Department of BSBE
Indian Institute Of Technology Guwahati

Introduction to microbiology and study of microorganisms:

cDNA and genomic DNA library construction

Dr. Sanjukta Patra

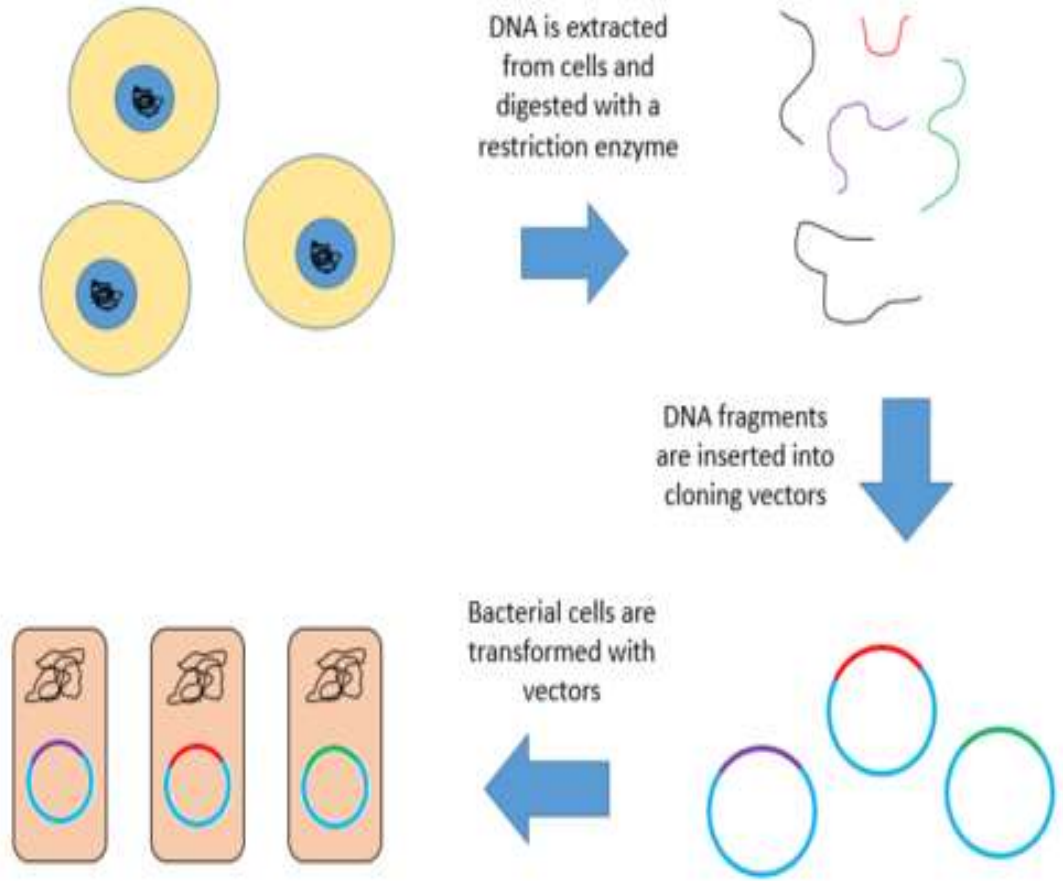
BT 207

Genetic Engineering

Jan-April 2023

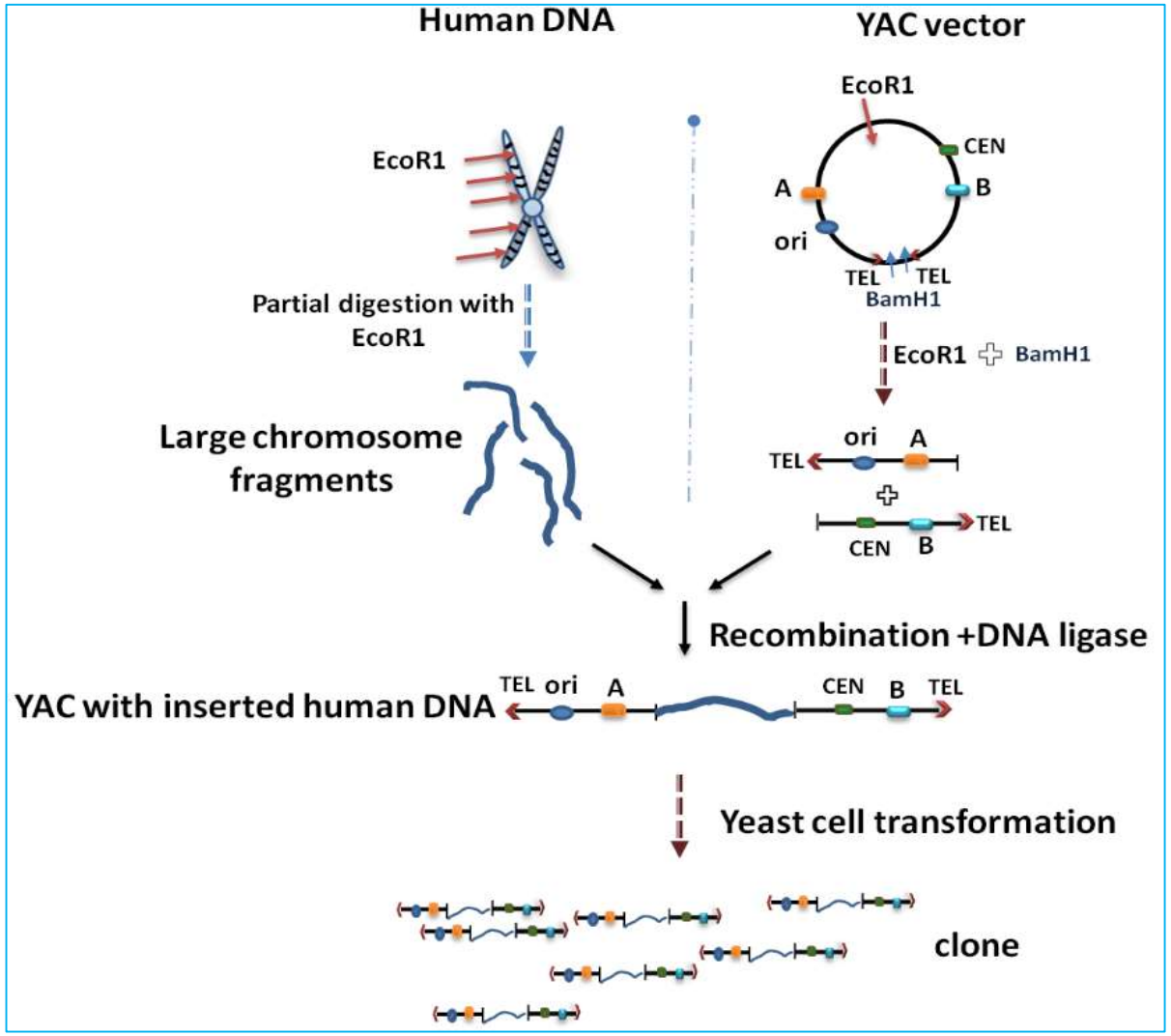
Genomic DNA Library

- 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.
- Depending upon organism and size of genome, this library is either prepared in BAC or YAC.



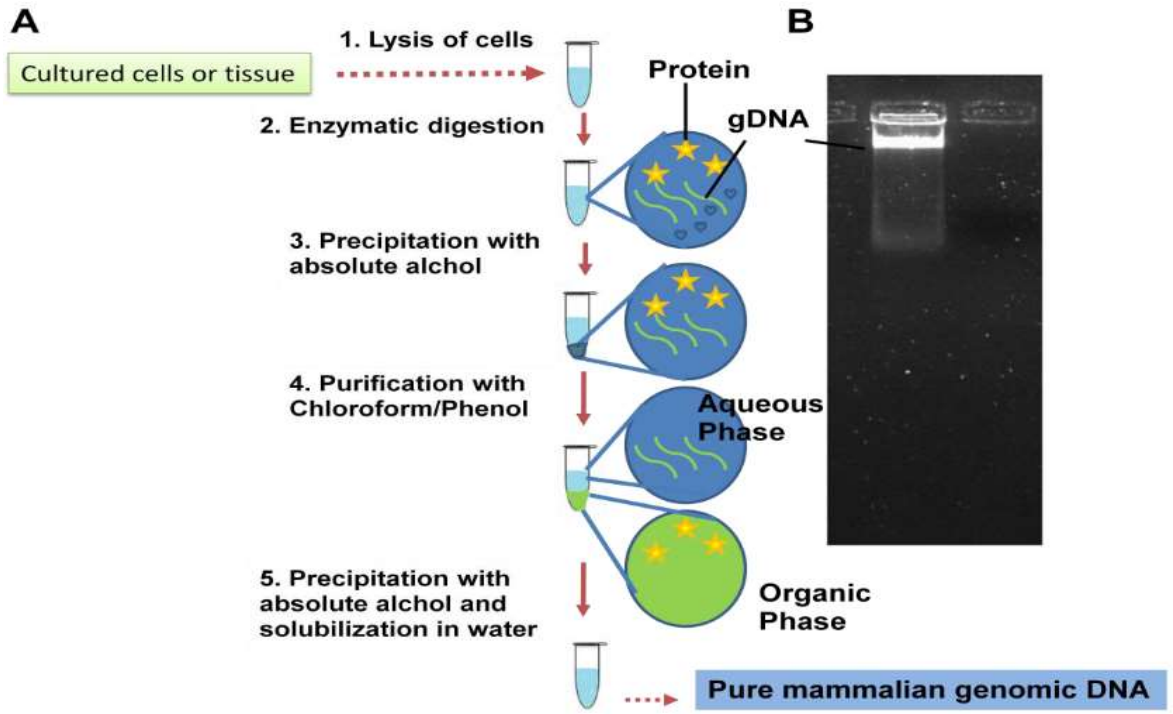
STEPS:

1. Isolation of genomic DNA
2. Generation of suitable size DNA fragments
3. Cloning in suitable vector system (depending on size)
4. Transformation in suitable host .



is the fractional proportion of the genome in a single recombinant

1. Isolation of genomic DNA



2. Generation of suitable size fragments

Restriction digestion: Genomic DNA can be digested with a frequent DNA cutting enzyme such as EcoR-I, BamH-I or sau3a to generate the random sizes of DNA fragments.

If a organism has a genome size of **$2x10^7$** kb and an average size of the fragment is 20kb, then no. of fragment, n= **$2x10^7$** kb / 20kb.

