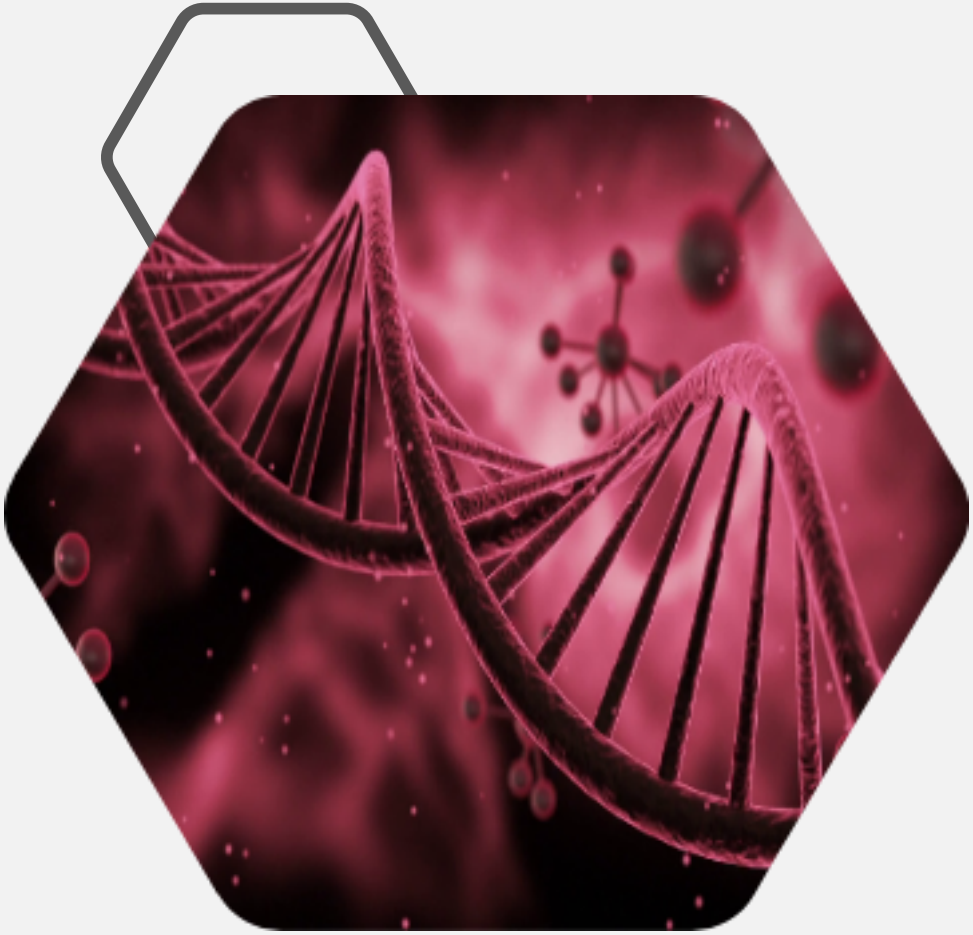
Gene Cloning Strategies and Construction of Libraries

Shreeya Santosh Mhade
M.Sc. Bioinformatics (Part I)
G. N. Khalsa College, Matunga

Contents

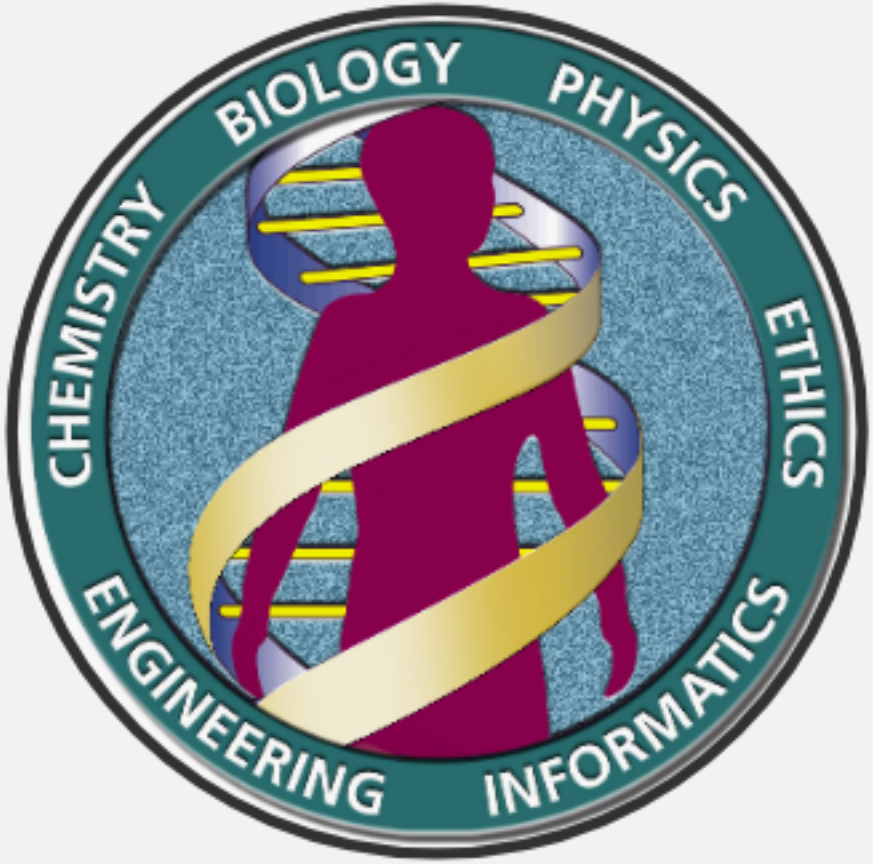


Key Questions to be answered:

- How is DNA cloned?
- What are genomic libraries?
- How can we find a specific gene in a library of cloned DNA?

Introduction

- **GENOMICS** is the science of obtaining and analyzing the sequences of complete genomes.
- At the core of genomics is recombinant DNA technology, the ability to construct and clone individual fragments of a genome, and to manipulate the cloned DNA in various ways, including sequencing it or expressing it in a foreign cell.
- The field of molecular genetics changed radically in the 1970s when procedures were developed that enabled researchers to construct recombinant DNA molecules and to clone (make many copies of) those molecules.
- Using recombinant DNA technology to manipulate genes for genetic analysis or to develop products or other applications is called **genetic engineering.**



DNA Cloning

In brief, DNA is cloned molecularly typically by the following steps:

- Isolate DNA from an organism.
- Cut the DNA into pieces with a restriction enzyme
- Introduce (transform) the recombinant DNA molecules into a host such as E. coli. Replication of the recombinant DNA molecule—the process of molecular cloning—occurs in the host cell, producing many identical copies called clones.

Traditional Cloning

- Traditional cloning relies on recombinant DNA methods that begin with preparing a vector to receive an insert DNA by digesting each with restriction enzymes.
- The digested fragments are then spliced together by an enzyme called ligase, in a process known as ligation, to form a new vector capable of expressing a gene of interest.
- This may be the simplest and oldest technique for traditional cloning and laid the foundation for researchers to develop novel cloning methods such as TA cloning™, TOPO™ cloning, PCR cloning, ligation-independent cloning, and gene assembly that exploit unique characteristics of other modifying enzymes.

