. Digested vector and insert DNA are ligated. Step 4. Recombinant vector is transformed into a host organism. Each vector contains a different fragment of genome. Step 5. Each plasmid multiplies in the host cell so that there are 50-100 copies of the plasmid in each cell.Step 6. Bacteria that have taken up plasmid are selected. Each cell bearing cloning vector multiplies. Each bacterial cell contains millions of vectors, all derived from transformed cell.Step 7.

Recombinant cloning vectors are isolated. Step 8. Insert DNA is sequenced.Step 9, 10 and 11. Sequence reads are aligned using computer programs to compare the sequences of genomic DNA inserts with each other. Step 12. Contiguous overlapping segments of a genome called contigsaregenerated.

Source: Author

How good is the constructed library

The number of clones required to include all sequences in the genome depends on the size of the genome being investigated and the average size of the insert DNA fragments inserted into the vector. The probability of finding at least one copy of any DNA sequence in the genomic library can be calculated from the following formula:

$$N = \ln (1-P) / \ln (1-f)$$

Where,
N is the necessary number of clones,
P is the probability of finding any particular sequence in the library,
*f*is the average size of cloned fragments in kilobase pairs divided by the size of the genomein kilobase pairs,
ln is natural logarithm.

This formula can alsobe used to determine the fraction of the genome present in a newly synthesized library, as the number of clones *N*, and average insert size can be calculated after a library is constructed, and the size of the genome is a known value.

For example, the table below shows the number of clones required for different genomes, if the subsequent fragments are cloned in cosmid vectors, calculated for a probability of 95%.

Table: Number of clones required for genomic library construction for different organisms.

Species	Genome Size (bp)	Number of clones if fragment size is 35 kb (for cloning in cosmid vectors)
E. Coli	4.6×10^6	410
Saccharomyces cerevisae	1.8×10^7	1500
Drosophila melanogaster	1.2×10^8	10,000
Rice	5.7×10^8	49,000
Human	3.2×10^9	274,000

Source: Author

Construction of a cDNA library

Researchers commonly use second type of DNA library called cDNA librarywhich is constructed with only the DNA that is transcribed as mRNA, and later translated to produce proteins. The first step in synthesis of a cDNA library is to isolate total RNA from thecell/tissue of interest. mRNA is present in small percentage in a mixture of total RNA which contains predominantlyrRNA and tRNA.Since mRNAs have a stretch of several adenines attached at their 3' end called poly A tails, they can be easily isolated by use of a columnwhich has oligodTs (thymidylate) linked to its matrix.

Construction of genomic and cDNA libraries

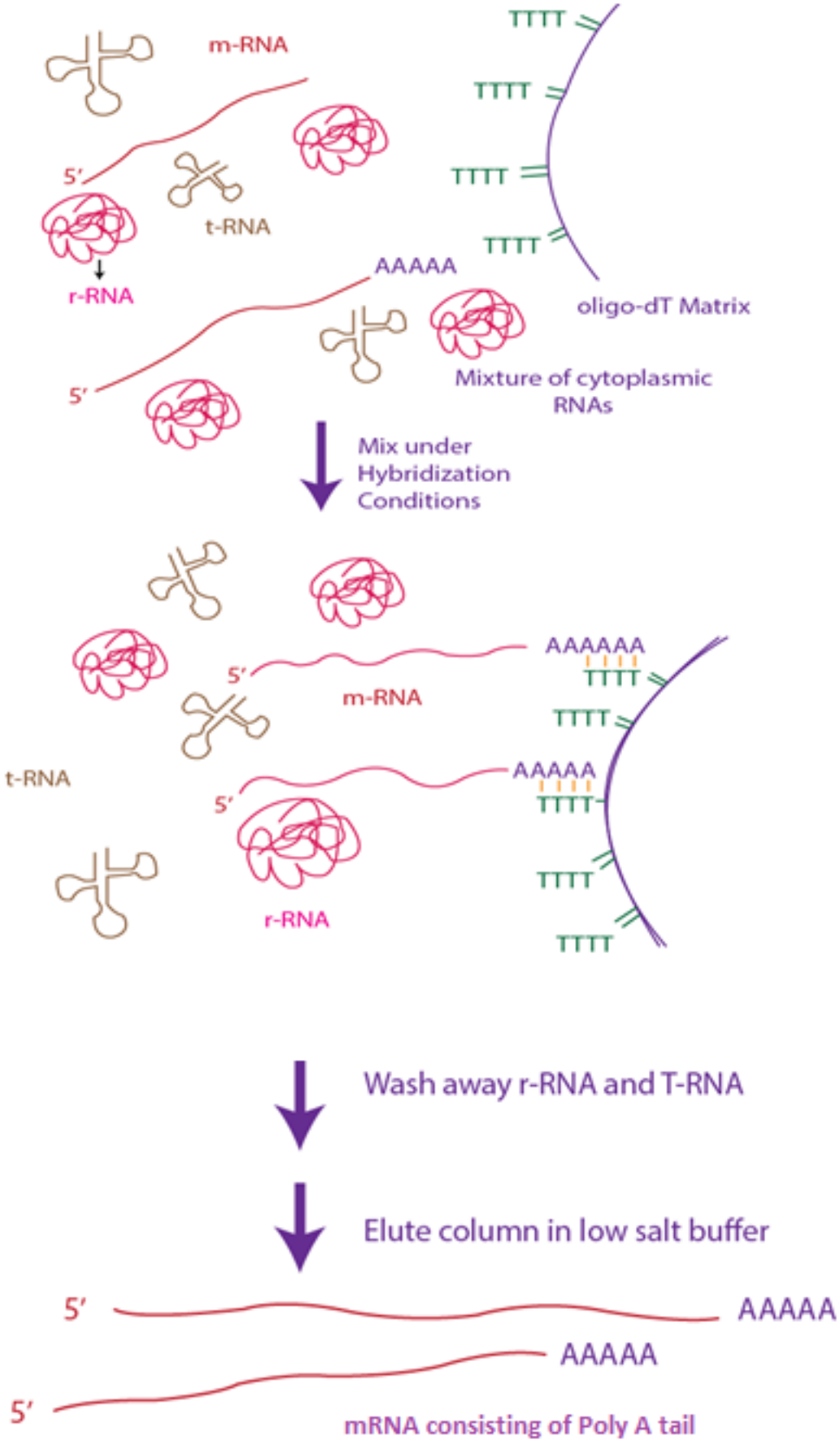


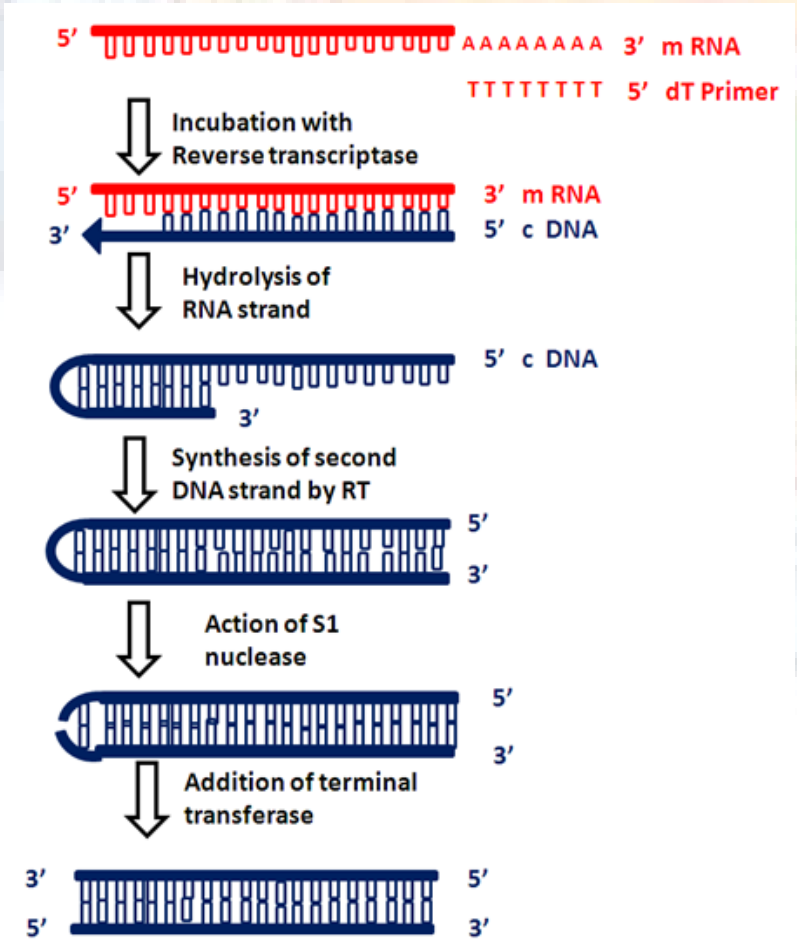
Figure: Isolation of mRNA using oligo-dT column chromatography

Source: http://nptel.ac.in/courses/102103013/19

mRNA cannot be cloned and is therefore converted to double stranded DNA. Enzyme reverse transcriptase, an RNA dependent DNA polymerase (found in retroviruses, AMV or MMLV) synthesizes a DNA copy complementary to each mRNA. This enzyme, like all DNA polymerases, can extend from a preexisting primer only. Therefore starting from oligodT primer, reverse transcriptase initiates synthesis of the cDNA strand. This is called as first strand synthesis resulting into mRNA cDNA hybrid.