The probability (P) of finding a particular genomic sequence in a random library of N independent clone is as follows:

Carbon and Clarke formula

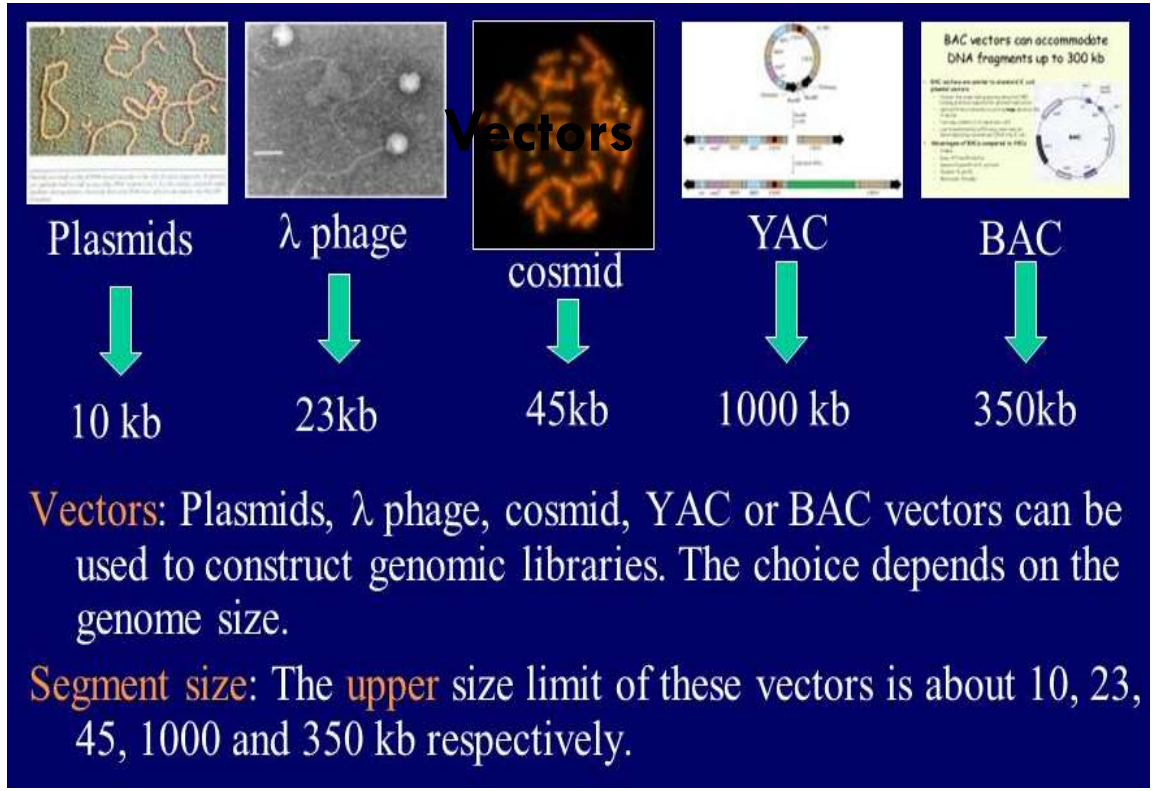
$$N=\ln(1-P)/\ln(1-f)$$

Where, N=number of clones, P=probability, f=fractional proportion of genome in a single recombinant

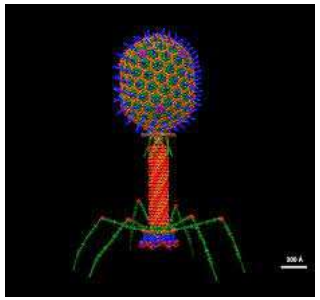
f=i/g where, i is the insert size & g is the genome size

3. Cloning into the suitable vector

- The suitable vector to prepare the genomic library can be selected based on size of the fragment of genomic DNA and carrying capacity of the vector Size of average fragment can be calculated and accordingly a suitable vector can be chosen.
- In the case of fragment generated by restriction enzyme, vector can be digested with the same enzyme and put for ligation to get clone.
- In one of the approaches, a adapter molecule can be used to generate sticky ends, alternatively a endonuclease can be used to generate sticky ends.
- In the case of mechanical shearing mediated fragment generation, putting these fragment needs additional effort.



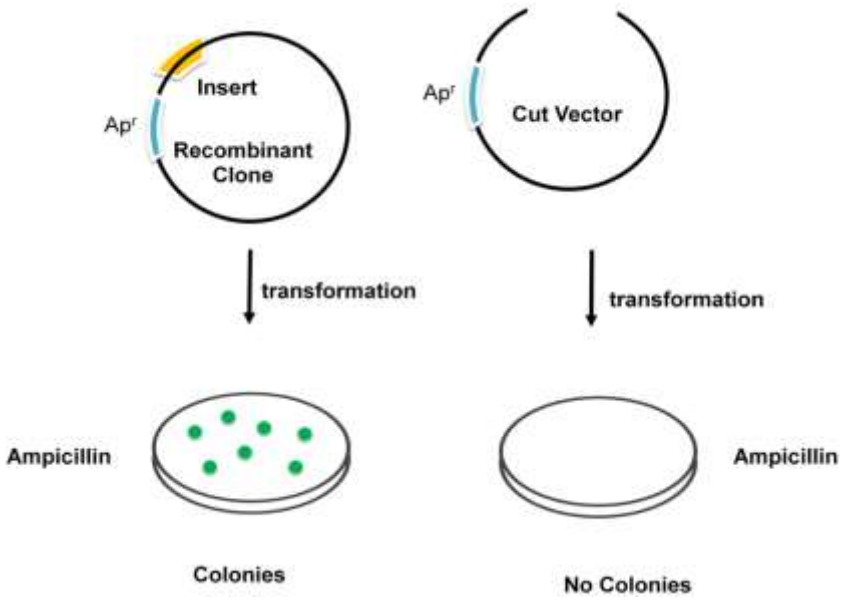
Can you tell the carrying capacity of bacteriophages?