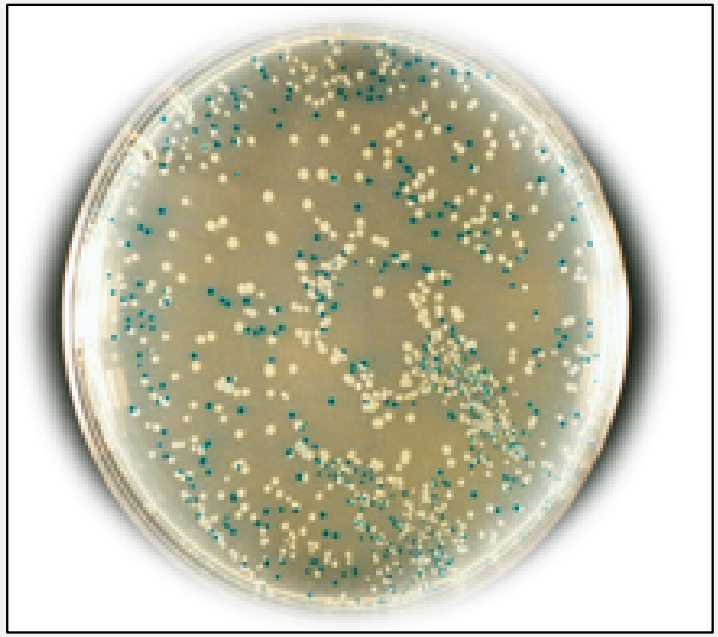


Figure 1. Blue – White Screening

A general workflow for traditional cloning includes the following steps (Figure 2):

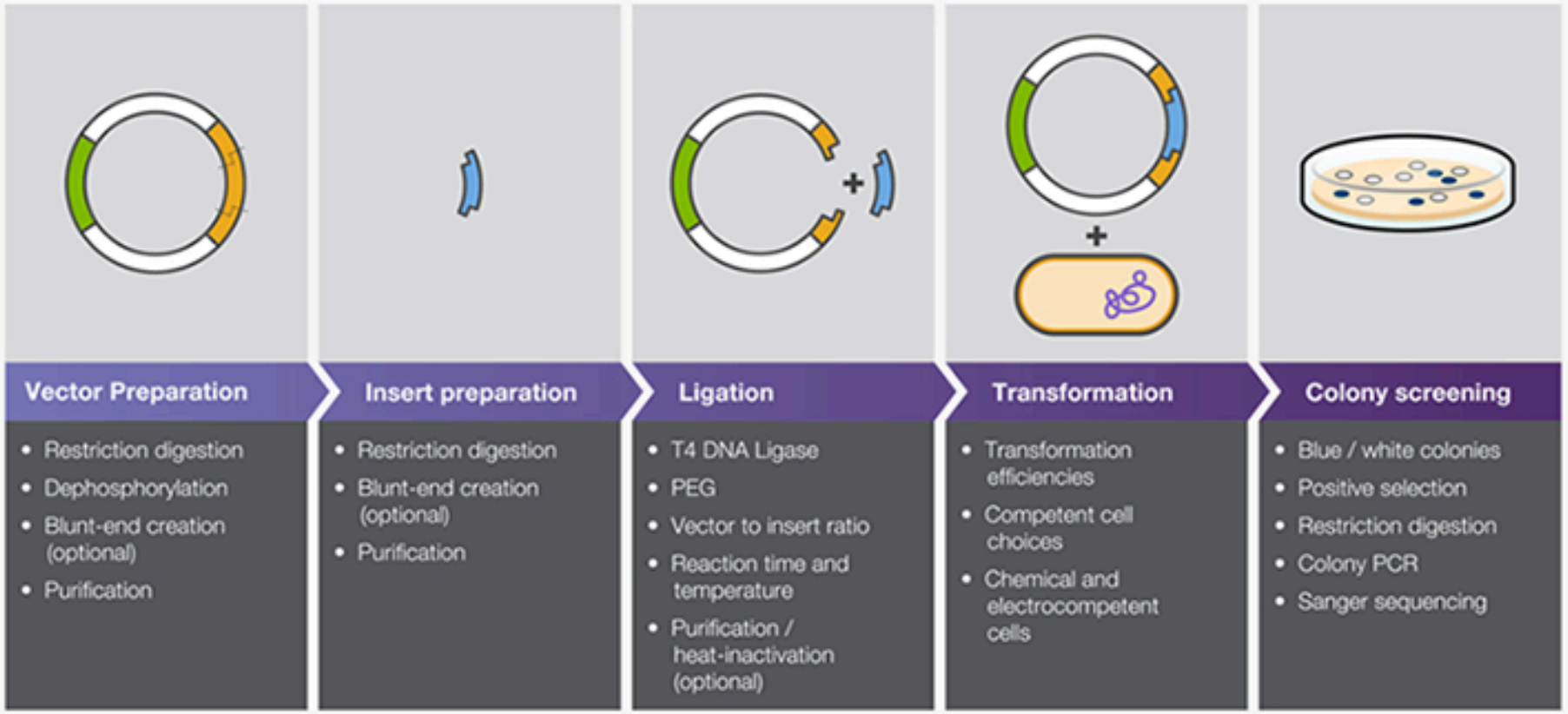


Figure 2. Traditional cloning workflow.

Vector Preparation

- The first step in preparing the vector for traditional cloning is to create an insertion site by restriction digestion.
- The choice of restriction enzymes depends upon the presence and location of their recognition sequences on the vector and the insert, and their compatibility for ligation.
- Vector restriction sites can be found on the vector map, or can be mapped using free online tools such as RestrictionMapper. The MCS, if available, is often the first choice for insertion, as the region is specifically designed for cloning.
- After restriction digestion, de-phosphorylation of the vector may be necessary to prevent self-ligation, especially if the resulting ends of vector digestion are compatible or blunt.

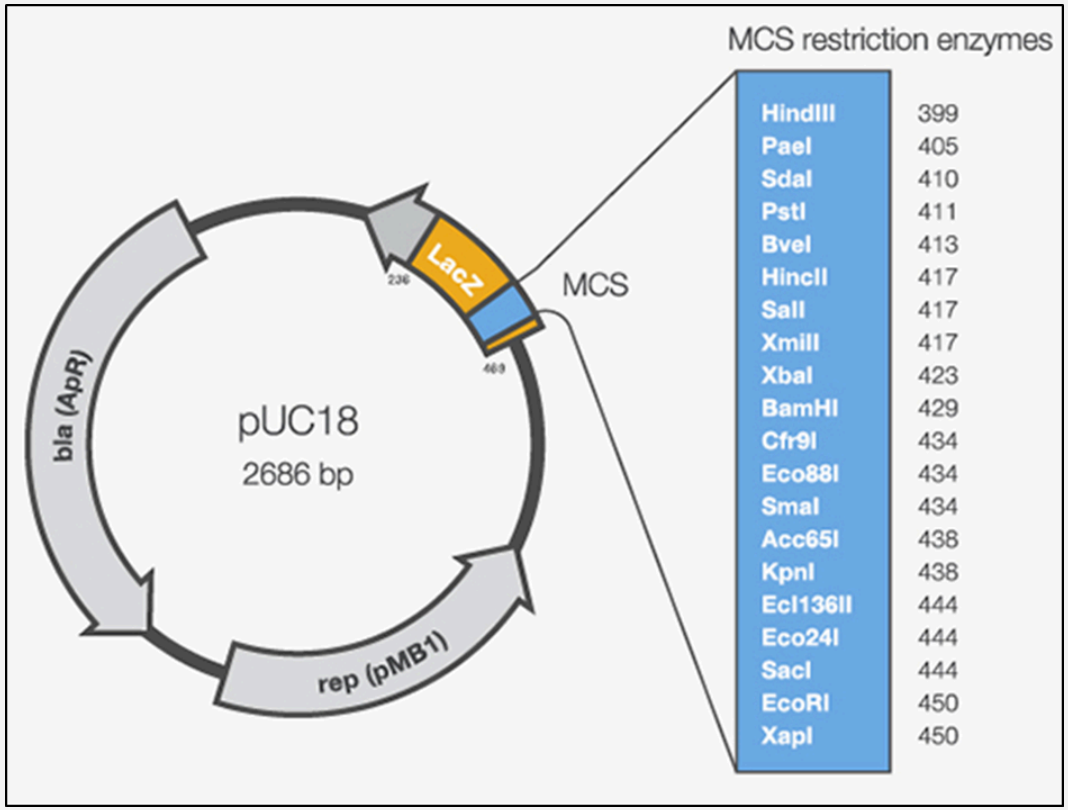


Figure 3. Map of pUC18 with its MCS.

