CONSTRUCTION OF GENOMIC LIBRARY

Sources:

1. Biotechnology By B.D. Singh,
2. Gene cloning and DNA analysis By TA Brown,
3. Advance and Applied Biotechnology By Marian Patrie,
4. Molecular Biotechnology By Glick et al and
5. Principle of Gene Manipulation By sandy b Primrose et al.

Dr. Diptendu Sarkar
diptendu81@gmail.com
Introduction

• A genomic library is an organism specific collection of DNA, covering the entire genome of an organism.
• A genomic library is a set of recombinant clones that contains all of the DNA present in an individual organism.
• Genomic libraries can be retained for many years, and propagated so that copies can be sent from one research group to another.
• It contains all DNA sequences such as:
  1. expressed genes,
  2. non-expressed genes,
  3. exons and introns,
  4. promoter and terminator regions and
  5. intervening DNA sequences.
Construction of genomic library

- Construction of a genomic DNA library involves
  - isolation,
  - purification and fragmentation of genomic DNA
  - followed by cloning of the fragmented DNA using suitable vectors.
- The eukaryotic cell nuclei are purified by digestion with protease and organic (phenol-chloroform) extraction.
- The derived genomic DNA is too large to incorporate into a vector, and needs to be broken up into desirable fragment sizes.
- Fragmentation of DNA can be achieved by physical method and enzymatic method.
- The library created contains representative copies of all DNA fragments present within the genome.
Mechanisms for cleaving DNA

(a) Physical method

It involves mechanical shearing of genomic DNA using a narrow-gauge syringe needle or sonication to break up the DNA into suitable size fragments that can be cloned. Typically, an average DNA fragment size of about 20 kb is desirable for cloning into λ based vectors. DNA fragmentation is random, which may result in variable sized DNA fragments. This method requires large quantities of DNA.

(b) Enzymatic method

- It involves use of restriction enzyme for the fragmentation of purified DNA.
- This method is limited by distribution probability of site prone to the action of restriction enzymes which will generate shorter DNA fragments than the desired size.
- If, a gene to be cloned contains multiple recognition sites for a particular restriction enzyme, the complete digestion will generate fragments that are generally too small to clone. As a consequence, the gene may not be represented within a library.
To overcome this problem, **partial digestion** of the DNA molecule is usually carried out using **known quantity of restriction enzyme** to obtain fragments of ideal size.

The two factors which govern the selection of the restriction enzymes are- **type of ends** (blunt or sticky) generated by the enzyme action and **susceptibility of the enzyme to chemical modification of bases like methylation** which can inhibit the enzyme activity.

The fragments of desired size can be **recovered by either agarose gel electrophoresis or sucrose gradient technique** and ligated to suitable vectors.

The complete (a) and partial (b) digestion of a DNA fragment using restriction enzymes.
Partial restriction digestion is achieved using restriction enzymes that produce blunt or sticky ends as described below-

i. Restriction enzymes generating blunt ends

The genomic DNA can be digested using restriction enzymes that generate blunt ends e.g. HaeIII and AluI.

Blunt ends are converted into sticky ends prior to cloning. These blunt ended DNA fragments can be ligated to oligonucleotides, that contain the recognition sequence for a restriction enzyme, called linkers or possess an overhanging sticky end for cloning into particular restriction sites called adaptors.

\[
\begin{align*}
\text{HaeIII: } & 5' - \text{GG CC-3'} \\
& 3' - \text{CC GG-5'} \\
\text{AluI: } & 5' - \text{AG CT-3'} \\
& 3' - \text{TC GA-5'}
\end{align*}
\]
Linkers

Linkers are short stretches of double stranded DNA of length 8-14 bp that have recognition site for restriction enzymes. Linkers are ligated to blunt end DNA by ligase enzyme. The linker ligation is more efficient as compared to blunt-end ligation of larger molecules. The ligated DNA can be digested with appropriate restriction enzyme generating cohesive ends required for cloning in a vector. The restriction sites for the enzyme used to generate cohesive ends, may be present within the target DNA fragment which may limit their use for cloning.

**Linkers and their use:** (a) the structure of a typical linker; (b) the attachment of linkers to a blunt-ended molecule.
Adapters

These are short stretches of oligonucleotide with cohesive ends or a linker digested with restriction enzymes prior to ligation. Addition of adaptors to the ends of a DNA converts the blunt ends into cohesive ends.