Do you know what are PACs?

4. Transformation to get colonies

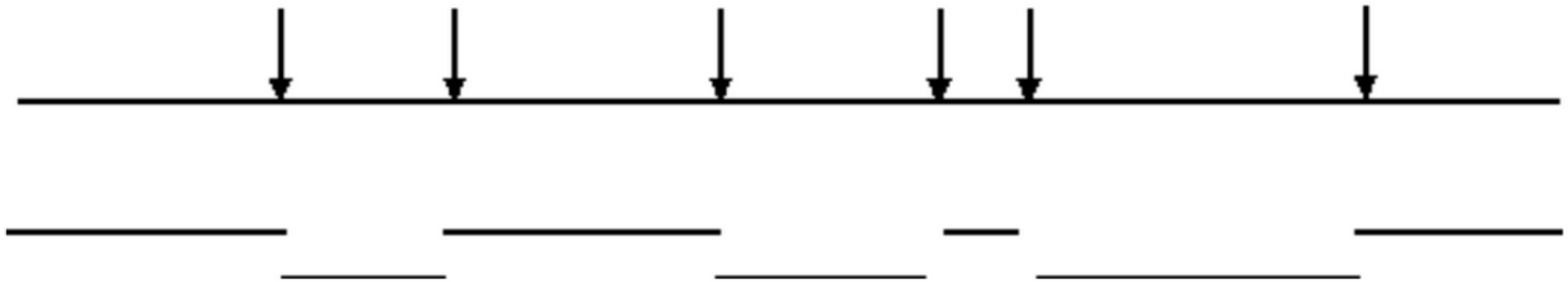
Post ligation, clones are transformed in a suitable host to get colonies. A suitable host can be a bacterial strain or yeast. Different methods of delivering clone into the host cell is present. Eg.: transformation, electroporation, chemical transfection, viral transduction,



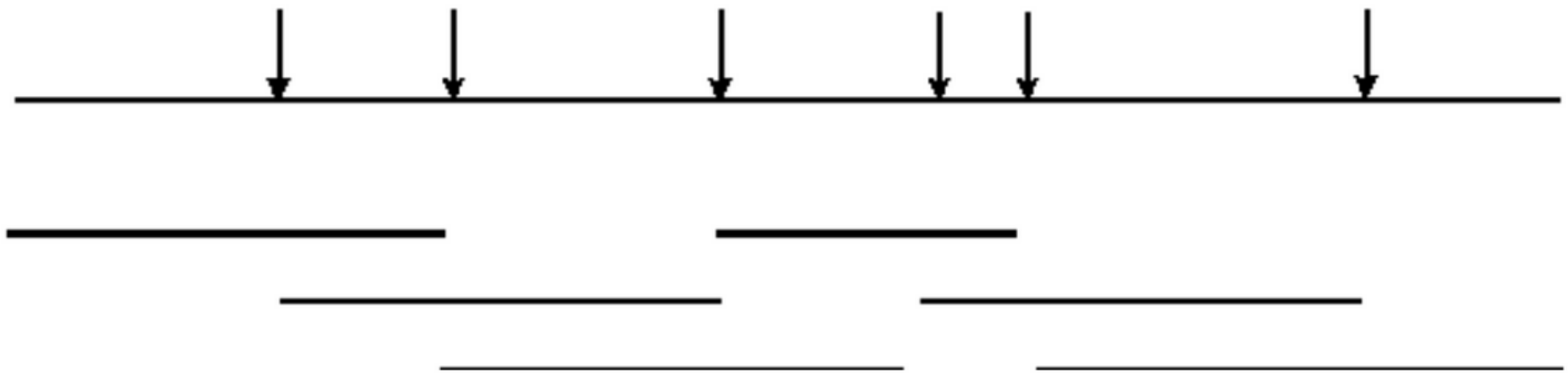
- After a library is created, the genome of an organism can be **sequenced** to elucidate how genes affect an organism or to compare similar organisms at the genome-level.
- The **genome-wide association studies** can identify candidate genes stemming from many functional traits. Genes can be isolated through genomic libraries and used on human cell lines or animal models to further research.
- Furthermore, creating high-fidelity clones with accurate genome representation and no stability issues would contribute well as intermediates for **shotgun sequencing** or the study of complete genes in functional analysis.

Incomplete Digestion of Genomic DNA will allow identification of sequence overlaps

Complete digestion with an endonuclease will result in a library containing *no overlapping fragments*:



However, *incomplete digestion* will result in a library containing overlapping fragments:



- Thus, the sequence information obtained from *one clone* will allow the isolation of clones containing *neighboring (overlapping) sequence information*.
- This can allow large contiguous stretches of sequence information to be obtained ("Chromosome Walking").

5. Probing libraries

Once a library (cDNA or genomic) has been constructed we want to be able to identify clones which contain DNA of interest.

•For example, from protein sequence information we can *deduce possible stretches of the corresponding DNA sequence* (there will however be ambiguity due to the degeneracy of codons).