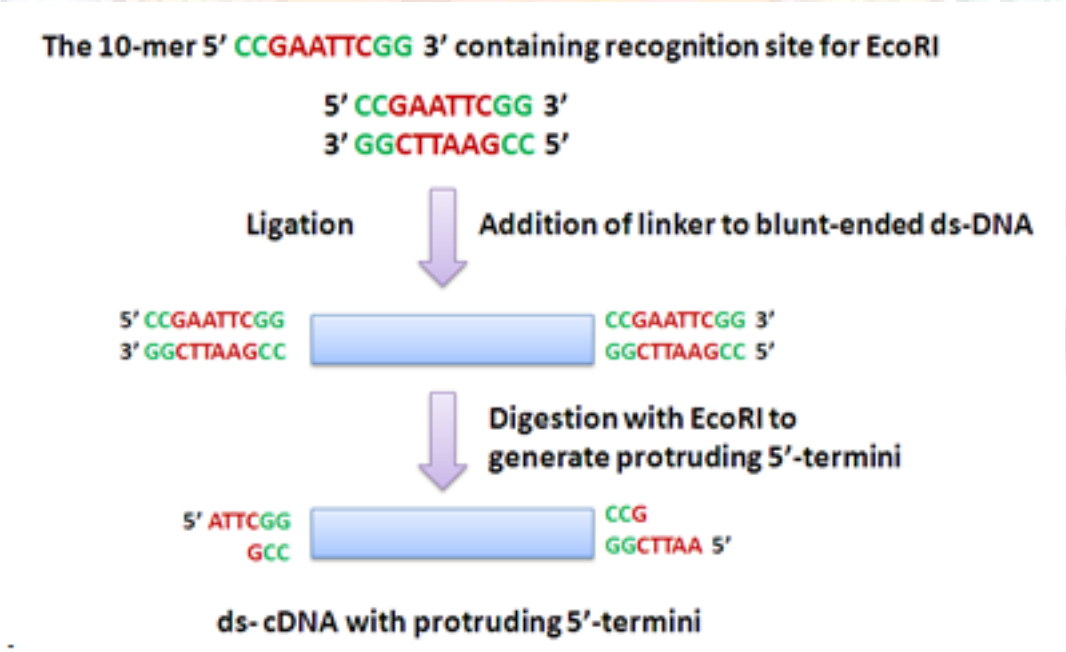
There are two methods for generating double stranded cDNA:

(1) A traditional method where in the 3' end of the newly formed cDNAloops back to form a hairpin loop structure and second strands were synthesized by the DNA polymeraseactivity followed by S1 nucleasedigestion to remove the hairpin structures.



Source:http://nptel.ac.in/courses/102103013/module4/lec3/images/3.png

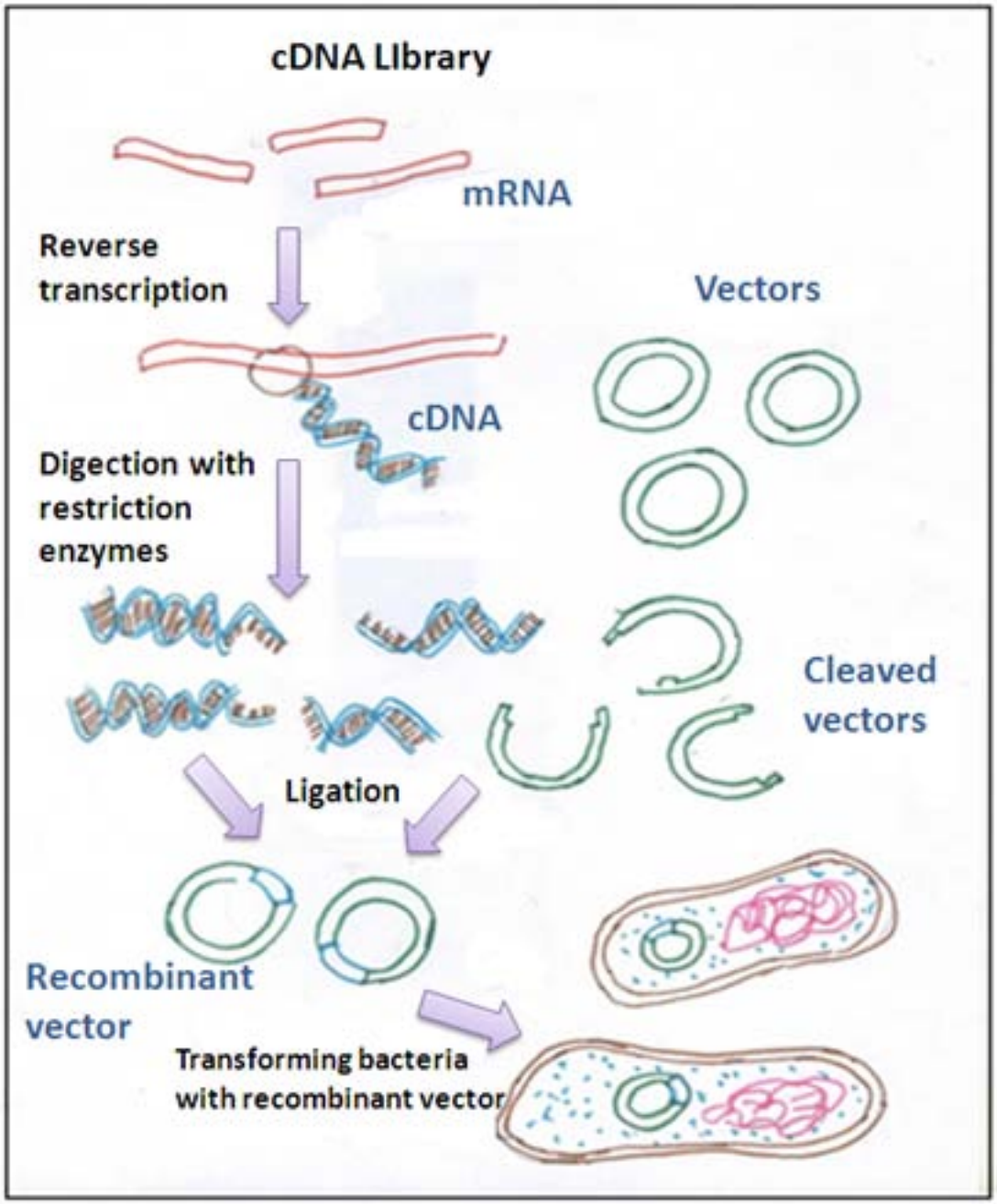
(2) Now a days, a second strategy is used in which themRNA strand of the mRNA-cDNA hybrid ischemically degraded by hydrolysis treatment or enzymatically digested by using an enzyme *RNase H*.DNA polymerase is used to make a second DNA strand complementary to the first.To do this several bases of C residue are added at the end of cDNA molecule through the action of terminal transferase enzymeDNA polymerase activity also requires a primer to initiate synthesis of the complementary strand. A short single-stranded primer which anneals to the short stretch of C's at the 3' end of the single-stranded DNA moleculeis used. This is known as second strand synthesis. Newly synthesized double stranded cDNA molecule contains a short dT-dA region at one end and dC-dG region at the other end. These double stranded DNA molecules are prepared for cloning by ligation of short linker which is oligonucleotideprimer of double stranded DNA containing recognition sequence for a particular restriction enzyme to both its ends.



Source:http://nptel.ac.in/courses/102103013/19

The importance of attaching a linker containing recognition sequence for a particular restriction enzyme is that a researcher doesn't know which restriction enzyme could be used to cut the insert DNA. These double stranded cDNA products and vector are digested with the same restriction enzyme, specific for the attached oligonucleotide

primer. The collection of cDNA with sticky ends and vector are joined covalently by ligase. Resulting recombinant molecules are transformed into *E.coli* cells to generate individual clones.



Source:http://nptel.ac.in/courses/102103013/19

The end result is a cDNA library consisting of a population of many different clones, each clone carrying a specific cDNA derived from a single mRNA. cDNAs are complementary to the mRNA that was transcribed in cells used as starting point for construction of the library. cDNAs which are frequently present in the library

represent the genes which were transcribed abundantly and cDNAs which were rare in the library represent the genes which were transcribed at low rate in the cell type being studied.

