Subject Name: Microbiology

Semester: IV

Name of the Teacher:

Dr. Kuntal Kanti Goswami

Topic:

Construction and Screening of Genomic and cDNA Libraries

(CC10; Unit 5)

B.Sc (HONOURS) MICROBIOLOGY (CBCS STRUCTURE)

CC-10: DNA Amplification and DNA Sequencing (THEORY)

SEMESTER -IV

Genomic DNA Library:

A genomic library is a set of recombinant clones that contains all of the DNA present in an individual organism.

An *E. coli* genomic library, for example, contains all the *E. coli* genes, so any desired gene can be withdrawn from the library and studied. Genomic libraries can be retained for many years, and propagated so that copies can be sent from one research group to another.

The big question is how many clones are needed for a genomic library? The answer can be calculated with the formula:

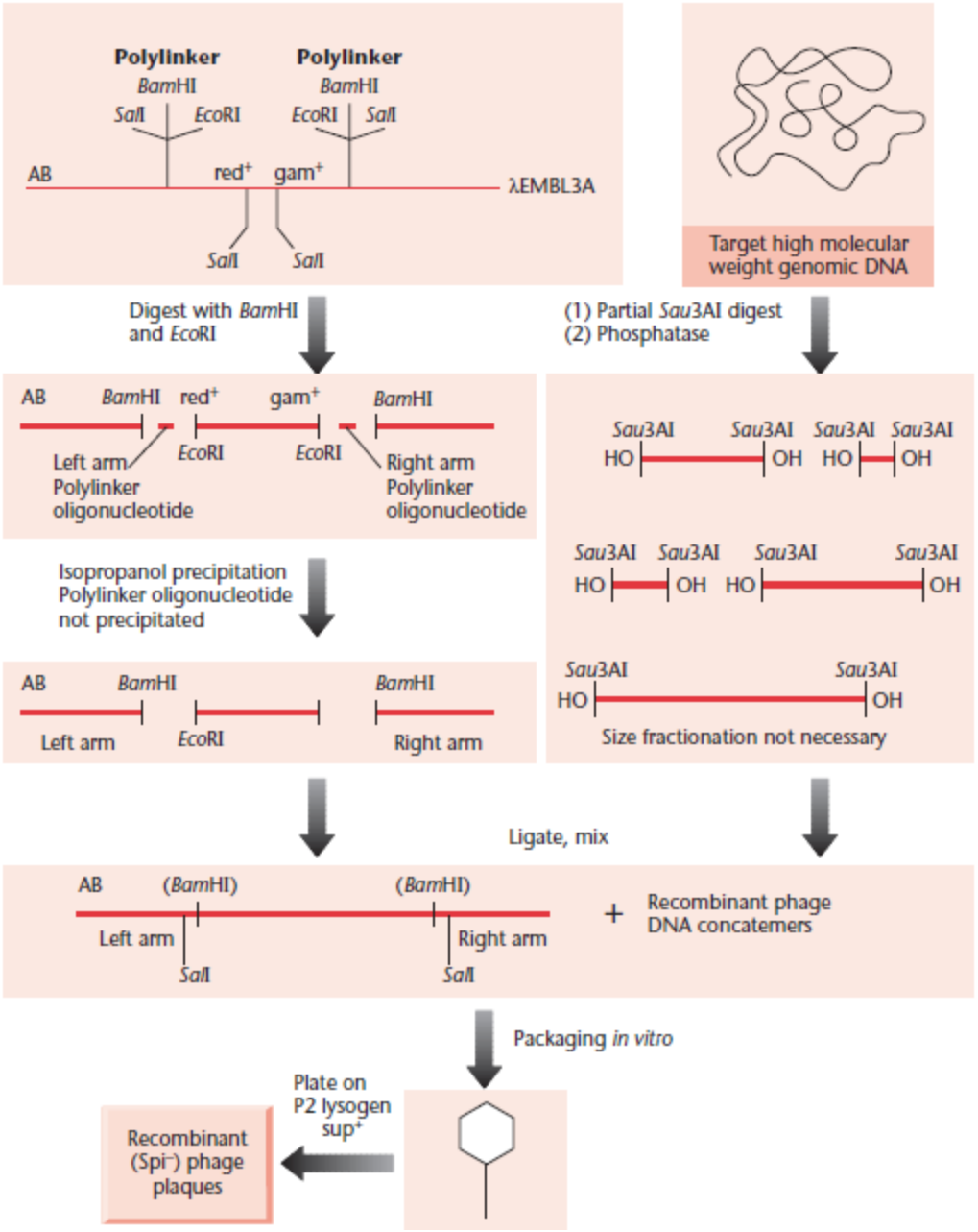
$$N = \frac{\ln(1-p)}{\ln\left(1-\frac{a}{b}\right)}$$

Where, **N** is the number of clones that are required, **p** is probability that any given gene will be present, **a** is the average size of the DNA fragments inserted into the vector, and **b** is the total size of the genome.