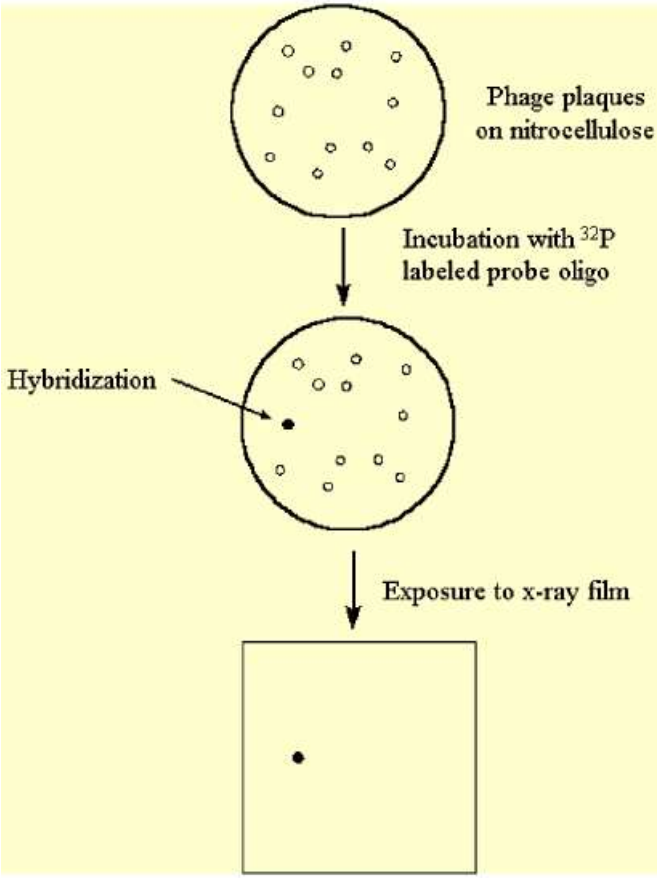
•If we can synthesize an oligonucleotide complementary to our DNA sequence of interest we can use it to specifically hybridize to the appropriate clone in our libraray (i.e. to **probe** our library).

In standard methodologies the oligonucleotide is *phosphorylated* at the 5' end with radiolabeled g^{32}P-ATP and *T4 polynucleotide kinase*.

•The probe is then incubated with individual phage plaques which have been fixed onto nitrocellulose and their DNA denatured by treatment with base.

•If the plaque contains complementary DNA to to probe sequence, the probe will hybridize.

•If the nitrocellulose (containing many individual plaques) is exposed to x-ray film, *only those plaques with hybridized probe will show up (as a dark spot)*



Radiolabeled plaque

Genomic DNA Library uses

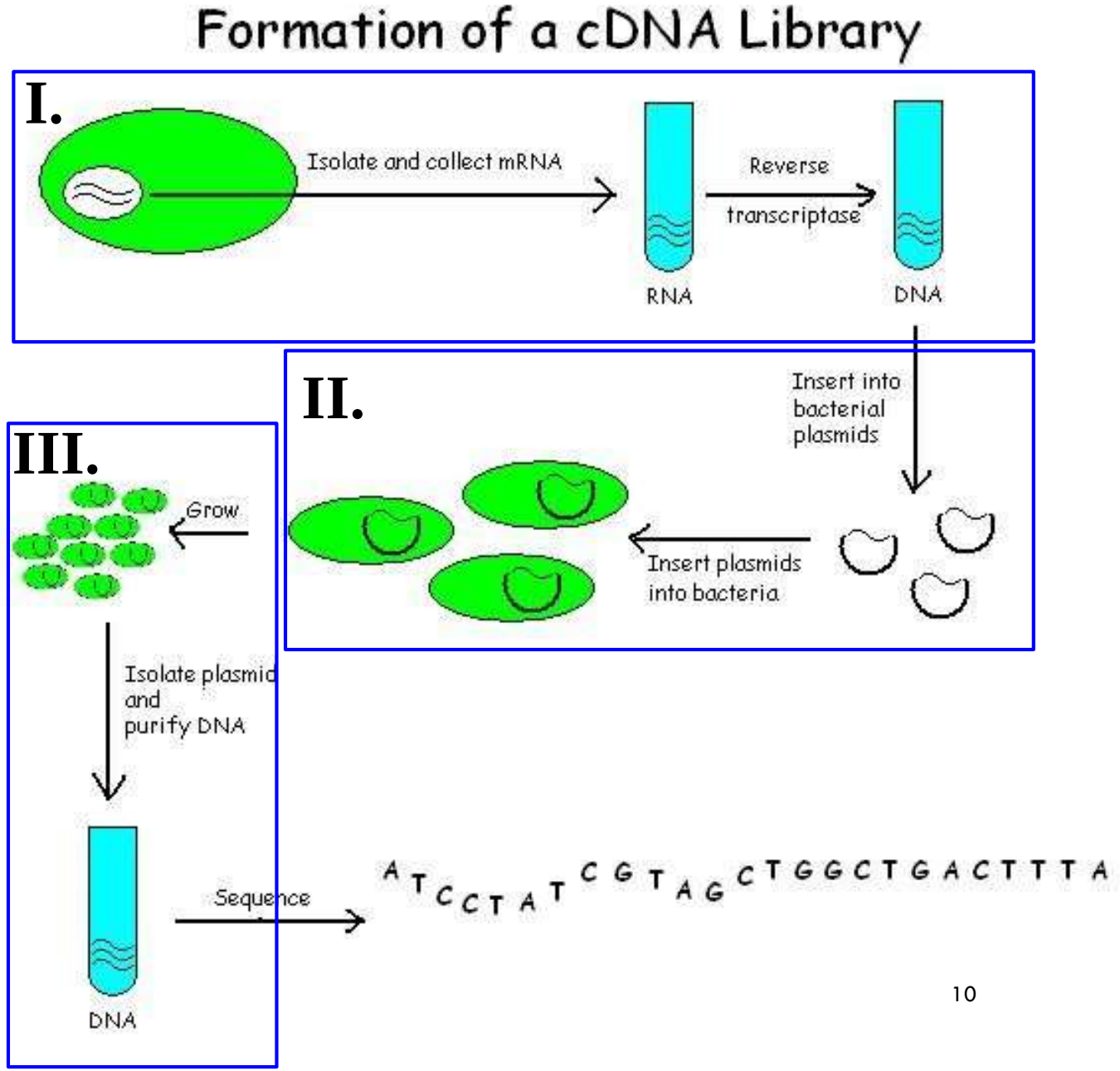
1. Genomic DNA projects – human genome project
2. Identification of new genes – Pharmaceutically important genes, enzymes etc
3. Helps in identification of genes which are silent in host
4. Study function of non-coding regions – Gene Regulatory regions