Construction of genomic and cDNA libraries

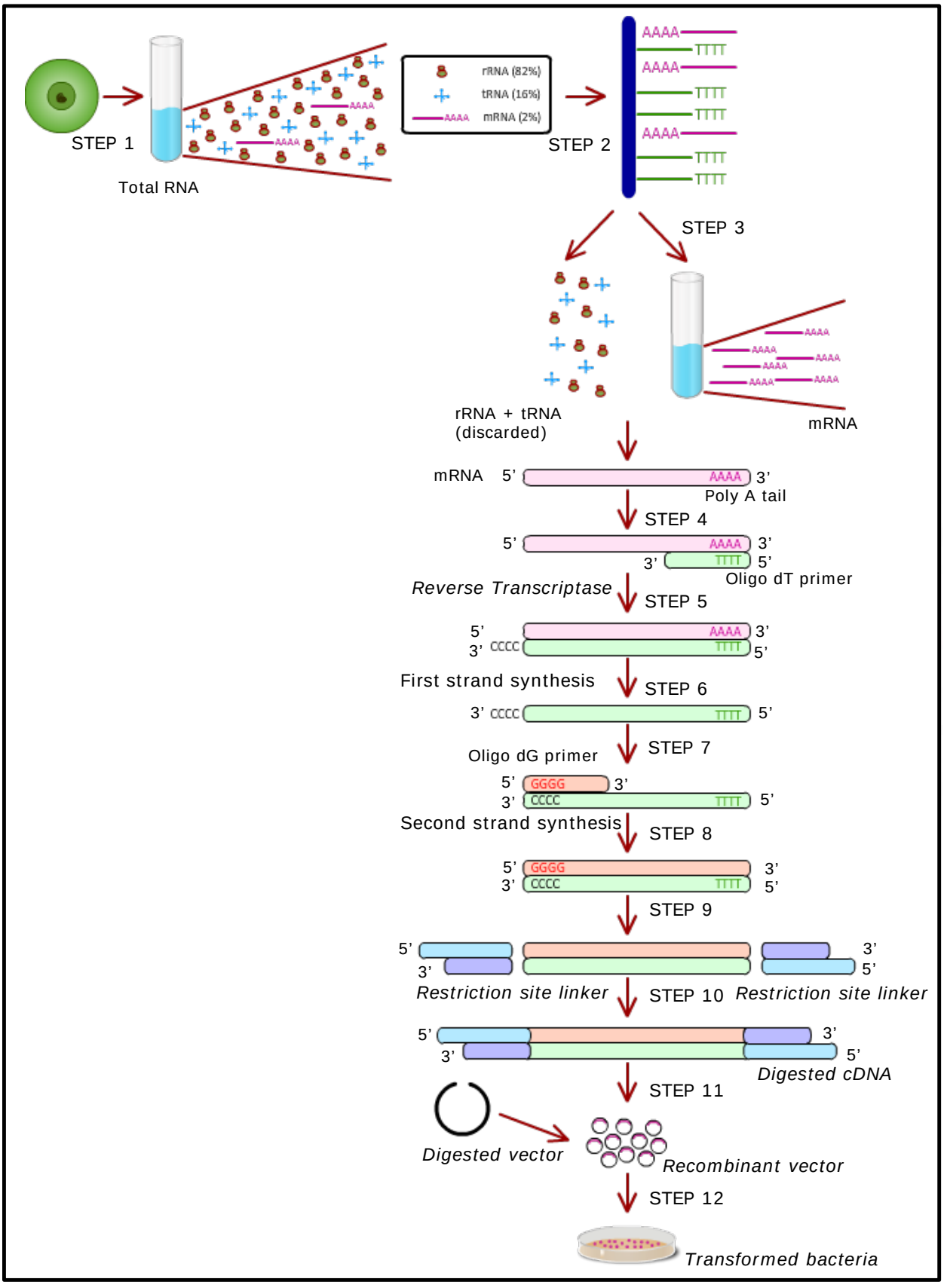


Figure: Steps in the construction of a cDNA library. Step 1. Isolate total RNA from the cell/tissue of interest.Step 2. mRNAs having poly A tails are isolated by use of a column which has oligodTs linked to its matrix.Step 4. Hybridize mRNA with oligodT primer. Step 5. Enzyme reverse transcriptase initiates synthesis of the first cDNA strand. Several bases of C are added to the end of cDNA molecule.Step 6. mRNA strand is degraded.Step 7. A single-stranded oligonucleotide dG primer anneals to the short stretch of C's at the 3' end of the single-stranded DNA molecule.Step 8. Second strand DNA molecule is synthesized, containing a short dT-dA region at one end and dC-dG region at the other end.Step 9. A short linker which is oligonucleotide primer of double stranded DNA containing recognition sequence for a particular restriction enzyme is ligated to both the ends of cDNA.Step 10. Double stranded cDNA products are digested with the restriction enzyme.Step 11. Digested vector molecules are ligated with the cDNA.Step 12. Resulting recombinant molecules are transformed into *E.coli* cells to generate individual clones. The end result is a cDNA library consisting of a population of many different clones, each clone carrying a specific cDNA derived from a single mRNA. Also show another figure for hair loop structure

Source: Author

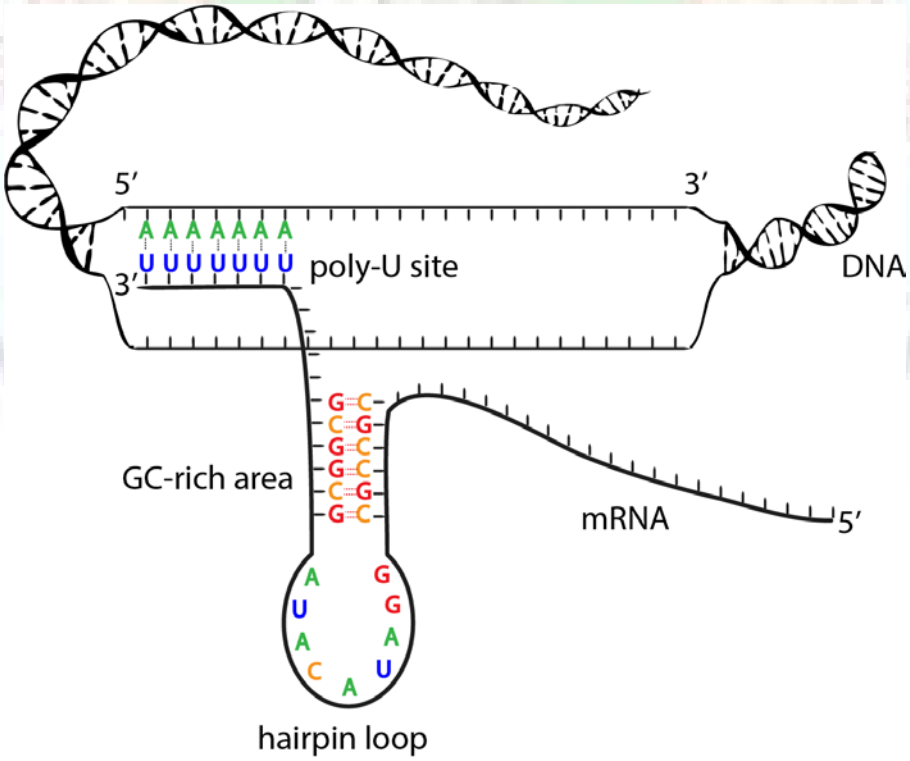


Figure: Diagram showing hair pin loop formation in mRNA

Source:https://adapaproject.org/bbk/tikiindex.php?page=Leaf%3A+How+do+cells+terminate+the+process+of+transcription%3F (CC)

Table: Comparing the genomic and cDNA library

Source:http://nptel.ac.in/courses/102103013/20

Feature	**Genomic library**	**cDNA library**
Sequences present	Ideally, all genomic sequences	Only structural genes that are transcribed
Contents affected by : (a) Developmental stage (b) Cell type	 No No	 Yes Yes
Features of DNA insert(s) representing a gene: (a) Size (b) Introns (c) 5´- and 3´- regulatory sequences	 As present in genome Present Present	 Ordinarily, much smaller Absent Absent
As compared to the genome (a) Enrichment of sequences (b) Redundancy in frequency (c) Variant forms of a gene	 In amplified genomic libraries In amplified libraries Not possible	 For abundant mRNAs For rare mRNA species For such genes, whose RNA transcripts are alternatively spliced

Genetic Selection of recombinants from the library

Genomic and cDNA libraries contain millions of individual clones, which are not distinguished from each other. Libraries can be searched by various methods to isolate the genes ofinterest. Genetic selection is required to identify recombinants, as large population of cells can be screened, differentiating between recombinant and non-recombinant vector. These methods rely on exploitation of certain genetic characteristics of the system, the traits usually encoded by the vector. All cloning vectors carry a selectable genetic marker which confers selectable growth advantage to the cell bearing that vector. Genetic selection methods therefore employ techniques to identify recombinant host cells in a population of

cells by applying selection pressure. Genetic selection methods can be simple employing antibiotic selection pressure on vector cells. A more sophisticated selection method utilizes chromogenic substrate X-gal for color difference between recombinant and non-recombinant cells.

Genetic selection by antibiotic resistance

Transformation is an inefficient process and recombinant vectors are stably transferred to a small population of cells only**.** Cloning vectors carry drug resistance genes or nutritional markers, which allow rare transformants to be selected with ease.These markers confer the cells containing vector, the ability to grow in presence of antibiotics (like ampicillin, kanamycin, tetracycline etc). Bacteria carrying parental plasmid are unable to form colonies under selective conditions. For example, if the cloning vector carries a kanamycin resistance gene (*kan*^{r}), it will confer resistance to the *E. coli* cells against kanamycin if the transformation process was successful. Therefore when the library clones are plated out on medium containing kanamycin, only cells containing recombinant plasmid (with an insert) would be resistant to the antibiotic and would be able to grow. One pit-fall of this method is that it doesn't differentiate between re-ligated vector (without the inserted gene of interest) and recombinant vector (with the inserted gene of interest) as in both cases drug resistance genes will confer the cells containing vector, the ability to grow in presence of antibiotics.