Insert Preparation

- The source of the insert for cloning may be genomic DNA, a portion of another plasmid, or a linear DNA fragment.
- As with vector preparation, restriction enzymes that are suitable for cloning of the insert into the vector are selected (Figure 4). One of the most popular strategies is to perform double digests of both the insert and vector for directional cloning.
- Two enzymes that generate non-compatible ends (EcoRI and KpnI) are used. Since vector and insert ends can join in only one orientation due to compatibility (EcoRI with EcoRI, KpnI with KpnI), this approach allows the insert to be cloned directionally (Figure 4A).
- In instances where suitable restriction enzymes are not available, the DNA ends created by the chosen restriction enzymes may be blunted (or “polished”) for cloning. Blunting will alter the original sequences around the DNA ends; (Figure 4B).
- In some instances, a single restriction enzyme may be chosen that cuts both the insert and vector DNA, generating complementary ends for ligation (Figure 4C). This method is commonly used in genomic DNA cloning.
- In situations when it is not possible to use a single restriction enzyme, a pair of enzymes that have different recognition sequences but generate compatible overhangs can be considered as an alternative. (Figure 4D).

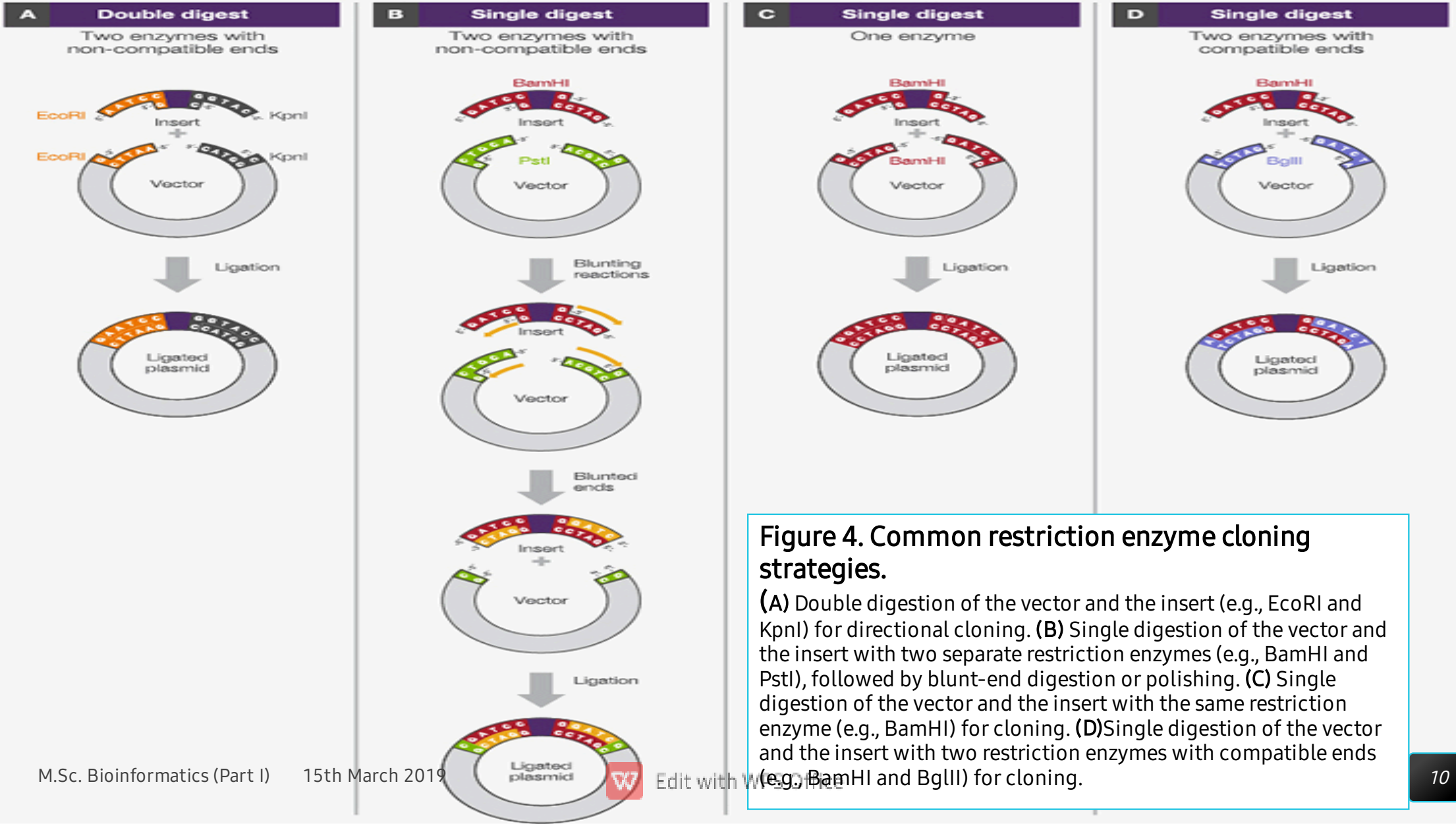


Figure 4. Common restriction enzyme cloning strategies.

(A) Double digestion of the vector and the insert (e.g., EcoRI and KpnI) for directional cloning. **(B)** Single digestion of the vector and the insert with two separate restriction enzymes (e.g., BamHI and PstI), followed by blunt-end digestion or polishing. **(C)** Single digestion of the vector and the insert with the same restriction enzyme (e.g., BamHI) for cloning. **(D)**Single digestion of the vector and the insert with two restriction enzymes with compatible ends (e.g., BamHI and BglII) for cloning.

Purification

- After restriction digestion of the insert and the vector (and subsequent blunting and dephosphorylation, if performed), the desired fragments can be purified by running the samples on an agarose gel and excising the fragments of interest.
- Extracted DNA should be highly pure for successful ligation. The simplest method to assess purity is to measure its absorbance: pure DNA has an A260/A280 ratio of >1.8 and an A260/A230 ratio of approximately 2.0.

Figure 5. Gel Electrophoresis

Ligation

- Once the fragments of interest are obtained, a ligation reaction can be set up to join the insert and the vector.
- The most common enzyme used for ligation is T4 DNA ligase, which links DNA ends between 5′ phosphate and 3′ OH groups.
- The T4 DNA ligase reaction requires ATP, DTT, and Mg^{2+}, which are generally supplied in the reaction buffer **(Figure 6)**.
- To improve the outcome of ligation, a general recommendation is to set up multiple reactions with varying insert:vector molar ratios, typically in the range of 1:1 to 5:1.
- For less efficient ligations, as with DNA fragments with blunt ends, the addition of inert macromolecules like polyethylene glycol (PEG) is often recommended to increase the effective concentration of reaction components and thus improve the ligation efficiency.