- A cDNA library is a combination of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which constitute some portion of the transcriptome of the organism and are stored as a **"library".**

- cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism at one point of time.

A "library" is a convenient storage mechanism of genetic information.

Library?



"**Reverse transcription**" is a mechanism whereby genetic information contained in mRNA is converted back into a double stranded DNA form! The enzyme responsible for this is an RNA dependent DNA polymerase called reverse transcriptase.

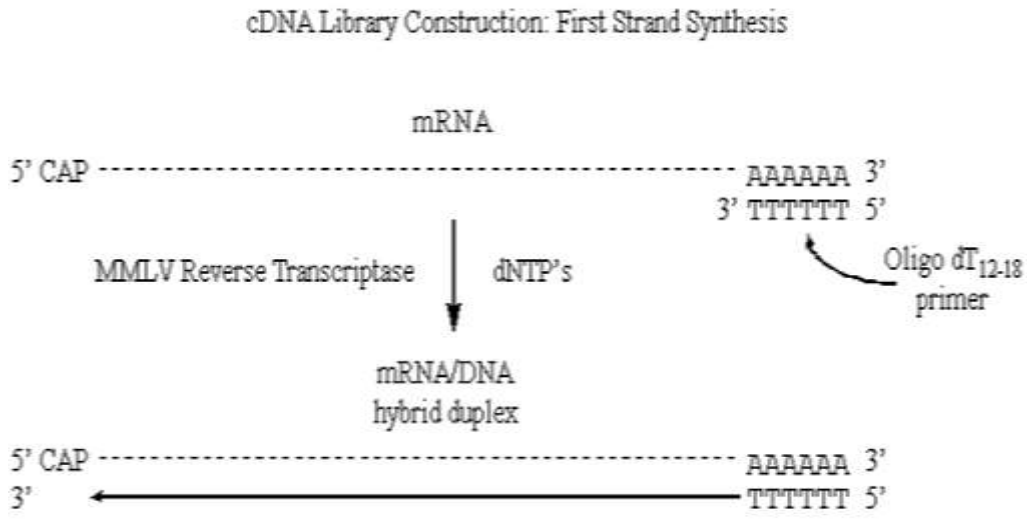
- One of the most common commercially available reverse transcriptases is Moloney murine leukemia virus (MMLV). Viruses whose genome is actually in an RNA form and must be converted to duplex DNA, isolated from them. These viruses typically carry a functional reverse transcriptase along with their mRNA genetic component when they infect cells.

- This RNA dependent DNA polymerase (as will all polymerases) add nucleotides to a nascent polynucleotide in the **5' to 3' direction using RNA as the template**. It does not contain any 3'->5' exonuclease (proofreading) activity.

- MMLV will use mRNA as a template, but requires a primer (it can extend a DNA primer but cannot synthesize one).

- One of the really neat things about eukaryotic mRNA's is the presence of the **3' poly A tracks**. Therefore we can use poly dT as a single primer for a variety of different eukaryotic mRNA's.