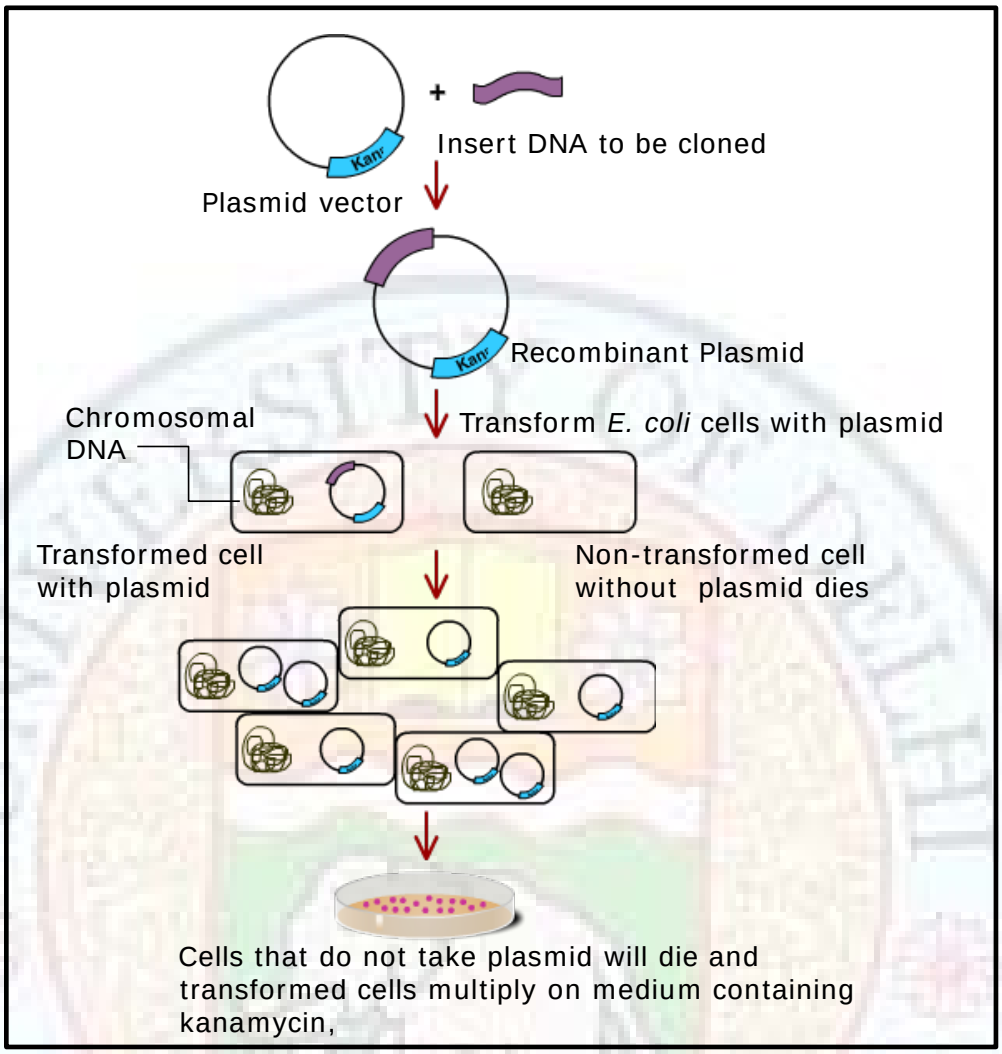


Figure: Genetic selection method of selecting recombinants by direct antibiotic resistance.

Source: Author

Genetic selection by blue white screening

Many general purpose vectors use a non-destructive histochemical procedure called blue white screeningto detect β galactosidase activity in transformed bacteria.These vectors carry coding and regulatory sequences necessary for production of*E. coliβ-galactosidase*protein which is encoded by *lacZ* geneunder the control of the *lac* promoter.Iflac repressor is expressed in the host *E. coli* strain, the expression of a *lacZ* gene on the vector could be induced using lactose analog IPTG (isopropyl- ß-D-

thiogalactopyranoside), resulting in formation of β-galactosidase. Its presence can be detected by use of synthetic soluble, colourlesssubstrate X-gal(5-bromo-4-chloro-3-indolyl-beta-d-galactoside), which is hydrolyzed to yield an insoluble, blue colored product (5,5'-dibromo-4,4'-dichloro indigo).The basis of the blue white selection is called α-complementation, wherein the *E. coli* strain carries the lacZ deletion mutant, the ω-peptide (N-terminal residues 11—41 are deleted) while the plasmids used carry the lacZα sequence which encodes the first 59 residues of β-galactosidase, the α-peptide. When the two peptides are expressed together, they form a functional β-galactosidase enzyme.However,neitherthe ω-peptide nor the α-peptide produce functional β-galactosidaseenzymeby itself. The *lacZα* sequence are designed to have multiple cloning site which are furtherdigestedby restriction enzymes to introduce the insert DNA into the vector during cloning procedure, thereby disrupting the gene and production of α-peptide. As a result, in cells containing the plasmid with an insert,functional β-galactosidasecannot be formed where as in cells without insert,β-galactosidasecan be formed. Cells harboring functional β-galactosidase are blue in color therefore blue colonies may contain a re-ligated vector without an insert (with an uninterrupted *lacZα).* White coloniesdon't express β-galactosidase and indicate the presence of an insert in *lacZα*and are likely to contain recombinant vector containing insert.

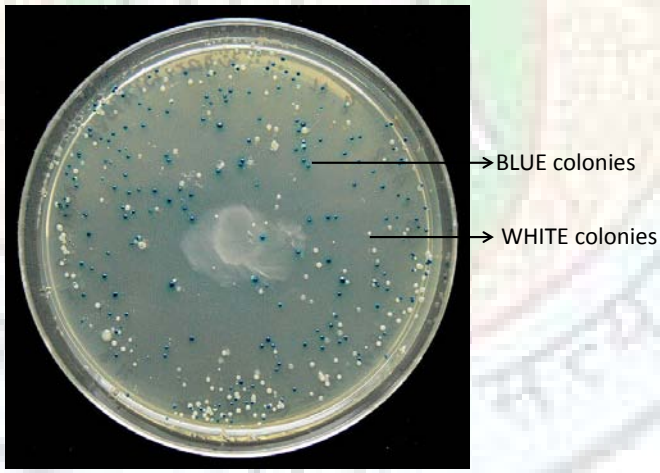


Figure: A petri plate showing blue and white colonies

Source:http://en.wikipedia.org/wiki/Cloning_vector#mediaviewer/File:Blue-white_test.jpg

Genetic selection by plaque morphology

Phage vectors replicate and spread to generate regions of cell destruction known as plaques. Plaque formation is itself a selection method in case of phages, which contain the *cI*gene. This genecodes for the cI repressor, which is responsible for the production of lysogens. If the *cI*gene is inactivated by cloning an insert into a restriction site within the cIgene, the plaques are clear and can be differentiated from the turbid non-recombinants plaques.

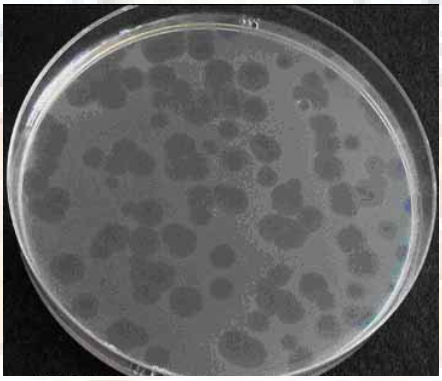


Figure: Plaques formed by bacteriophage upon infection of bacteria

Source:http://ocw.mit.edu/courses/biological-engineering/20-109-laboratory-fundamentals-in-biological-engineering-fall-2007/labs/mod1_2/

Screening DNA libraries to obtain gene of interest

Library screening relies on information available about vector and genes of interest. In this, clones of the library are subjected to analysis to identify desired sequences. Differentmethods areusedto identify clones harboring fragments of genes of interest in the library. DNA sequences can be screenedfor the gene sequence of interest among the clones of library with oligonucleotide probes that match to the sequence of the clone of interest.Functional screening for the encoded protein products by the expression of the gene of interest can also be performed to identify it.

Colony Hybridization

The method employing labelled oligonucleotideDNA probe to screen a library is known as colony hybridization.It was applied for the first time by M.Grunstein and D. S.Hogness in 1975. The basis of hybridization is complementarity between single stranded DNA or RNA molecules to associate with each other. In this method single stranded fragment of labelled DNA or RNA known as oligonucleotide probeis used to detect the gene of interest on the basis of homology. A library contains large numbers of clones, each containing a common vector backbone with unique insert. The oligonucleotide probe is used to label the colonies of cells containing the region of gene whose sequence matches with the probe. Firstly, recombinant vectors are transformed into suitable bacterial host cells. The cells are spread on nutrient agar plates containing solid growth medium. Each cell multiplies and is converted to a colony of cells, with each colony containing same cells with similar vector and insert from the library. Pieces of nitrocellulose membrane are pressed on top of petri plates containing colonies of cells to produce a replica of the plate. This ensures that nitrocellulose filter retains some cells from each colony in the same position as they were on the petri plate. This is important because if the desired clone is discovered, the colony of cells containing that specific clone can be identified and insert DNA can be recovered. The petri plate acts as a reference for the nitrocellulose membrane.

The cells on the nitrocellulose membrane are treated with chemicals which open up cells and release the DNA on the filter. DNA is then denatured to make it single-stranded. Nitrocellulose filters are incubated with labelled oligonucleotide probe, whose sequence is complementary to the gene of interest, under conditions favoring hybridization. The hybridization step is performed at non-stringent conditions ensuring that the probe will bind to any clone harboring complementary sequence. These non-stringent hybridization conditions allow non-specific binding of probe to clones which have even little complementary to the probe.

Hybridization is followed by washing step, which is performed under stringent condition. This ensures that any excess probe that has not hybridized or bound non-specifically is washed away.

Membrane is then assayed and bound probes can be detected. If the probe has been labelled using radioactive isotopes, X-Ray film is overlaid on the membrane. Radioactive probe in DNA-DNA hybrid molecule bound to the membrane will expose the X-Ray film, producing black spots on the film. Membrane maintains the same pattern of colonies as it

was there on the petri-plate, therefore black spots on the film can provide details about corresponding colonies on the petri-plate. This technique of colony hybridization can be employed to screen both genomic and cDNA libraries without the expression of inserted sequence. Large number of recombinant clones can be screened by this method to determine the recombinant plasmid containing the sequence of interest. At the same time, this method is time consuming and can be followed for microbes only, which can withstand plating process.