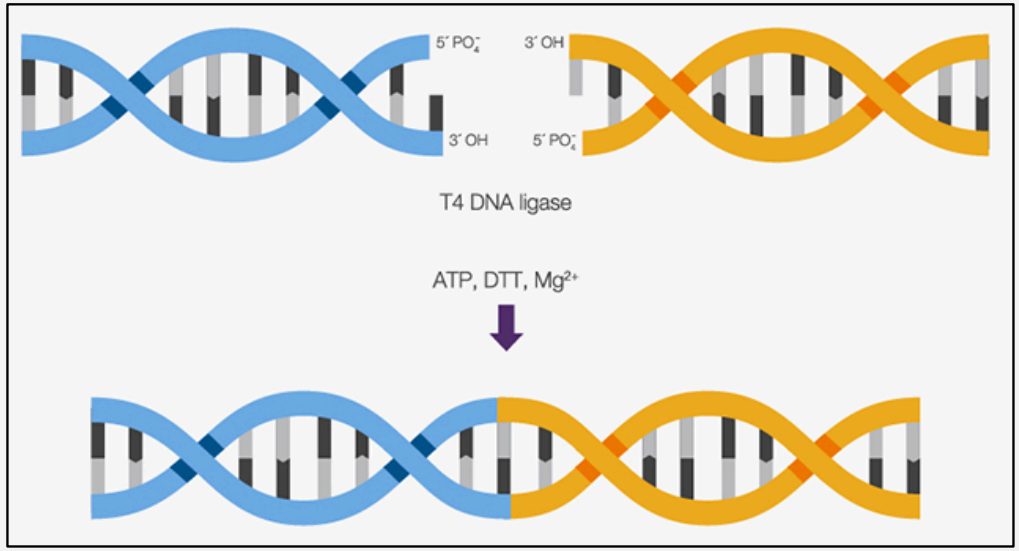


Figure 6. T4 DNA ligase reaction.

Transformation and Colony Screening

- Transformation is a naturally occurring process in which bacterial cells take up foreign DNA at a low frequency.
- The most common approach to prepare bacteria to be competent for transformation is to treat log-phase bacterial cells with calcium chloride.
- Transformed bacteria (after heat shock or electroporation) are then plated on an agar plate with an appropriate antibiotic, and screened (by blue-white screening or another method) for colonies that carry the desired plasmid with insert.
- The transformation reaction contains a mix of cells with no vector, the vector with no insert, the insert alone, and the successfully ligated vector and insert.
- Bacteria without the vector lack the antibiotic resistance gene and will not grow, whereas bacteria transformed with the vector (with or without the insert) survive due to the expressed antibiotic resistance gene **(Figure 7)**. Thus, the antibiotic resistance allows selection for uptake of an intact plasmid.

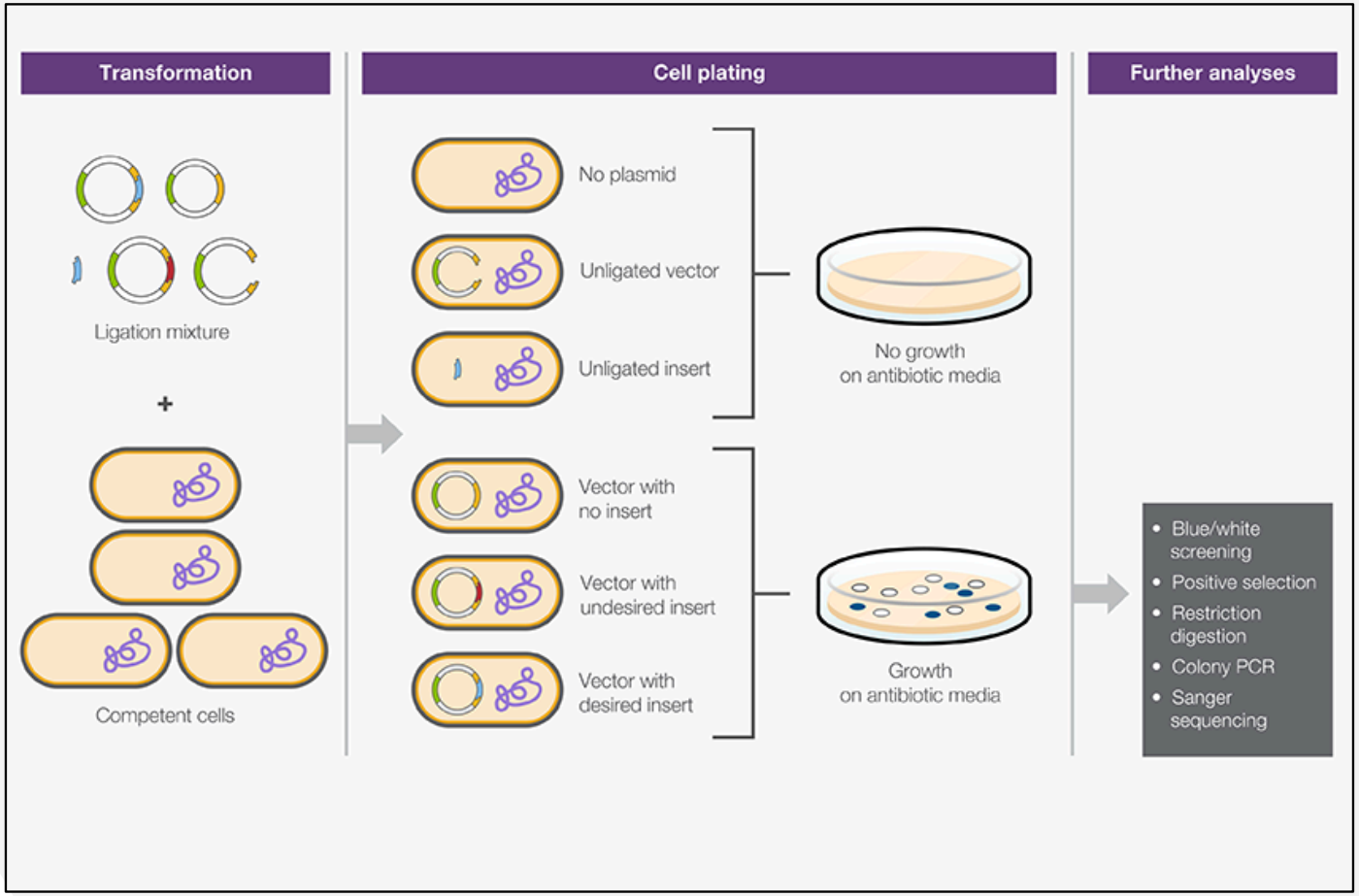


Figure 7. Mixture of bacteria after transformation and their phenotypes (growth/no growth/blue-white) on an antibiotic selection media plate.

Cloning Applications and Methodologies

Cloning methods rely on molecular biological processes that occur in nature. The techniques are continually being refined and simplified; therefore, many strategies nowadays permit cloning of sequences of interest from their sources more efficiently. These cloning strategies include:

1. **PCR cloning strategies**
2. **Subcloning basics**
3. **Library construction essentials**
4. **Shotgun cloning and sequencing method**

- All gene-cloning experiments have four basic steps:
  1. Isolation of a DNA fragment
  2. Joining of the fragment to a cloning vector
  3. Introduction of the cloning vector, along with the inserted DNA fragment, into host cells
  4. Identification of cells containing the recombinant DNA molecule

Cloning Strategies

- In developing a cloning strategy, a number of factors must be taken into consideration.

Table 1. Considerations in developing a cloning strategy[7]

Step in Gene Cloning	Considerations
1. Isolation of DNA fragment	a. The purpose of cloning (is expression required?). Is the entire sequence needed? b. What is known about the gene and the protein (if any) that it encodes? c. The size of the gene. d. Is the chromosomal location of the gene known? e. Size of the genome from which the gene is isolated.
2. Joining DNA fragment to vector	a. Type of cloning vector used.
3. Transfer of recombinant vector to host cell	a. Type of cloning vector used.
4. Identification of cells carrying recombinant molecule	a. Known information about the gene.