For humans and other mammals, several hundred thousand clones are required. It is by no means impossible to obtain several hundred thousand clones, and the methods used to identify a clone carrying a desired gene can be adapted to handle such large numbers, so genomic libraries of these sizes are by no means unreasonable. However, ways of reducing the number of clones needed for mammalian genomic libraries are continually being sought.

One solution is to develop new cloning vectors able to handle longer DNA inserts. The most popular of these vectors are bacterial artificial chromosomes (BACs), which are based on the F plasmid. The F plasmid is relatively large and vectors derived from it have a higher capacity than normal plasmid vectors. BACs can handle DNA inserts up to 300 kb in size, reducing the size of the human genomic library to just 30,000 clones. Other high-capacity vectors have been constructed from bacteriophage P1, which has the advantage over λ of being able to squeeze 110 kb of DNA into its capsid structure. Cosmid-type vectors based on P1 have been designed and used to clone DNA fragments ranging in size from 75 to 100 kb. Vectors that combine the features of P1 vectors and BACs, called P1-derived artificial chromosomes (PACs), also have a capacity of up to 300 kb.

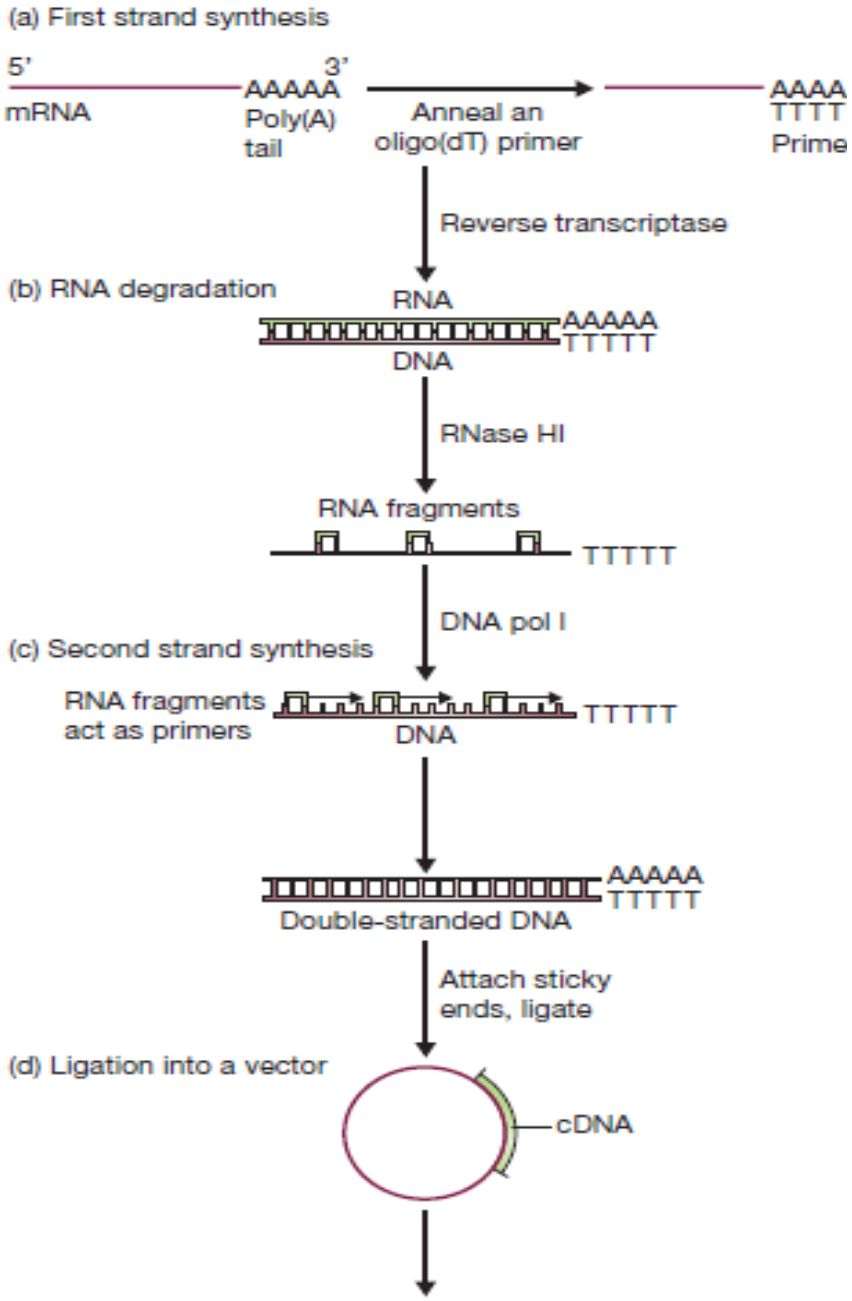
Organism	Genome size (kb) (haploid where appropriate)
Escherichia coli	4.0×10^3
Yeast (*Saccharomyces cerevisiae*)	1.35×10^4
Arabidopsis thaliana (higher plant)	1.25×10^5
Tobacco	1.6×10^6
Wheat	5.9×10^6
Zea mays	1.5×10^7
Drosophila melanogaster	1.8×10^5
Mouse	2.3×10^6
Human	2.8×10^6
Xenopus laevis	3.0×10^6