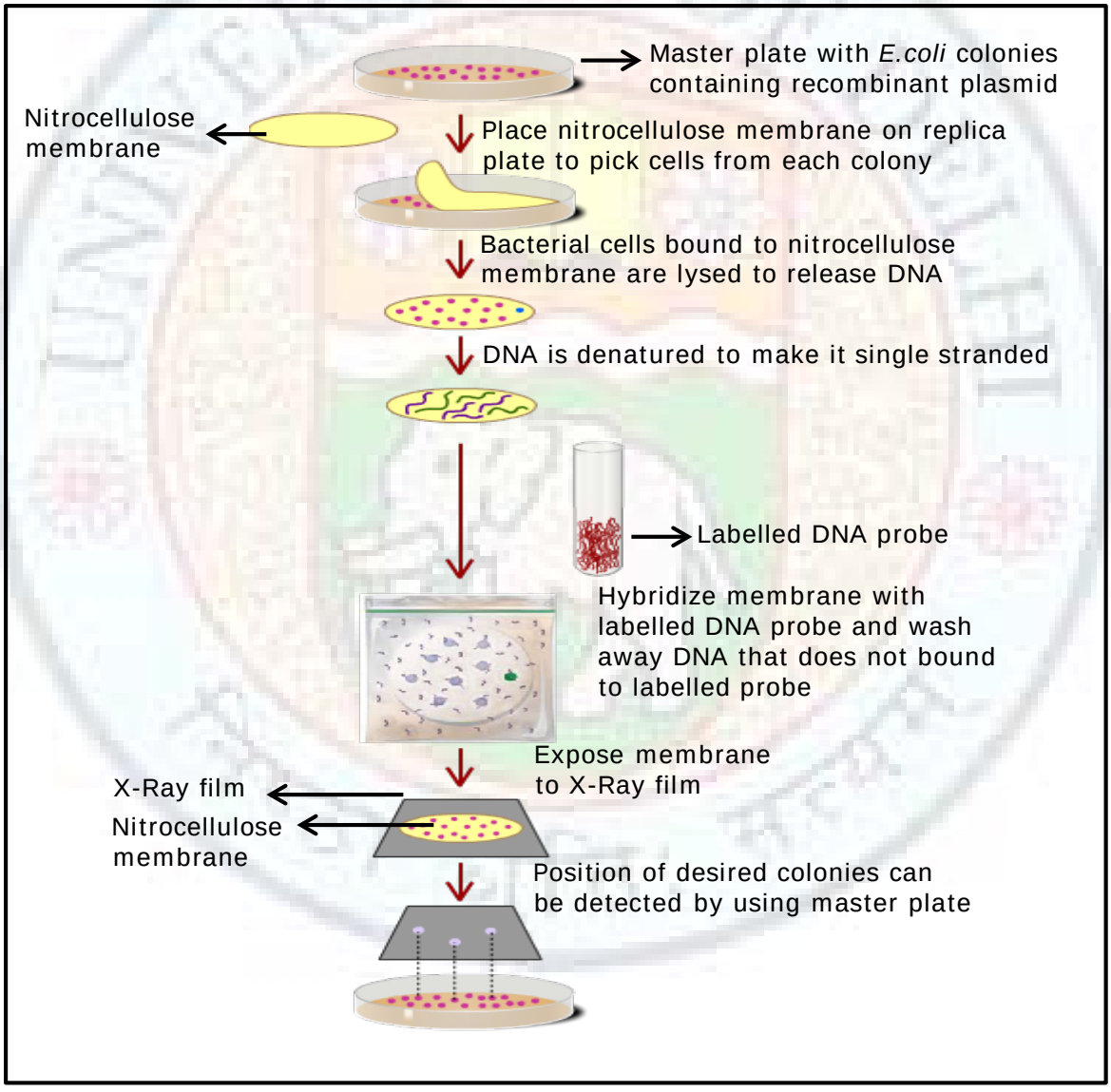


Figure: DNA libraries can be screened by colony hybridization with a radiolabeled probe to detect a clone of interest

Source: Author

Nowadays, with the concern over the usage of some non-radioactive materials which undergo chemical or color reactions are used to label the probe. For example, digoxigenin is a non – radioactive antibody based probe and labelling withdeoxyuridine triphosphate (dUTP)nucleotides and labelling with the enzyme horseradish peroxidase are becoming increasingly popular.

Polymerase Chain Reaction

Colony hybridization is a laborious and time-consuming process and requires several rounds of plating and filter hybridization. It is also prone to false positives results. Polymerase chain reaction(PCR) has emerged as a robust technique in the area of molecular biology and is also efficient method for library screening. PCR allows isolation of DNA fragments from genomic DNA by selective amplification of a specific region of DNA. Although it is tedious to screen thousands of clones by but becauseof the specificity and sensitivity of PCR (if primer pair for gene of interest is availablei.e. the sequence of the desirable fragment is known), it can be a useful way to identify clones of interest. Individual clones of the library can be subjected to PCR analysis.

Typically the PCR procedure consists of a three-step cycle at discrete temperatures to bring about achain reaction that produces an amplified product of identical DNA molecules. First regular cycling step is denaturation where the reaction mixture is heated at 94–98 °C for 20–30 seconds to separatethetwoDNA strands. Second step is annealing wherein the reaction temperature is lowered to 50–65 °C for 20–40 seconds so that short single-stranded DNA primers complementary to sequences on opposite strands anneals at each end of the target sequence. Lastly, a heat-stable DNA polymerase extends the primers in the 5' ---> 3' direction and synthesizes a new DNA strand complementary to the DNA of the template strand. The PCR technique is explained elaborately in the chapter Polymerase chain reaction.

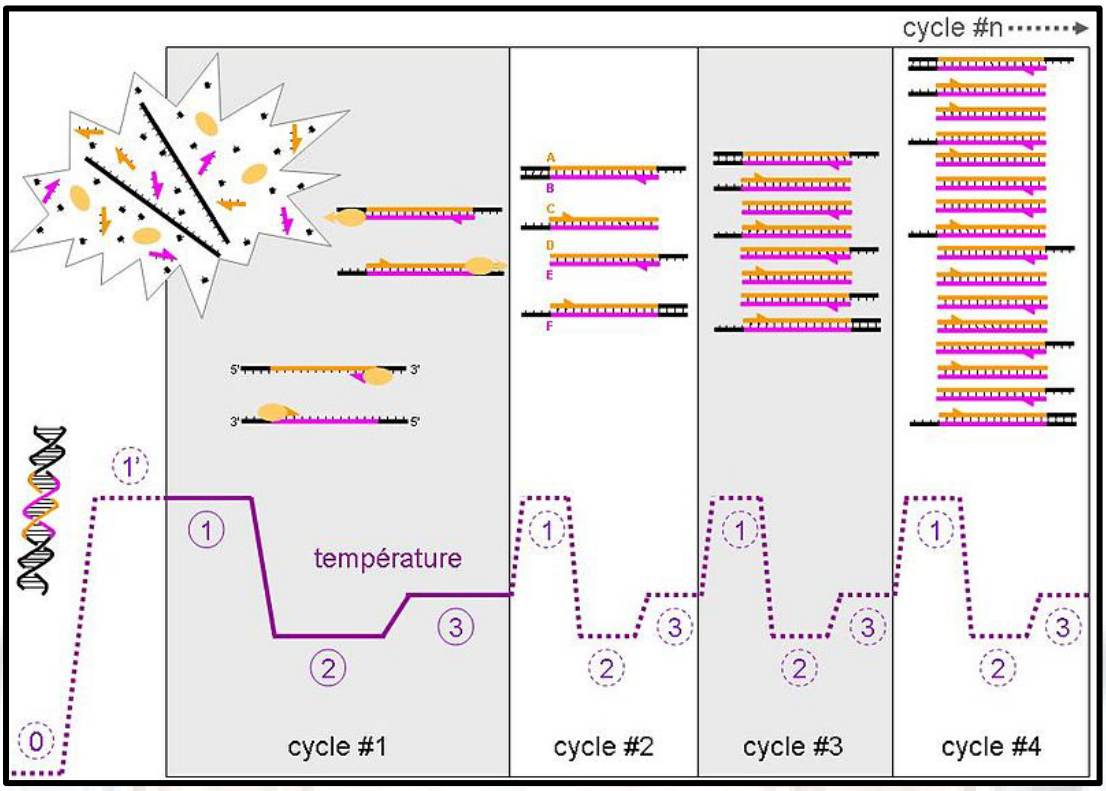


Figure: Schematic representation of the basic principle of PCR
Source:http://commons.wikimedia.org/wiki/File:PCR_basic_principle1.jpg

Following video link shows how to set up a PCR reaction for screening a library to isolate the gene of interest.

http://www.youtube.com/watch?v=q6MHnuznzpk

For library screening, multiple clones canbe pooled and a PCR reaction can beset on the pooled samples. This is known as combinatorial screening. The high sensitivity and specificity of PCR allows the detection of target sequences from pooled samples as well. Samples from clones in the rows and columns of the plate are pooled respectively and a PCR is performed on the mixed samples. If a clone is present in a sample from row pool, it must be in one of the wells in that row. Likewise, if a clone is present in a sample from column pool, it must be the clone in that column. If a positive clone is obtained, the clone is identified by PCR of independent samples of the row and column.

Immunological Screening

The desired gene can also be identified by the activities of the protein product of the encoded gene. This type of strategy is used for cDNA expression libraries. The fragments are cloned into special cloning vectors which allow the functional expression of cloned DNA fragments. The protocol of immunological screening is similar to hybridization screening except the fact that antibodies are used as probe instead of nucleic acid. Clones of the library are spread on replica plates.Transformed bacterial colonies on petri plates are lysed to release protein (antigen) from the positive colonies. Protein samples are transferred to similar position on the nitrocellulose membrane so thatthe pattern of clones remains same on the membrane and the replica plates. Released bacterial and expressed proteins are bound to the membrane. Membrane is overlaid with primary antibody which binds to protein of interest. A secondary antibody with a detection system is then overlaid on the membrane to detect protein-primary antibody complex. Unbound antibodies are then washed out. Bound antibodies are detected as it produces a colored compound. The colonies with positive results are isolated to identify the vector and correct gene of interest. The limitation of this technique is that it can be used only when specific antibody is available.