PCR Cloning

- PCR cloning is a method in which double-stranded DNA fragments amplified by PCR are ligated directly into a vector. PCR cloning offers some advantages over traditional cloning which relies on digesting double-stranded DNA inserts with restriction enzymes to create compatible ends, purifying and isolating sufficient amounts, and ligating into a similarly treated vector of choice

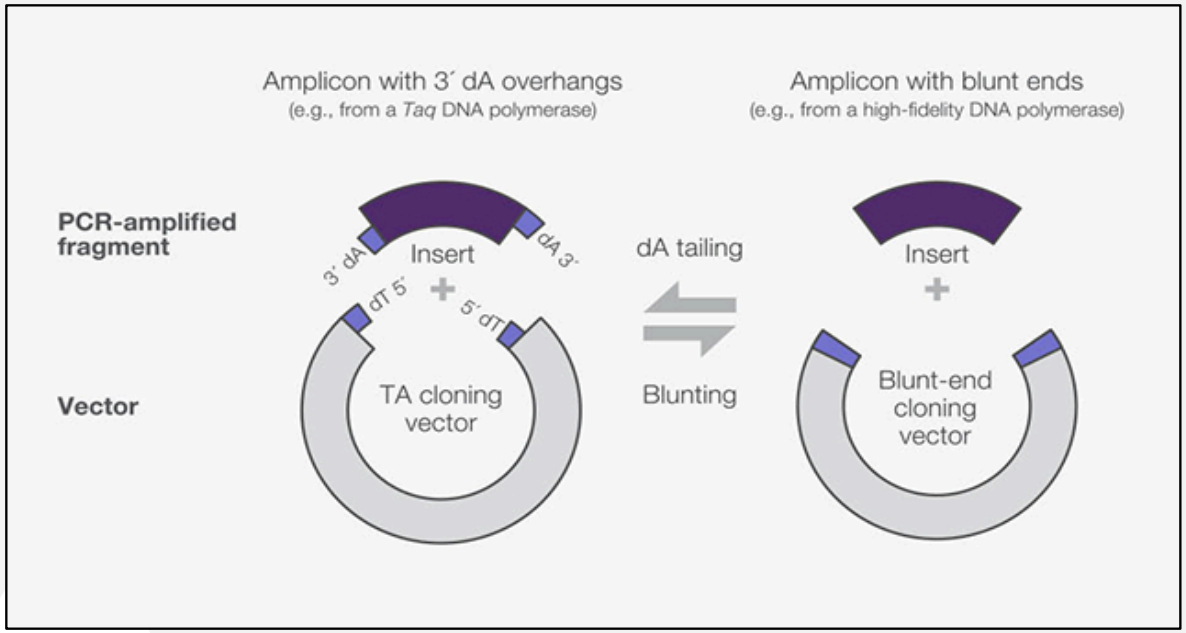


Figure 8. Common PCR cloning strategies.

Advantages:

- Much less starting template materials.
- Simpler workflow

Sub cloning Basics

- Sub cloning refers to moving one fragment of a plasmid into another plasmid that can serve as a vector.
- There are a variety of reasons why it is necessary to transfer the fragment of interest into a different vector backbone.
- For instance, the new vector may possess a specific marker for antibiotic selection or fluorescent expression.
- Sub cloning may also be performed to move a cloned fragment to an expression vector of a more suitable host for the study (e.g., bacteria, mammals, insects, plants, etc.); to place the gene of interest under a different expression promoter (e.g., a constitutive to inducible promoter); or to tag or fuse the experimental gene with another protein or a marker.
- Whatever the goal of the experiment may be, the two most common approaches to sub cloning rely on restriction digestion and/or PCR cloning.
- Sub cloning by restriction digestion is the more traditional of the two methods. In this workflow, fragments from the vector and the insert are double-digested with two restriction enzymes that generate sticky or cohesive ends **(Figure 9)**.

Sub cloning by restriction digestion

Figure 9. Sub cloning by restriction digestion strategies

PCR Sub cloning

Figure 10. **Schematic workflow of PCR subcloning in combination with restriction digestion (RE = restriction enzyme site).**

- A second popular approach uses PCR to amplify the region of interest from the plasmid.
- The resulting PCR product is then cloned into the desired vector. TA cloning or blunt-end cloning methods can be used as described in the PCR Cloning section.
- Following the PCR reaction, PCR products are restriction digested, purified, and sub cloned into the restriction sites of the vector.

Gateway cloning

- Other subcloning strategies have been devised to take advantage of special vectors that do not require the use of restriction enzymes or a ligase.
- One such example is Invitrogen Gateway cloning, which exploits unique recombination activities of the family of Invitrogen Clonase enzymes **(Figure 11)**.
- This method involves use of specially designed Gateway-specific plasmids and Gateway-compatible insert ends (*att* sites) for recombination.

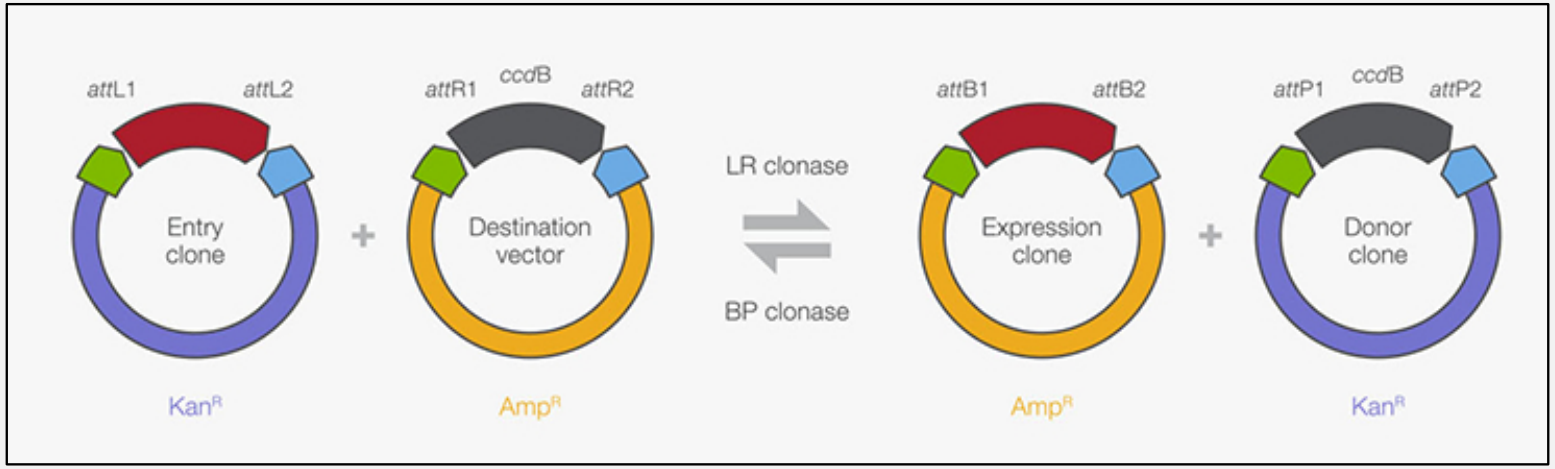


Figure 11. Gateway cloning strategies. *ccdB* is a toxic gene used in bacterial cell selection.

Library construction essentials

- In molecular cloning, DNA library construction refers to the creation of clones that carry DNA fragments representing the complete genomic DNA (gDNA) of a species, or the complementary DNA (cDNA) of RNA transcripts representing the expressed genome.
- By constructing DNA libraries, thousands of genetic fragments can be conveniently archived and expanded for downstream applications, such as genotyping and phenotypic screening.
- gDNA libraries serve as helpful tools to study the genetic composition of different species or gene mutations that occur in diseases such as cancer.
- cDNA libraries, on the other hand, are useful for expression analyses of genes and transcript variants based on the cell type and tissue origins (spatial), as well as time points (temporal).