Creation of a genomic DNA library using the phage-λ vector EMBL3A. High-molecular-weight genomic DNA is partially digested with *Sau*3AI. The fragments are treated with phosphatase to remove their 5′ phosphate groups. The vector is digested with *Bam*HI and *Eco*RI, which cut within the polylinker sites. The tiny *Bam*HI/*Eco*RI polylinker fragments are discarded in the isopropanol precipitation, or alternatively the vector arms may be purified by preparative agarose gel electrophoresis. The vector arms are then ligated with the partially digested genomic DNA. The phosphatase treatment prevents the genomic DNA fragments from ligating together. Non-recombinant vector cannot re-form because the small polylinker fragments have been discarded. The only packageable molecules are recombinant phages. These are obtained as plaques on a P2 lysogen of sup^+ *E. coli*. The Spi^- selection ensures recovery of phage lacking *red* and *gam* genes. A sup^+ host is necessary because, in this example, the vector carries amber mutations in genes *A* and *B*. These mutations increase biological containment, and can be applied to selection procedures, such as recombinational selection, or tagging DNA with a sup^+ gene. Ultimately, the foreign DNA can be excised from the vector by virtue of the *Sal*I sites in the polylinker. (*Note*: Rogers *et al.* (1988) have shown that the EMBL3 polylinker sequence is not exactly as originally described. It contains an extra sequence with a previously unreported *Pst*I site. This does not affect most applications as a vector.)

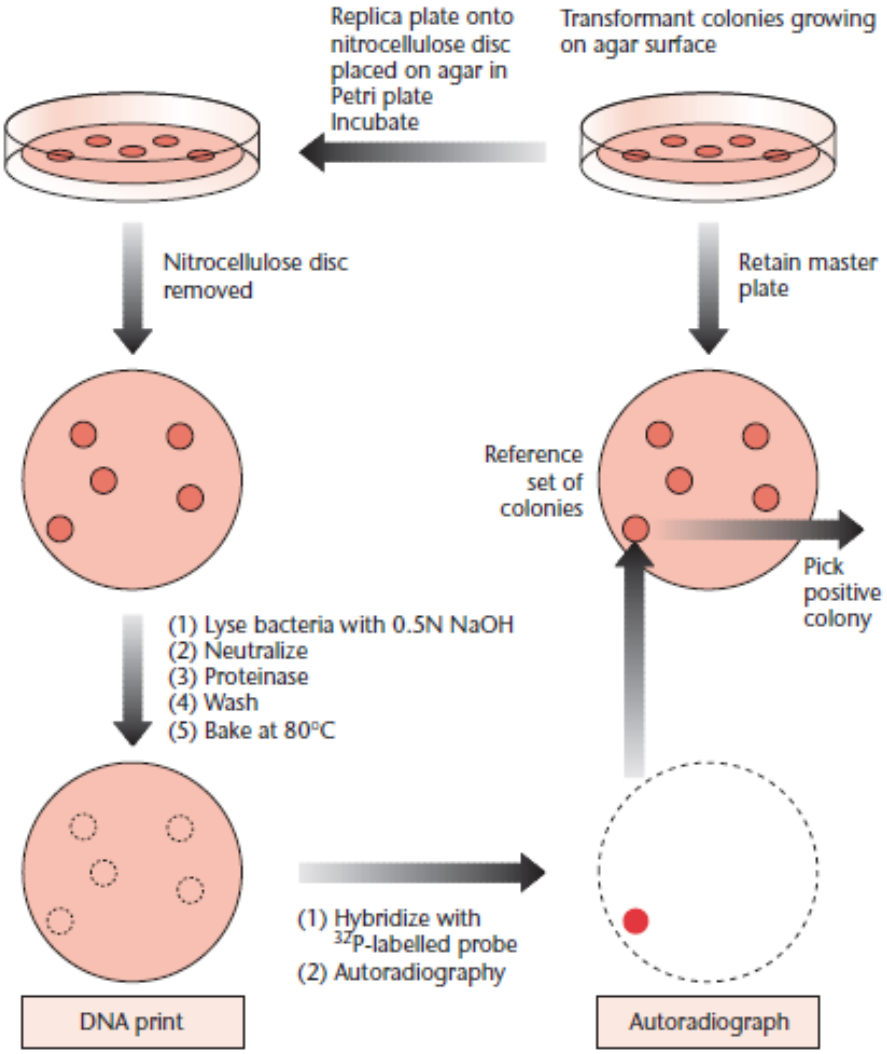
cDNA Library:

Messenger RNA cannot itself be ligated into a cloning vector. However, mRNA can be converted into DNA by complementary DNA (cDNA) synthesis. The key to this method is the enzyme reverse transcriptase, which synthesizes a DNA polynucleotide complementary to an existing RNA strand. Once the cDNA strand has been synthesized the RNA member of the hybrid molecule can be partially degraded by treating with ribonuclease (RNase) H. The remaining RNA fragments then serve as primers for DNA polymerase I, which synthesizes the second cDNA strand, resulting in a double-stranded DNA fragment that can be ligated into a vector and cloned. The resulting cDNA clones are representative of the mRNA present in the original preparation. In the case of mRNA prepared from wheat seeds, the cDNA library would contain a large proportion of clones representing gliadin mRNA. Other clones will also be present, but locating the cloned gliadin cDNA is a much easier process than identifying the equivalent gene from a complete wheat genomic library.



e) Transformation and Screening

Colony Hybridization:



Colony PCR:

Colony PCR involves designing PCR primers that will yield a specific product of known size only if the desired construct is present, and ideally, a product of a different size if the site of the desired manipulation remains unaltered. This technique takes advantage of the high sensitivity of PCR – the small amount of template DNA that is required to give an easily visualized band on an agarose gel following PCR amplification can be recovered from a very crude preparation of cells. Colony PCR is thus a powerful tool for rapidly and easily screening through potentially large numbers of colonies to distinguish true positives from false positives. In most cases it is a superior alternative to the older strategy of growing small cultures from several colonies, preparing microgram quantities of DNA from each culture, and performing restriction digests to verify that the desired construct is present.

Primers designed to specifically target the insert DNA can be used to determine if the construct contains the DNA fragment of interest. Alternatively, primers targeting vector DNA flanking the insert can be used to determine whether or not the insert is the correct molecular size. Insert specific primers can provide information on both the specificity and size of the insert DNA while the use of vector specific primers allows screening of multiple constructs simultaneously. Colony PCR can also be used to determine insert orientation. PCR amplification of the plasmid using an insert specific primer paired with a vector specific primer can be designed to produce an amplicon of a specific size only if the insert is in the correct orientation.

Chromosome walking and Chromosome Jumping:

The principle of chromosome walking is that overlapping clones will hybridize to each other, allowing them to be assembled into a contiguous sequence. This can be used to isolate genes whose function is unknown but whose genetic location is known, a technique known as positional cloning.

To begin a chromosome walk, it is necessary to have in hand a genomic clone that is known to lie very close to the suspected location of the target gene. In humans, for example, this could be a restriction fragment length polymorphism that has been genetically mapped to the same region. This clone is then used to screen a genomic library by hybridization, which should reveal any overlapping clones.

These overlapping clones are then isolated, labelled and used in a second round of screening to identify further overlapping clones, and the process is repeated to build up a contiguous map. If the same library is used for each round of screening, previously identified clones can be distinguished from new ones, so that walking back and forth along the same section of DNA is prevented. Furthermore, modern vectors, such as λDASH and λFIX, allow probes to be generated from the end-points of a given genomic clone by in vitro transcription, which makes it possible to walk specifically in one direction.

In Drosophila, the progress of a walk can also be monitored by using such probes for in situ hybridization against polytene chromosomes. Monitoring is necessary due to the dangers posed by repetitive DNA. Certain DNA sequences are highly repetitive and are dispersed throughout the genome. Hybridization with such a sequence could disrupt the orderly progress of a walk, in the worst cases causing a ‘warp’ to another chromosome. The probe used for stepping from

one genomic clone to the next must be a unique sequence clone, or a subclone that has been shown to contain only a unique sequence.

Chromosome walking is simple in principle, but technically demanding. For large distances, it is advisable to use libraries based on high-capacity vectors, such as BACs and YACs, to reduce the number of steps involved. Before such libraries were available, some ingenious strategies were used to reduce the number of steps needed in a walk. In one of the first applications of this technology, Hogness and his co-workers cloned DNA from the Ace and rosy loci and the homoeotic Bithorax gene complex in Drosophila. The number of steps was minimized by exploiting the numerous strains carrying well-characterized inversions and translocations of specific chromosome regions. A different strategy, called **chromosome jumping**, has been used for human DNA. This involves the circularization of very large genomic fragments generated by digestion with endonucleases, such as NotI, which cut at very rare target sites. This is followed by subcloning of the region covering the closure of the fragment, thus bringing together sequences that were located a considerable distance apart. In this way a jumping library is constructed, which can be used for long-distance chromosome walks.