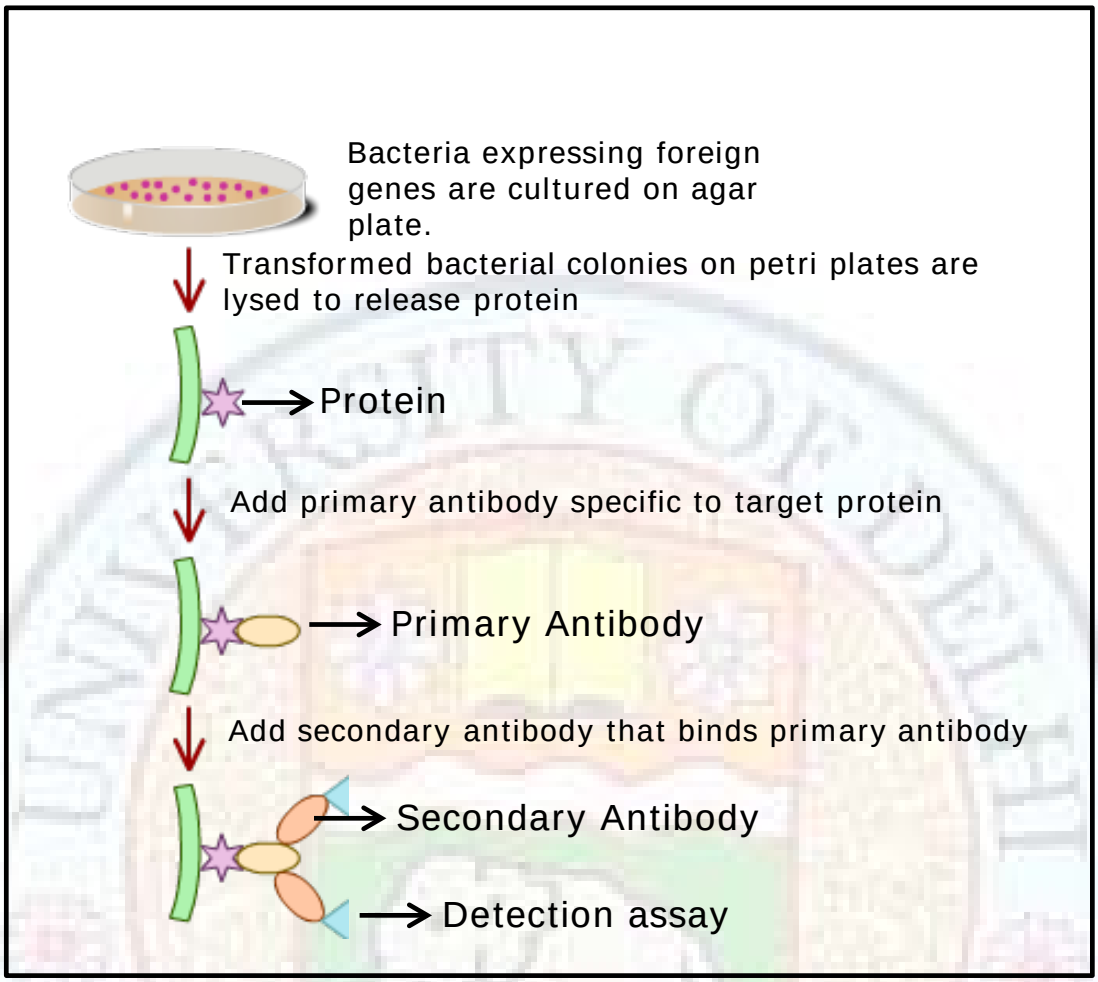


Figure:Immunological screening of a functional protein from DNA library. The figure shows one reference protein.

Source:Author

Complementation

Genomic libraries can also be screened for the ability of the cloned gene to expressa functional protein. This approach is effective if the functional protein complements a recessive mutation and is known as functional complementation. Yeast is a model eukaryotic organism, which is easy to grow and perform genetic manipulations. In this method, a yeast mutant defective in the gene of interest is chosen and a yeast genomic library is transformed in bacterial cells. Yeast mutant is complemented when expression of transgene from library is induced.

Construction of genomic and cDNA libraries

Yeast genes do not contain multiple introns and therefore entire gene sequences can be inserted into a plasmid vector. Functional complementation is tested for yeast genescloned in plasmid vector and therefore shuttle vectors (which are capable of propagation in two different hosts, yeast and *E. coli* in the present case) are used for screening of recombinant plasmids. This shuttle vector contains essential elements to allow cloning of yeast DNA fragments in *E. coli*. Additionally it contains origin for DNA replication in yeast (ARS, Autonomously replicating sequence), yeast centromere (CEN) to allow faithful segregation of plasmid during yeast cell division and a yeast gene encoding a selectable marker URA3 fororotidine-5'phosphate decarboxylase, an enzyme which is required for the synthesis of uracil. A particular yeast DNA sequence is digested to produce overlapping restriction fragments.Shuttle vector is cleaved with the same restriction enzyme to produce sticky ends complementary to the DNA sequence. Vector is transformed to *E. coli* cells, and cells that grow after selection for ampicillin resistance contain single type of yeast cDNA fragment.

A yeast genomic library can be cloned in shuttle vector and screened to identify the wild type gene corresponding to recessive and temperature sensitive *cdc28* mutations. These cell cycle mutants can grow easily at 23°C but are unable to make colonies at 36°C. The assay utilizes double mutants, which requires uracil for growth due to *ura3* mutation and is temperature sensitive due to cdc28 mutation. A genomic DNA clone that complements this mutation can be identified by transforming recombinant plasmids isolated from yeast genomic library with yeast double mutants. Two steps are important to prove that insert present on the plasmid contains wild-type *CDC28* gene. Firstly, transformed yeast cells will carry a plasmid borne selectable marker, a gene for synthesis of uracil, and can be selected by their ability to grow in absence of uracil. Transformed colonies are then allowed to grow under permissive conditions ie at 23°C. Replica plates of transformants are then transferred to non- permissive conditions ie at 36°C. Only the yeast colonies that carry dominant form of *CDC28* gene will be grow in non- permissive conditions. Plasmid DNA can be extracted from yeast cells and further analyzed by sequencing.

Construction of genomic and cDNA libraries

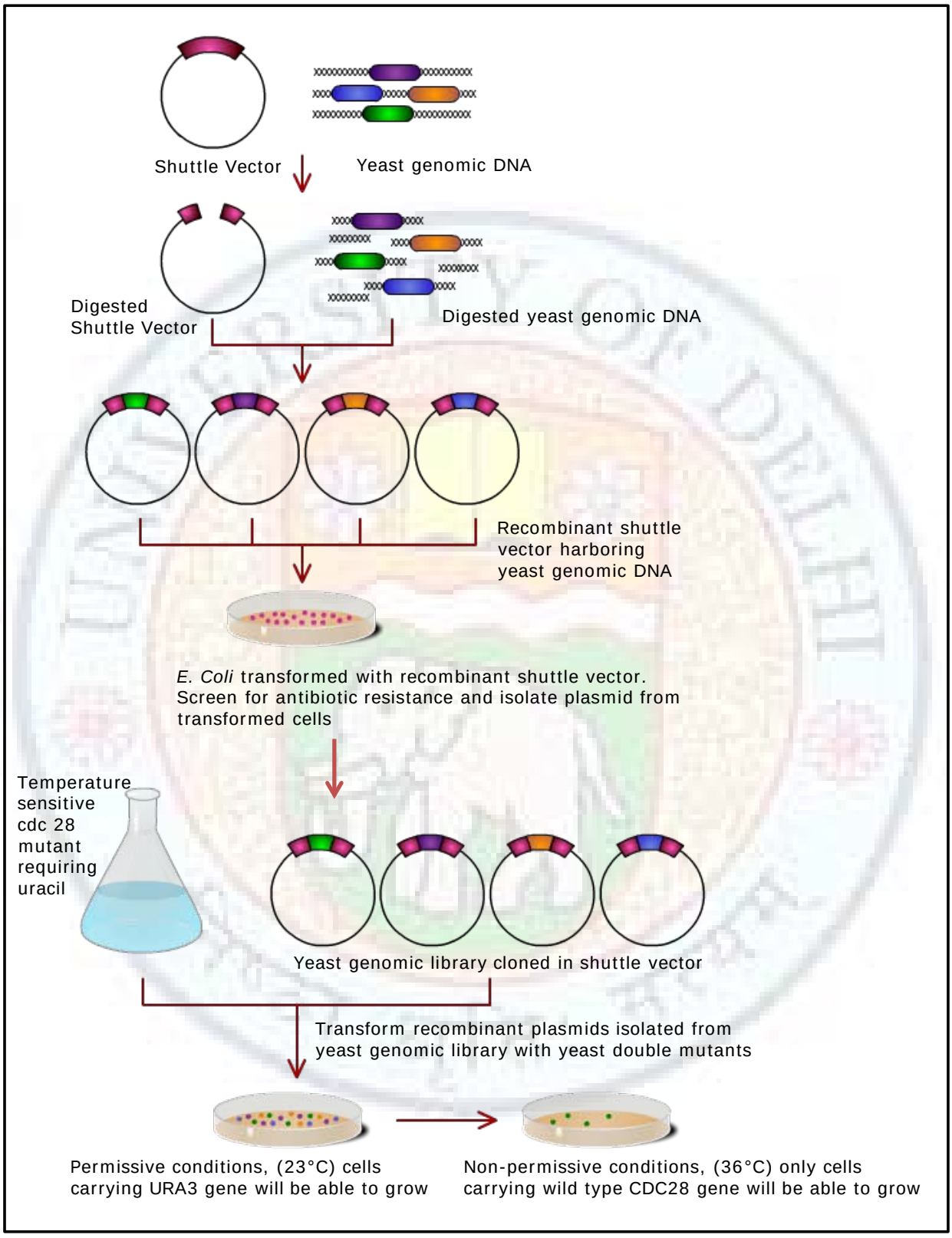
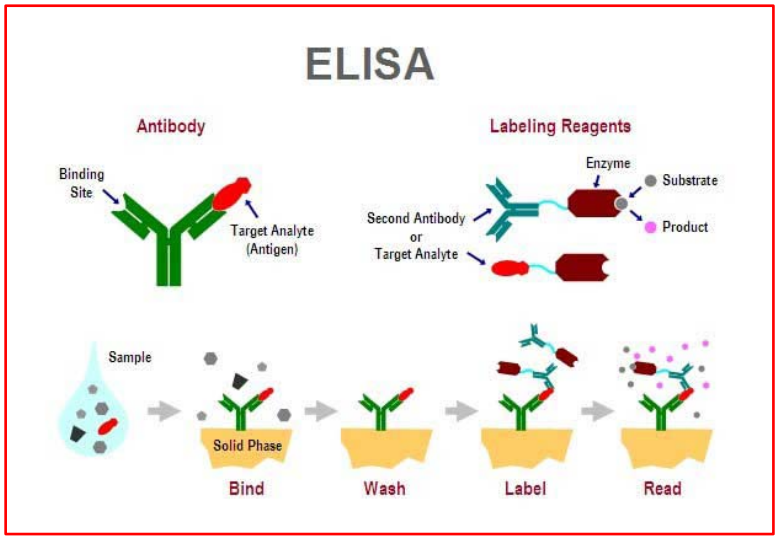