Adaptors and the potential problem with their use.
(a) A typical adaptor. (b) Two adaptors could ligate to one another to produce a molecule similar to a linker, so that (c) after ligation of adaptors a blunt-ended molecule is still blunt-ended and the restriction step is still needed.
The use of adaptors: (a) the actual structure of an adaptor, showing the modified 5′-OH terminus; (b) conversion of blunt ends to sticky ends through the attachment of adaptors. After the adaptors have been attached, the abnormal 5′-OH terminus is converted to the natural 5′-P form by treatment with the enzyme polynucleotide kinase, producing a sticky-ended fragment that can be inserted into an appropriate vector.
ii. Restriction enzymes that generate sticky ends

Genomic DNA can be digested with commonly available restriction enzymes, that generate sticky ends.

For example, digestion of genomic DNA with the restriction enzyme _Sau3AI_ (recognition sequence 5’-GATC-3’) generates DNA fragments, that are compatible with the sticky end produced by _BamHI_ (recognition sequence 5’-GGATCC-3’) cleavage of a vector. Once the DNA fragments are produced, they are cloned into a suitable vector.

**Construction of genomic library**

- **Bacteriophage DNA**
  - Size: 49 kb
  - Replaceable region
  - Cut and remove replaceable region
- **Human DNA**
- **Partial Digestion**
- **DNA with sticky ends**
- **Vector DNA with sticky ends**
- **Mix and seal with DNA ligase**
  - **Recombinant DNA**
- **Package in vitro**
  - **Recombinant phage containing human DNA**
Cloning of genomic DNA

Various vectors are available for cloning large DNA fragments. λ phage, yeast artificial chromosome, bacterial artificial chromosome etc. are considered as suitable vectors for larger DNA and λ replacement vectors like λDASH and EMBL3 are preferred for construction of genomic DNA library. T4 DNA ligase is used to ligate the selected DNA sequence into the vector.

(1) λ replacement vectors

The λEMBL series of vectors are widely used for genomic library construction. The multiple cloning sites of these vectors flanking the stuffer fragment contain opposed promoters for the T3 and T7 RNA polymerases. The restriction digestion of the recombinant vector generates short fragments of insert DNA left attached to these promoters. This generates RNA probes for the ends of the DNA insert. These vectors can be made conveniently, directly from the vector, without recourse to sub-cloning.
1. Preparation of arms and genomic inserts

2. Ligation

3. Packing with a mixture of the phage coat proteins and phage DNA-processing enzymes

4. Infection and formation of plaques

Library constructed
(2) High-capacity vectors

The high capacity cloning vectors used for the construction of genomic libraries are **cosmids**, **bacterial artificial chromosomes (BACs)**, **P1-derived artificial chromosomes (PACs)** and **yeast artificial chromosomes (YACs)**. They are designed to handle longer DNA inserts, much larger than for λ replacement vectors. So they require lower number of recombinants to be screened for identification of a particular gene of interest.

The recombinant vectors and insert combinations are **grown in E. coli** such that a single bacterial colony or viral plaque arises from the ligation of a single genomic DNA fragment into the vector.
### Vectors used for cloning genomic libraries.

<table>
<thead>
<tr>
<th>VECTOR</th>
<th>INSERT SIZE</th>
<th>FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>λ phages</strong></td>
<td>Up to 20-30 kb</td>
<td>Genome size-47 kb, efficient packaging system, replacement vectors usually employed, used to study individual genes.</td>
</tr>
<tr>
<td>2. <strong>Cosmids</strong></td>
<td>Up to 40 kb</td>
<td>Contains cos site of λ phage to allow packaging, propagate in E. coli as plasmids, useful for sub-cloning of DNA inserts from YAC, BAC, PAC etc.</td>
</tr>
<tr>
<td>3. <strong>Fosmids</strong></td>
<td>35-45 kb</td>
<td>Contains F plasmid origin of replication and λcos site, low copy number, stable.</td>
</tr>
<tr>
<td>4. <strong>Bacterial artificial chromosomes (BAC)</strong></td>
<td>Up to 300kb</td>
<td>Based on F- plasmid, relatively large and high capacity vectors.</td>
</tr>
<tr>
<td>5. <strong>P1 artificial chromosomes (PACs)</strong></td>
<td>Up to 300 kb</td>
<td>Derived from DNA of P1 bacteriophage, combines the features of P1 and BACs, used to clone larger genes and in physical mapping, chromosome walking as well as shotgun sequencing of complex genomes.</td>
</tr>
<tr>
<td>6. <strong>Yeast artificial chromosomes (YAC)</strong></td>
<td>Up to 2000kb</td>
<td>Allow identification of successful transformants (BAC clones are highly stable and highly efficient)</td>
</tr>
</tbody>
</table>
Number of clones required for a library