- The construction of gDNA and cDNA libraries shares many similarities but also some important differences.
- As the starting materials are different between the gDNA library and the cDNA library, their purification and preparation employ different approaches;

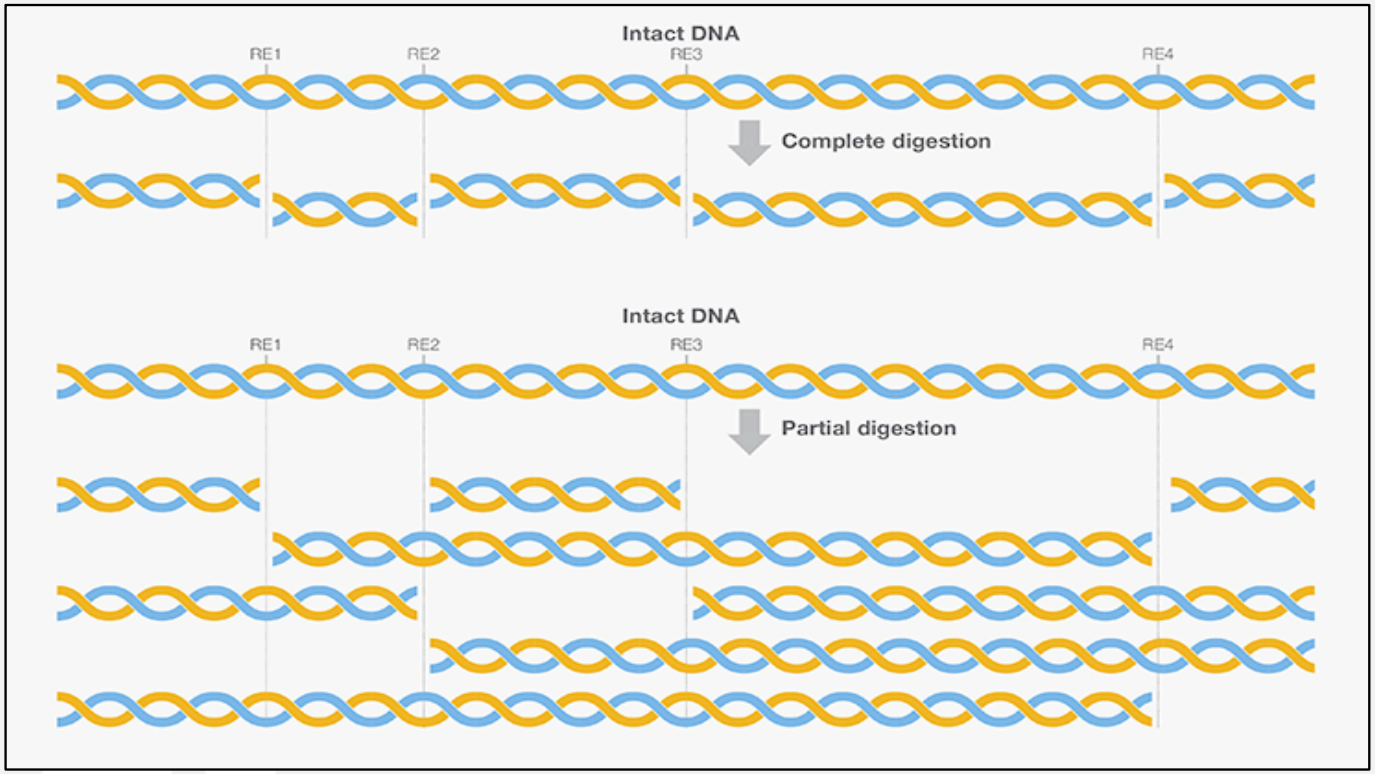


Figure 12. Schematic diagram of complete vs. partial digestion of a fragment by a restriction enzyme with four cutting sites.

- For genomic library preparation, gDNA is purified from the organism, tissues, or cells of interest.
- Extracted gDNA is then digested, isolated, and ligated into the vector of interest with compatible ends.
- Partial digestion of the genome is often carried out with a restriction enzyme with prevalent cutting sites to allow sequence overlaps between fragments for mapping of the cloned inserts **(Figure 12)**.

- Vector selection for gDNA libraries is an important consideration because the gene fragments used in the library constructions are often large (e.g., >20 kb). The choice of cloning vector, in turn, determines the method to deliver insert-carrying vectors into the host **(Table 2)** [1].

Table 2. Common vector types, cloned fragment lengths, and vector delivery methods in library construction.

Vector type	Cloned DNA (kb)	Vector delivery method
Plasmid	20	Transformation
λ phage	25	Transduction
Cosmid	45	Transduction
P1 phage	100	Transduction
BAC (bacterial artificial chromosome)	300	Electroporation
YAC (yeast artificial chromosome)	1,000	Transformation (yeast)

- Ligation products or recombinant DNA can be introduced directly into bacterial cells via transformation or packaged into bacteriophage for infection or “transduction” of the host cells **(Figure 13)**. The transformed or transduced cells are intended for subsequent archiving, expansion, and sequencing in downstream experiments.

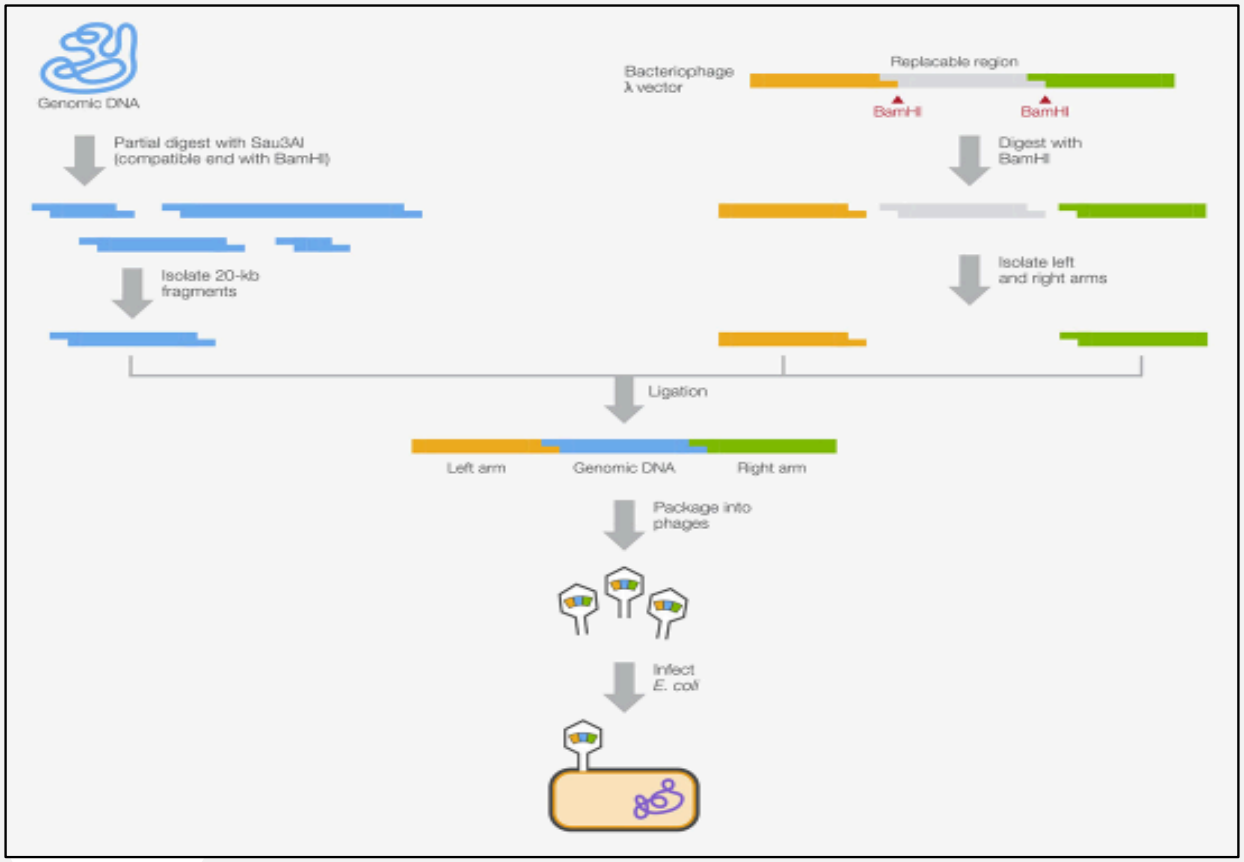


Figure 13. Schematic workflow of genomic library preparation using a λ phage vector. A genomic DNA sample is partially digested with Sau3AI, after which ~20-kb fragments (ideal size for viral packaging) are isolated for ligation with the viral gene fragments. The left and right arms of the λ vector comprise essential components for viral growth in the bacterial cells.