Step 1



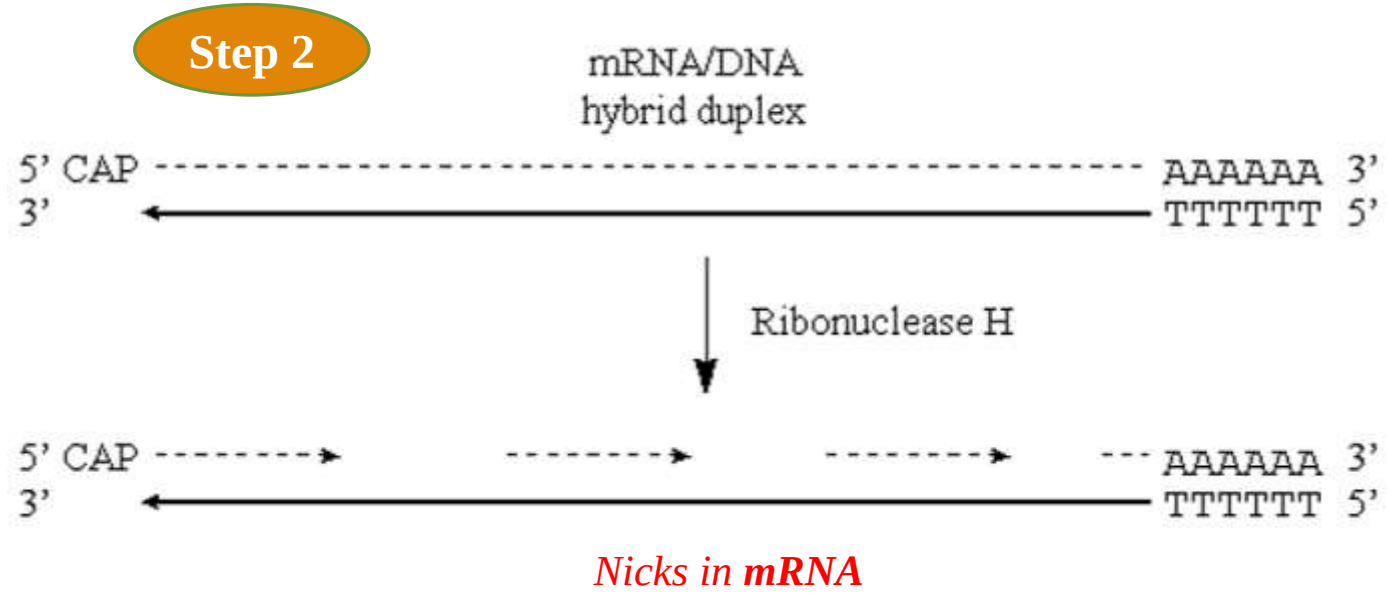
First strand synthesis
*Note that **complementary DNA (or cDNA)** is produced to the original mRNA strand*

•Nicks in the RNA half of the molecule can be introduced via the action of the enzyme **RNAse H.**

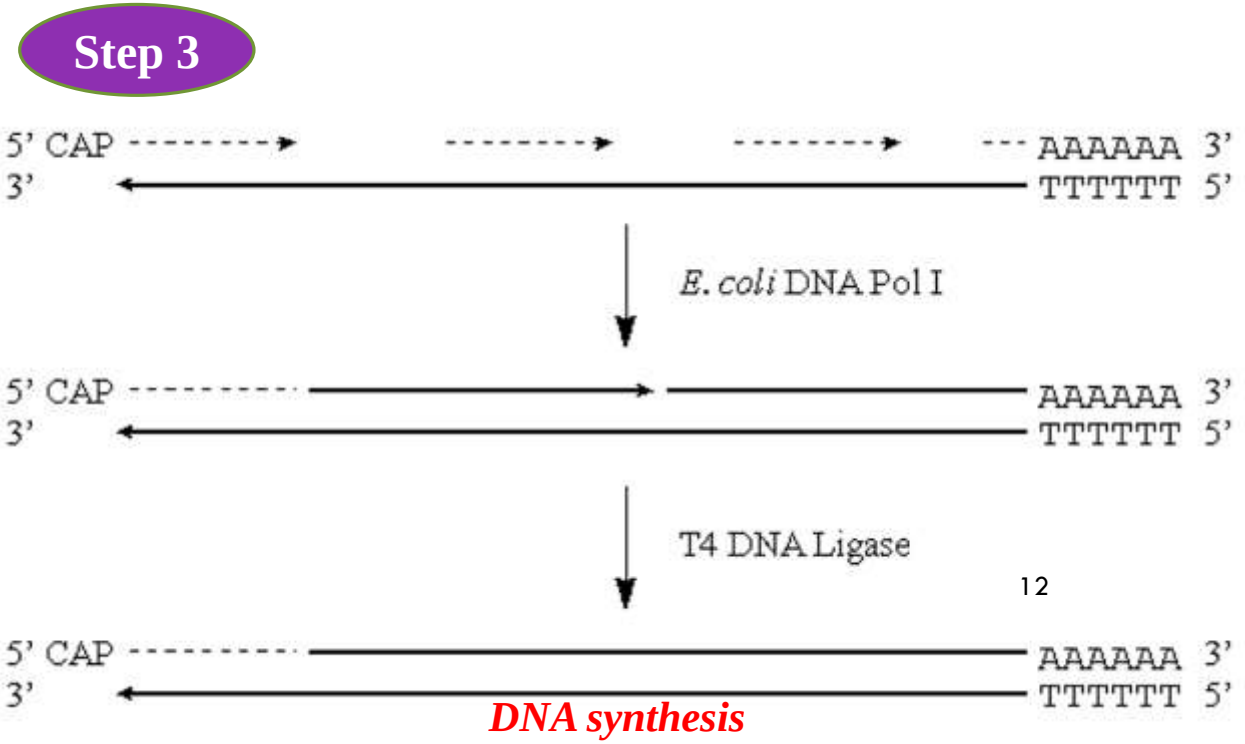
•This enzyme exhibits endonucleolytic cleavage of the RNA moiety of RNA/DNA hybrids, as well as 5'->3' and 3'->5' exoribonuclease activity.