- The number of clones to be pooled depends upon the size of the genome and average size of the cloned DNA.
- Let $f$ be the size of the genome relative to a single cloned fragment. The minimum number of clones required can be calculated as: $f = \frac{\text{genome size}}{\text{fragment size}}$.
- For the *E. coli* genome (4.6 Mb) with an average cloned fragment size of 5 kb, $f$ will be 920. Thus, for the human genome (2.8 $\times 10^6$ kb) and an average cloned fragment size of 20 kb, $f = 1.4 \times 10^5$.
- The number of independent recombinants required in the library must be greater than $f$, as sampling variation leads to the several times inclusion and exclusion of some sequences in a library of just $f$ recombinants.
- In 1976, Clarke and Carbon derived a formula to calculate probability ($P$) of including any DNA sequence in a random library of $N$ independent recombinants. The actual number of clones required can be calculated as: $N = \frac{\ln(1 - P)}{\ln(1 - 1/f)}$; where $N$ = number of clones and $P$ = probability that a given gene will be present.

Therefore, to achieve a 95% probability ($P = 0.95$) of including any particular sequence in a random human genomic DNA library of 20 kb fragment size: $N = \frac{\ln(1 - 0.95)}{\ln(1 - 1/1.4 \times 10^5)} = 4.2 \times 10^5$.
- Notice that a considerably higher number of recombinants is required to achieve a 99% probability, for here $N = 6.5 \times 10^5$.
- Bigger the library better will be the chance of finding the gene of interest.
- The pooling together of either recombinant plaques or bacterial colonies generates a primary library.
Amplified library

• The primary library created is usually of a low titer and unstable. The stability and titer can be increased by amplification. For this, the phages or bacterial colonies are plated out several times and the resulting progenies are collected to form an amplified library.

• The amplified library can then be stored almost indefinitely due to long shelf-life of phages.

• It usually has a much larger volume than the primary library, and consequently may be screened several times.

• It is possible that the amplification process will result in the composition of the amplified library not truly reflecting the primary one.

• Certain DNA sequences may be relatively toxic to E. coli cells. As a consequence bacteria harboring such clones will grow more slowly than other bacteria harboring non-toxic DNA sequences. Such problematic DNA sequences present in the primary library may be lost or under-represented after the growth phase required to produce the amplified library.
Sub genomic library

- **Sub genomic library** is a library which represents only a fraction of the genome.
- Enhancing the fold of purification of target DNA is **crucial for sub genomic DNA libraries** which can be achieved by multiple, sequential digestion when information of the restriction map of the sequences of interest is known.
- After initial purification of a given fragment, the purification can further be increased by redigestion with another enzyme generating a smaller (clonable) fragment relative to original DNA.
Advantages of genomic libraries

• Identification of a clone encoding a particular gene of interest.
• It is useful for prokaryotic organisms having relatively small genomes.
• Genomic libraries from eukaryotic organisms are very important to study the genome sequence of a particular gene, including its regulatory sequences and its pattern of introns and exons.

Disadvantages of genomic library

• Genome libraries from eukaryotes having very large genomes contain a lot of DNA which does not code for proteins and also contain non-coding DNA such as repetitive DNA and regulatory regions which makes them less than ideal.
• Genomic library from a eukaryotic organism will not work if the screening method requires the expression of a gene.
Applications

1. To determine the complete genome sequence of a given organism.
2. To study and generate transgenic animals through genetic engineering, serving as a source of genomic sequence.
3. To study the function of regulatory sequences \textit{in vitro}.
4. To study the genetic mutations.
5. Used for genome mapping, sequencing and the assembly of clone contigs.