Figure: Construction of yeast genomic libraries in shuttle vector and screening by functional complementation to identify clones carrying the normal form of a mutant yeast gene.

Source: Author

An Enzyme Linked Immunosorbent Assay (ELISA)

An Enzyme Linked Immunosorbent Assay (ELISA) is a common quick and simple biochemical technique and involves specific and non-specific interactions by sequential binding to a solid surface. It detects the presence of an antibody or an antigen in the sample. The procedure of the ELISA results in a colored end product which correlates to the amount of specific substance present in the original sample. Steps of ELISAs begin with a coating step, where an antigen containing solution is adsorbed to a polystyrene 96 well plate. A small proportion of protein coats the surface of tubes or wells. Large number of samples can be run together.Fisrt step of coating is followed by washing and detection steps as shown in the schematic diagram below.The procedure uses surface binding for separation; several washing steps are performed between each ELISA step to remove any unbound substance. After the unbound antigens are washed away, the samples of antibody (known or unknown) are incubated in the antigen coated wells. Antibody which remains bound to immobilized antigen, after washing can be detected using labeled anti-immunoglobulin or immunoglobulin binding protein. These assays employ enzyme labelled detecting agent are therefore known as enzyme linked immunosorbent assays. These assays can also be quantified by using a standard solution of known antibody content.



Source: http://www.socmucimm.org/elisa-method/

Chemical synthesis of DNA Probes and DNA oligonucleotides

DNA probesare purified fragments of single-stranded DNA ranging from 25 to several thousand in length. They are indispensable tools in molecular biology and biotechnology research to isolate a specific DNA segment, to construct recombinant clones of genes of interest, orto sequence a desired DNA fragment.

DNA probes can be generated directly from a known DNA sequence or indirectly from a polypeptide sequence. Colony hybridization relies heavily of availability of complementary oligonucleotide probe. A unique 20 nucleotide long sequence occurs in every 4^{20} (~ 10^{12}) nucleotides therefore for a probe to bind uniquely to the clone of interest, it should be minimum 20 nucleotides long. Apart from use in hybridization experiments, these custom designed oligonucleotide DNA probes can be used to introduce recognition sequence of a particular restriction enzyme into cloned DNAs and thereby help to construct recombinant clones of genes of interest.

DNA probes can be homologous or heterologous. Homologous DNA probes share high degree of complementarity with the gene of interest. Synthesis of homologous DNA probes requires information about protein or DNA sequence of the gene of interest. Thesehomologous genes probesare similar to enable hybridizationeven under high stringency conditions. Sometimes, probes that are not perfectly homologous are used to screen for a particular gene of interest. These are heterologousprobes and are extremely useful for library screening especially when probe from one organism is used to detectclones prepared from DNA libraryof a second organism.

DNA oligonucleotides are short linear sequences of DNA about 10–20 nucleotides long. Theycan be synthesized chemically by DNA synthesizers using the precursor phosphoamidines. These are chemically protected molecules. The process of synthesis is automated and performed on solid supports. DNA synthesizers can be programmed to synthesize oligonucleotides up to 60-100 bases by just typing the desired sequence into computer controlling the machine. The direction of chain growth is by addition to 5'end of molecule.

These oligonucleotides have critical utility as primers in PCR and sequencing reaction. Custom designed oligonucleotides are helpful tools in site directed mutagenesis to create

directed mutations in the cloned DNA. Newly synthesized oligonucleotides having a mismatch to a segment of cloned DNA is hybridized to the cloned DNA and used as a primer for DNA synthesis with cloned DNA acting as a template. The newly amplified double stranded molecule bears mismatch at customized position. The complementary strands of newly synthesized DNA are denatured and desired mismatch is amplified further.

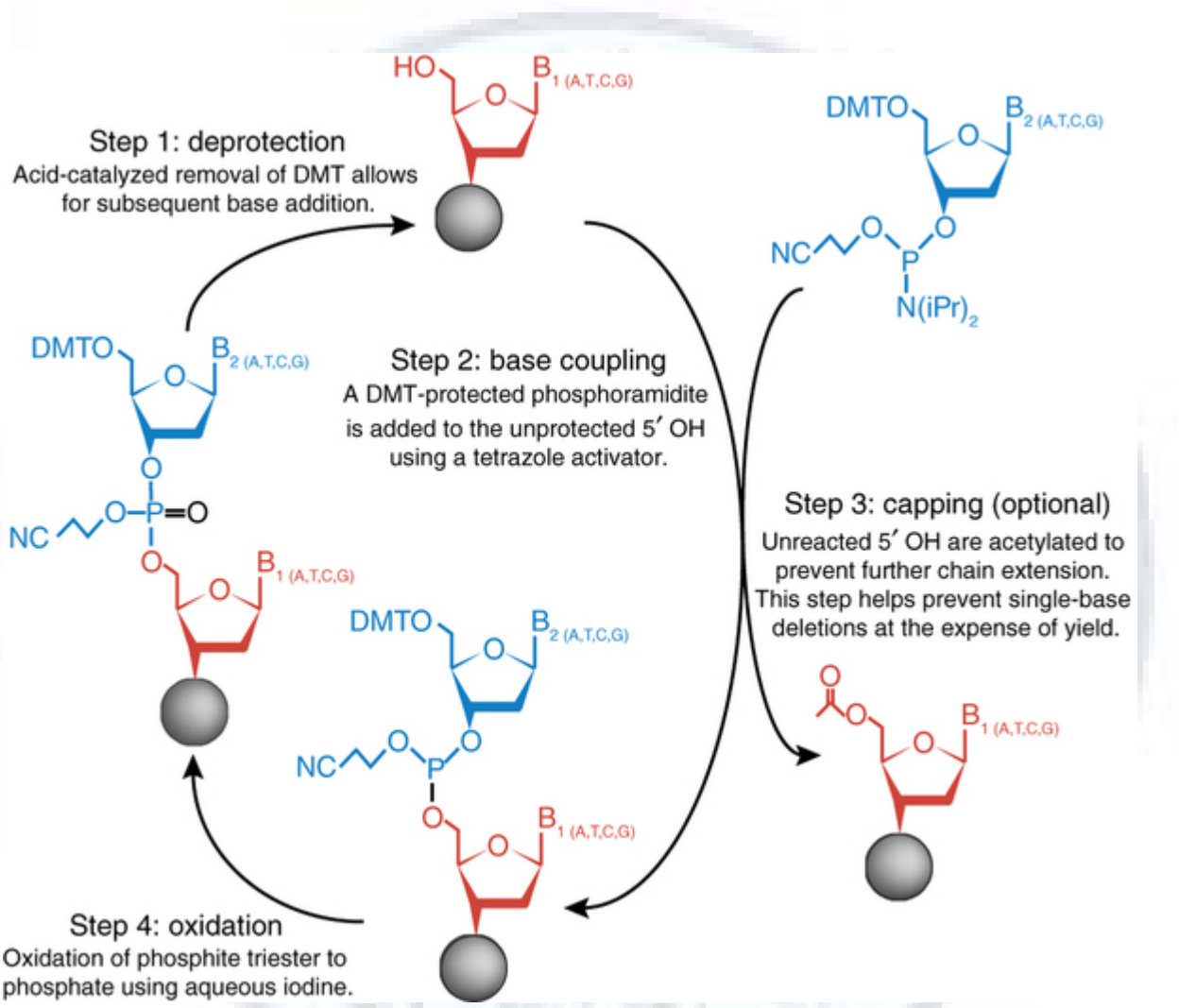


Figure : Four main steps in artificial synthesis of DNA oligonucleotides.