•In other words, it will nick the RNA and then proceed to digest back in both directions:

•These RNA fragments can now serve as primers for DNA synthesis by *E. coli* **Pol I**. This enzyme will also translate the "nicks" to effectively remove the RNA primers:



*Nicks in **mRNA***



DNA synthesis

Insertion of cDNA into plasmid

To complete our construction of a useful cDNA library we need a way to *maintain and propagate* our cDNA.

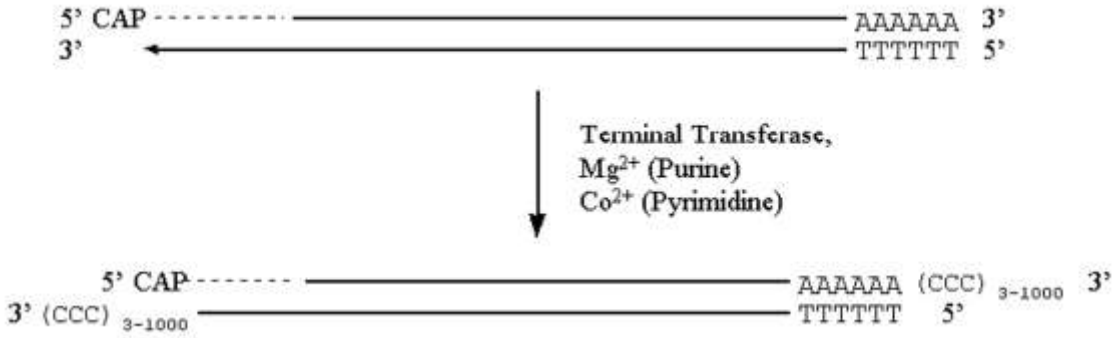
•We can accomplish this by inserting the cDNA into an appropriate **plasmid**.

•There are two classical ways of accomplishing this feat:

1.Homopolymeric tailing

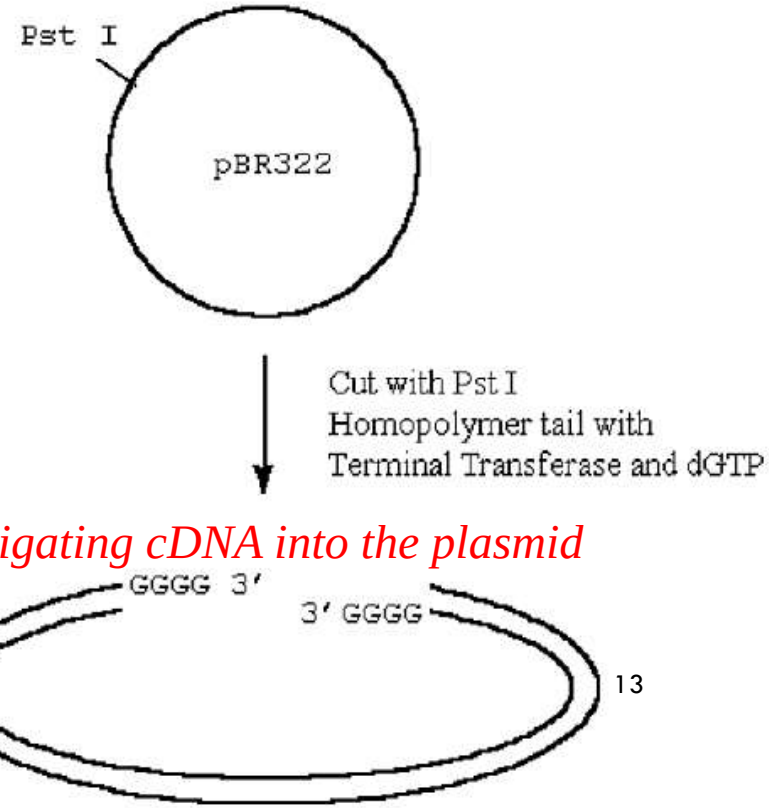
2.Linker addition

Step 4



Terminal transferase activity

Step 5



Ligating cDNA into the plasmid

Homopolymeric tailing

Terminal transferase is an unusual DNA polymerase found only in a type of eukaryotic cell called a *prelymphocyte*.

•In the presence of a divalent cation the enzyme catalyzes the addition of dNTP's to the 3'-hydroxyl termini of DNA.

•If we cut the plasmid and also treat it with terminal transferase, except now by adding the complementary base to the one added to the cDNA, we can **anneal and ligate the cDNA into the plasmid**

cDNA Library

- cDNA libraries are commonly used when reproducing eukaryotic genomes, as the amount of information is reduced to remove the large numbers of non-coding regions from the library.
- cDNA libraries are used to express eukaryotic genes in prokaryotes. Prokaryotes do not have introns in their DNA and therefore do not possess any enzymes that can cut it out during transcription process.
- cDNA does not have introns and therefore can be expressed in prokaryotic cells.
- cDNA libraries are most useful in reverse genetics where the additional genomic information is of less use.
- Additionally, cDNA libraries are frequently used in functional cloning to identify genes based on the encoded protein's function.
- When studying eukaryotic DNA, expression libraries are constructed using complementary DNA (cDNA) to help ensure the insert is truly a gene.

cDNA Library vs. Genomic DNA Library

cDNA library lacks the non-coding and regulatory elements found in genomic DNA.
Genomic DNA libraries provide more detailed information about the organism, but are more resource-intensive to generate and keep.

For example, how large a library (i.e. how many clones) would you need in order to have a 99% probability of finding a desired sequence represented in a library created by digestion with a 6- and 8- cutter?