- For cDNA library preparation, total RNA is extracted from a biological source (e.g., cells, tissue, etc.), after which mRNA is reverse transcribed into complementary DNA (cDNA). This process is known as first-strand cDNA synthesis. The second strand is then synthesized to obtain the double-stranded cDNAs. The resulting double-stranded fragments may be ligated directly into a blunt-end cloning vector (random cloning), or "tagged" at the ends with restriction sites for directional cloning **(Figure 14)**.

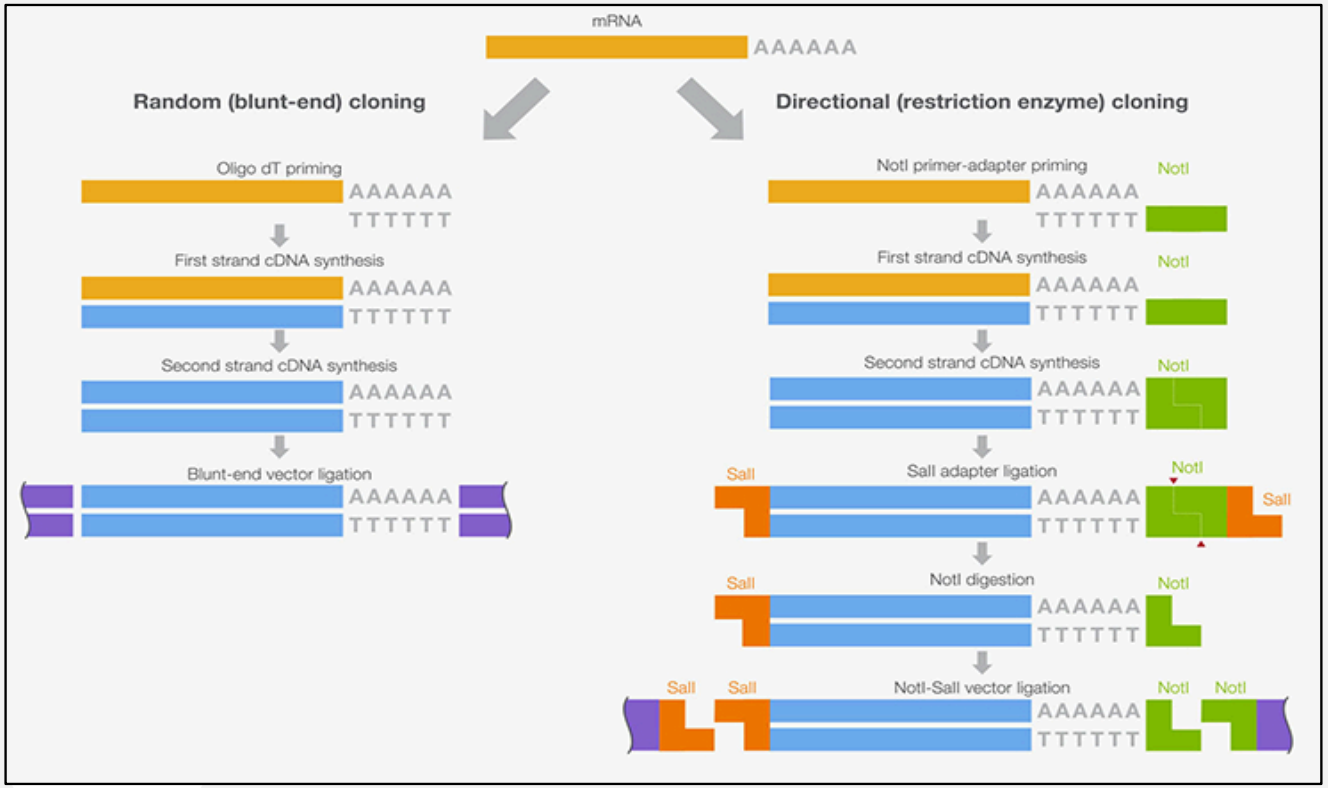


Figure 14. cDNA cloning strategies using mRNA with a poly-A tail. In random (non-directional) cloning, double-stranded cDNA are ligated directly to a blunt-end cloning vector. In directional cloning, adapters with rare restriction sites (e.g., NotI and SalI) are ligated to the double-stranded cDNA ends to clone into a vector with compatible ends.

Shotgun cloning and sequencing method

- Following library construction, one of the goals is to characterize the clones by sequencing the inserts. Insert sizes represented within these libraries can often range from 25 kb to 300 kb, depending on the type of vectors and the genome size of the organism of interest [1].
- For Sanger sequencing, once the most widespread method for DNA sequencing, the upper limit of a sequencing reaction with good-quality reads is generally less than 1 kb.
- To overcome this dilemma, researchers can turn to shotgun cloning and sequencing. In this approach, the large cloned inserts are further fragmented by physical or enzymatic means and sub cloned into another vector; the smaller cloned fragments are then sequenced.
- These sequences are reassembled thereafter based on sequence overlaps (termed contiguous or “contigs”) using bioinformatics programs to ultimately obtain the original long sequence (**Figure 8**).