Source: Kosuri, Sriram, and George M. Church. "Large-scale de novo DNA synthesis: technologies and applications." *Nature methods* 11.5 (2014): 499-507.

http://www.nature.com/nmeth/journal/v11/n5/abs/nmeth.2918.html (Displayed with permission)

Link for complete steps in DNA synthesis -

Construction of genomic and cDNA libraries

http://ipmbgazette.weebly.com/uploads/1/0/3/0/1030249/mary_naluguza.pdf

Summary

1. DNA libraries are powerful tools in plant biotechnology research for isolating and studying genes of interest in specific cells/tissues/ organism.
2. Depending upon the starting material used for library construction, DNA libraries can be of two types: Genomic library and cDNA library.
3. Genomic library contains DNA sequences from entire genome of the organism and provide a catalog of all the genes present in the organism.
4. cDNA libraries are constructed from complementary DNA synthesized by mRNA and provide instant snapshot of transcriptionally active genes in the tissue used to make library.
5. Identification of the target gene/cDNA from the constructed library is utmost important and genetic selection and screening methods are employed at the same time to achieve success. Genetic selection and screening methods require expression or repression of traits of both vector and gene of interest. Libraries can bescreened for the presence of specific genes of interest. Genetic selection remains a prerequisite tool to identify recombinant vector. A common method of plasmid-based genomic libraryscreening is colony hybridization. The basis of approach is complementarity between single stranded DNA or RNA molecules to associate with each other. PCR is also an efficient method for library screening. The procedure of immunological screening is similar to hybridization screening except the fact that antibodies are used as probe instead of nucleic acid. Identification of a gene by utilizing mutant strains of yeast by complementation provides a rapid and simple means of screening yeast genomic library.
6. DNA probes can be generated directly from a known DNA sequence or indirectly from a polypeptide sequence. These oligonucleotides are indispensable tools in colony hybridization, cloning, PCR, sequencing reaction and in site directed mutagenesis.

Exercises

1. Define the following terms

(i) Genome (ii) Contig (iii) Phosphoamidines (iv) Reverse Transcriptase (v) Site directed mutagenesis (vi) Probe (vi) Genomic library (vii) Plaque
(viii) Hybridization (ix) Complementation

2. Name the scientists who coined the terms genome and genomics?
3. Why are DNA libraries constructed?
4. Explain the method used to construct a genomic library?
5. List important differences between genomic and cDNA libraries?
6. Describe importance of the use of a genomic library to provide information that acDNA library cannot?
7. Why reverse transcriptase is used to construct a eukaryotic expression library?
8. Which characteristic property mRNA permits their easy isolation from total RNA?
9. List the steps you will follow to construct a cDNA library?
10. What is the importance of attaching a short linker to double stranded DNA molecules, during cDNA library preparation?
11. Genetic selection methods are a prerequisite tool in identification of recombinants in population of cells. Comment?
12. Briefly explain genetic selection methods employed to select recombinants in the library?
13. Discuss the methods followed to identify clones harboring fragments of genes of interest from the library?
14. How immunological screening differs from colony hybridization?
15. Write short note on (i) Blue –White screening (ii) Colony Hybridization
16. Gene identification from yeast genomic library can be done with the help of yeast mutants defective in gene of interest. Explain the process in detail?
17. What are DNA oligonucleotides? How are they synthesized? Comment on the utility of DNA oligonucleotides in molecular and biotechnology research?
18. Differentiate between homologous and heterologous DNA probes?

Glossary

Antibody:A protein molecule used for detection of antigen by identifying and binding to a unique part of antigen.

Autoradiography: It is a technique used tolocalize radioactively labeled compounds within cells and tissuestoproduce a photographic image by radioactive decay.

β-galactosidase:A bacterial protein, encoded by the *lacZ*gene.

Bacterial artificial chromosome (BAC):A vector for transforming and cloning large DNA fragments up to about 200 kb long in bacteria.

Bacteriophage: They are also called phages and are virus that infects bacteria.

cDNA (complementary DNA): DNA synthesized from an mRNA template in a reaction catalyzed by the enzyme reverse transcriptase.

cDNA library:A collection of cloned complementary DNAs (cDNA) sequences with each clone carrying a specific cDNA derived from a single mRNA.

CEN: Nucleotide sequence of DNA in yeast centromere to allow faithful segregation of plasmid during yeast cell division.

Chromosome: A circular DNA molecule containing the organism's genome in prokaryotes. In eukaryotes, it is a thread like structure visible during mitosis and meiosis and composed of linear DNA molecule complexedwith RNA and proteins.

Complementarity: Hydrogen bonding between nitrogenous bases of DNA or RNAwhen they are placed antiparallel to each other.

Complementation test:A test used to determine whether two mutations that confer the same phenotype occur within the same gene. In this method, a yeast mutant defective in the gene of interest is chosen and a yeast genomic library is transformed in yeast cells. Yeast mutant is complemented when expression of transgene from library is induced.

Clone: Identical copies of a molecule, cells, or organisms.

Cloning: The production of exact copies of a DNA molecule by transformation and replication in a suitable host.

Contig:A contigous DNA sequence reconstructed from overlapping DNA sequences by sequence alignment.

deoxyribonucleic acid (DNA): The carrier of genetic information in all living cells. It isapolymer consisting of antiparallel chains of nucleotide which are held together by hydrogen bonds.

DNA ligase: An enzyme that catalyzes covalent bond formation between the 5'end of one DNA chain and the 3'end of another DNA chain during DNA replication andDNA repair.

Exon:A protein-coding region of a gene that specifies an amino acid sequence and is retained in mRNA after splicing.

Gene:A nucleotide sequence of DNA coding for a single polypeptide and represents the fundamental functional and physical unit of heredity.

Construction of genomic and cDNA libraries

Genome: Genes present in a chromosome set of a particular species determines its genome. For eukaryotic organisms, genome refers to amount of genetic material in haploid set of chromosomes of a particular organism.

Genomic library:A repertoire of cloned DNA fragments representing all the DNA sequences of an organism's genome.

Genomics: Thestudy of entire genomes of the organisms to gain insights about its structure and function.

Hybridization: Complementarity between single stranded DNA or RNA molecules to associate with each other.

Intron:A non-coding region of DNA that is present between coding regions and is removed by splicing and is not represented in the polypeptide.

messenger RNA (mRNA): DNA after transcription produces mRNA, which is then translated to amino acid sequence of a protein. Itis present in small percentage in a mixture of total RNA which contains prevalently rRNA and tRNA. They have several adenines attached at their 3' end called poly A tails.

Oligonucleotide:A short linear sequence of DNA about 10–20 nucleotides.

Plaque: A clear zone in an opaque lawn of bacteriabecause ofgrowth and reproduction of phage.

Plasmid: Adouble stranded extra chromosomalcircular DNA molecule that replicates autonomouslyof the host chromosome.

Polylinker: It is also known as multiple cloning site and is a segment of DNA that has been constructed to contain sites for multiple restriction enzymes.

Polymerase chain reaction (PCR): A procedure for producing exact multiples of a specific DNA sequence.

Primer:A short single-stranded DNA sequence for initiating synthesis by the enzyme polymerases.

Probe: Apurified fragment of single-stranded DNA ranging from 25 to several thousand in length used to isolate specific DNA segment, construct recombinant clones of genes of interest, to amplify and sequence a desired DNA sequence. Probes can be labeled and can be detected.

Restriction endonuclease:A bacterial enzyme that recognizes within or near a specific nucleotide sequences called restriction site in a DNA molecule, and cleaves the double stranded DNA at those sites.

Restriction site:A DNA sequence, within or near to which restriction endonuclease binds and cleaves the DNA.

Restriction site linker:A8 to 12 base pairs long double-stranded oligodeoxyribonucleotide that harbors the restriction site for a specific restriction enzyme.

Retrovirus:A virus that uses RNA as its genetic material. It possessesthe enzyme reverse transcriptase.

Reverse genetics: An experimental approach used to deduce thegene function after the gene sequence has been identified.An approach reverse of*forward genetics*.

Reverse transcriptase:A RNA dependent RNA polymerase that uses RNA as a template to make a single-stranded DNA molecule as a product.

Shuttle vector:A cloning vector that can replicate in two or more host organisms (e.g.*E. coli*and yeast).

Site-directed mutagenesis:A process that creates directed mutations in the cloned DNA by using a synthetic oligonucleotide containing a mutant base in its sequence.

Transformant: A recombinant recipient cell generated by transformation in bacteria.

Transformation: A process by which genetic information is transferred by exogenous DNA as a heritable change.

Vector: It is a double-stranded DNA molecule which can replicate autonomously in a host cell and into which a foreign DNA segment can be inserted.In essence it isa delivery agent/vehicle.

YAC:A cloning vector in the form of a yeast artificial chromosome, for cloning large DNA fragments synthesized using chromosomal components includinga centromere region (CEN), origin of replication (ARS),and marker genes from yeast.

LINK FOR ANIMATION: For construction of a DNA library

http://www.sumanasinc.com/webcontent/animations/content/dnalibrary.html

References

1. Lodish, H., Berk, A., Zipursky, S.l., Matsudaira, P., Baltimore, D. and James Darnel (2000) Molecular Cell Biology, W. H. Freeman and Company.6th edition.

2. Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., Losick, R. (2007). Molecular Biology of the Gene, Pearson Benjamin Cummings, CSHL Press, New York, U.S.A. 6th edition.

3. Snustad, D.P. and Simmons, M.J. (2010). Principles of Genetics. John Wiley and Sons, U.K. 5th edition.

4. Klug, W.S., Cummings, M.R., Spencer, C.A. (2009). Concepts of Genetics. Benjamin Cummings, U.S.A. 9th edition.

Web links

5. http://classes.midlandstech.edu/carterp/Courses/bio225/chap09/lecture5.htm
6. http://www.rvc.ac.uk/review/DNA_1/1c_DNA_Lib.cfm
7. http://www.rvc.ac.uk/review/DNA_1/1d_Screening.cfm