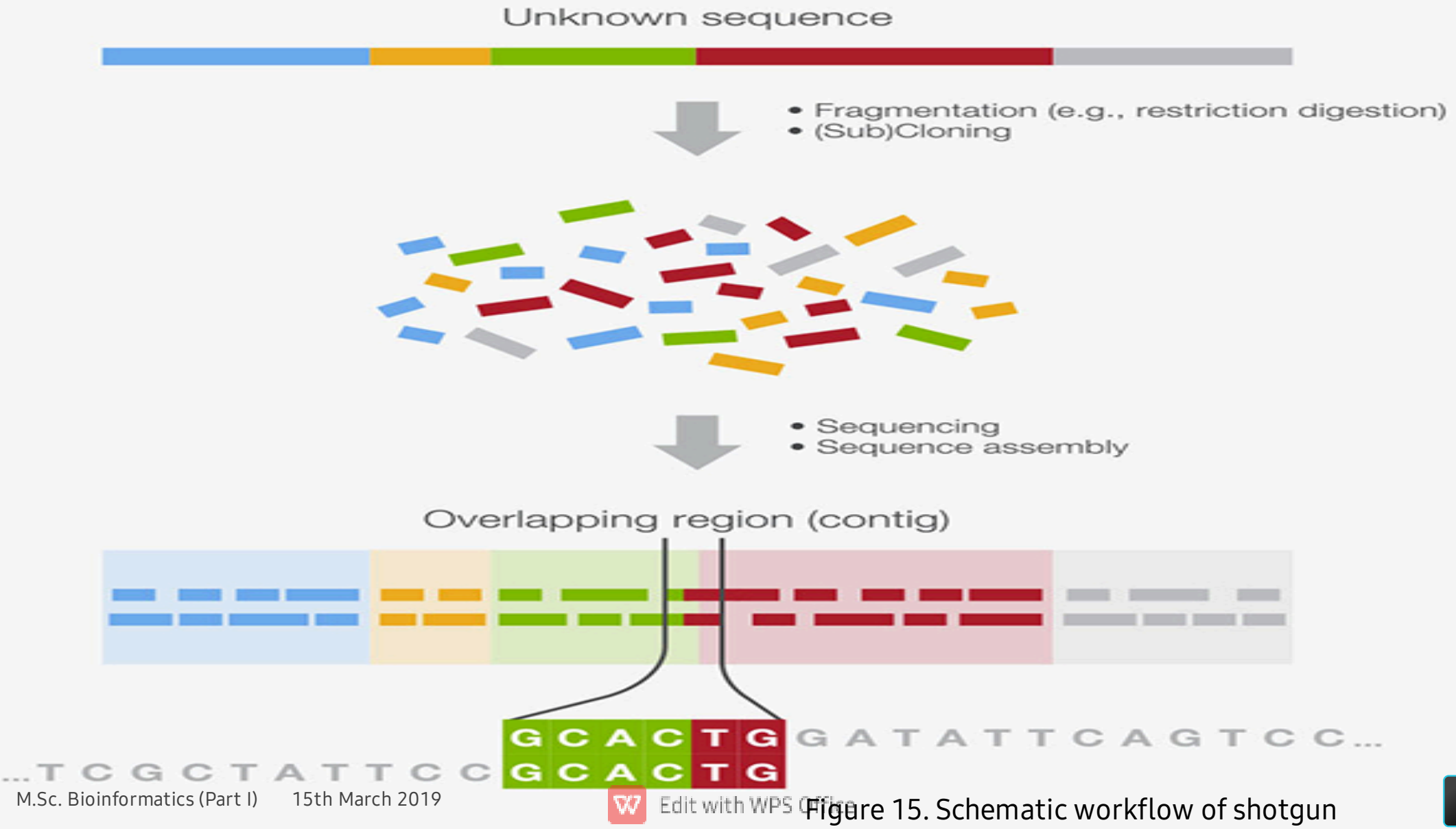


Figure 15. Schematic workflow of shotgun cloning

- Shotgun sequencing is instrumental in whole-genome sequencing of many organisms, ranging from viruses and bacteria to human.
- The method can be used to sequence the genome *de novo*, as well as improve quality of already-sequenced genome by verifying reads and filling in gaps.
- During the first sequencing of the human genome, the publicly funded Human Genome Project employed shotgun sequencing of large gene fragments that had been cloned into a bacterial artificial chromosome or BAC vector.
- The genomic positions of the cloned fragments had been defined prior to shotgun cloning, making their shotgun sequence assembly easier. Hence, this method is known as **hierarchical shotgun sequencing**(**Figure 16A**).
- It is also called **clone-by-clone sequencing** due to the use of BAC clones as a source [3,4].

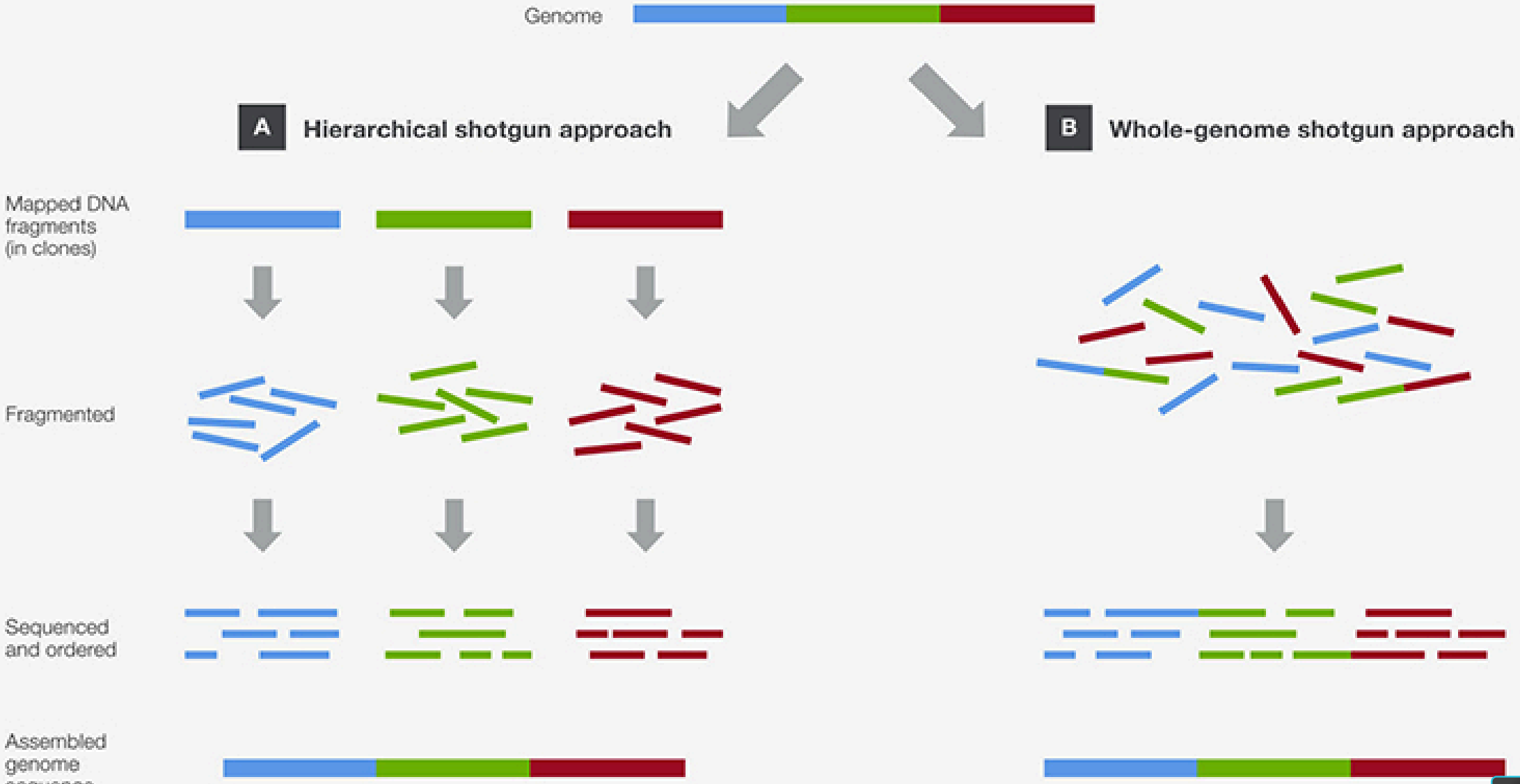


Figure 16. Schematic workflow of two shotgun sequencing approaches used in whole human genome sequencing.

- Concurrent with the Human Genome Project, another privately funded whole genome sequencing project led by Craig Venter used shotgun sequencing strategies directly on the human genome DNA (instead of cloned fragments that had already been mapped). This process is known as the **whole-genome shotgun approach** (**Figure 16B**) [5].
- In theory, shotgun sequencing requires no prior information about the genome or genetic maps, and would save time and resources.
- Nevertheless, it is helpful to have reference genetic maps during sequence assembly because a large amount of computational power is required in the whole-genome shotgun approach, especially for organisms with sizable genomes.
- Genetic mapping or fingerprinting is routinely carried out using restriction enzymes [4], as in the methods of RFLP and AFLP .

References

1. Lodish H, Berk A, Zipursky SL et al. (2000) Section 7.2, Constructing DNA Libraries with λ Phage and Other Cloning Vectors. Molecular Cell Biology. 4th edition. New York: W. H. Freeman.
2. Brown TA (2002) Chapter 6, Sequencing Genomes. Genomes. 2nd edition. Oxford: Wiley-Liss.
3. Lander ES, Linton LM, Birren B et al. (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921.
4. Olson MV (2001) The maps. Clone by clone by clone. Nature 409(6822):816–818.
5. Venter JC, Adams MD, Myers EW et al. (2001) The sequence of the human genome. Science 291(5507):1304–1351.
6. Peter J. Russell, iGenetics - A Molecular Approach, Third Edition. San Francisco: Pearson Benjamin Cummings; 2010.
7. Benjamin A. Pierce, Genetics - A Conceptual Approach, Fourth Edition. New York: W. H. Freeman and Company; 2012.