الجمهوريـــــة الجزائـريـــة الديمـقراطيـــة الشعـبيـــة
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة محمد الصديق بن يحيى– جيجل
Université Mohammed Seddik Benyahia-Jijel

كلية علوم الطبيعة والحياة
قسم بيولوجيا جزيئية و خلوية

Faculté des Sciences de la Nature et de la Vie
Département de Biologie Moléculaire et Cellulaire

Biotechnological Applications of Recombinant DNA

Polycopié de cours

2ème année Master de Biologie Moléculaire et Cellulaire

(UEM5)

Préparé par Dr. Selma HAMIMED

Maitre de Conférences « B »

2025/2026

Ref……

Table of content

Objectives of the course 1

Introduction to the course 1

I - Chapter I: Expression vectors 4

1. Introduction 4

2. Definition of expression vectors 4

3. Types of expression vector 7

3.1. Plasmid 7

3.2. Yeast expression vectors 9

4. Transformation methods 11

4.1. Physical methods 11

4.2. Chemical methods 11

4.3. Biological methods 11

5. Advanced technology in recombinant DNA Crispr (Cas9) 15

5.1. Production of Crispr RNA 15

5.2. Targeting The Crispr RNAs 16

5.3. Restoration a er cleavage 17

6. Selectable and screening markers 19

6.1. The lac operon 19

6.2. The tryptophan operon 22

7. Application of expression vector 25

II - Chapter II: Production of fusion proteins 28

1. Definition of tag protein 28

2. Types of Tag 30

2.1. Polyarginine-tag (Arg-tag) 30

2.2. Polyhistidine-tag(His-tag) 30

2.3. FLAG-tag 30

2.4. Strep-tag 31

2.5. S-tag 31

2.6. Cellulose-binding domain 31

2.7. Maltose-binding protein 31

3. Uses of protein tags 32

3.1. Purification of recombinant proteins 32
3.2. Enhancing the solubility of recombinant proteins 32
3.3. Membrane or lipid-binding fusion proteins 34
3.4. Toxin fusion proteins 34
3.5. Enzyme fusion proteins 34
3.6. Targeted delivery fusion proteins 34
3.7. Fusion tags as reporters 34
3.8. Combinatorial tagging 34
4. Reporter genes 35
4.1. Transcriptional fusion 36
4.2. Translational fusion 36
5. Removal tags 36
5.1. Chemical cleavage 36
5.2. Self -cleaving 37
5.3. Enzymatic cleavage 37
6. Purification 38
III - Chapter III: Biological engineering via directed mutagenesis and agricultural applications 39
1. Overview 39
2. What is mutagenesis? 40
3. Mechanism of mutagenesis 42
3.1. Endogenous mutagenesis 42
3.2. Exogenous mutagenesis 43
4. Mutagen and their types 45
5. Difference between mutation and mutagenesis 47
6. Site-Directed Mutagenesis 47
7. Types of Site-Directed Mutagenesis 49
7.1. Oligonucleotide SDM 49
7.2. Cassette mutagenesis 50
7.3. PCR-SDM 51
7.4. TALENS vs. Zinc Finger Nucleases 52

8. Plant genetic improvement technologies 53
8.1. Introduction 53
8.2. Genetically modified organisms (GMOs) 55
8.3. Application of mutation breeding 56
8.4. Application of genome editing technologies in the improvement of plant 57
IV - Chapter IV: Genetic engineering in medical applications 60
1. Introduction 60
2. Diagnostic tools 60
2.1. Polymerase Chain Reaction 60
2.2. DNA sequencing 64
3. New drugs and new therapies for genetic diseases: case of gene therapy 66
3.1. Types of gene therapy 66
3.2. *In vivo* gene therapy 68
3.3. *Ex vivo* gene therapy 68
3.4. Gene therapy strategies 69
4. New drugs and new therapies for genetic diseases: Case of cell therapy 71
5. Combating infectious diseases 72
5.1. Plasmid profiling 72
5.2. RFLP (Restriction Fragment Length Polymorphism) 73
5.3. AFLP (Amplified Fragment Length Polymorphism) 73
5.4. RAPD (Random Amplified Polymorphic DNA) 73
5.5. Pulsed field gel electrophoresis 74
5.6. Ribotyping 74
6. Genetic fingerprints and its application in forensic medicine 75
6.1. Short Tandem Repeats (STRs) 75
6.2. Single Nucleotide Polymorphisms (SNPs) 75
6.3. The use of mitochondrial genes 76

6.4. Fingerprinting by Y chromosome 77
V - Chapter V: Somatic cloning 78
1. Introduction 78
2. What is Somatic cell cloning? 78
3. Factors affecting somatic cell cloning 79
3.1. Effect of donor age 80
3.2. Effect of cell culture duration 80
3.3. Effect of modification of pre-existing epigenetic marks in donor cells 80
4. Example: Dolly ‘’ same to her mother” 81
5. Therapeutic cloning 81
VI - Chapter VI: Modification of metabolism by genetic engineering 83
1. Introduction 83
2. How to conduct modification of metabolism? 83
2.1. Setting up metabolic pathways 83
2.2. Analyzing a metabolic pathway 84
2.3. Determining the optimal genetic manipulations 84
2.4. Experimental measurements 84
2.5. Resulting approaches 84
3. Case of non-canonical amino acids 84
References 86

Objectives of the course

Recombinant DNA technology has proven important in producing vaccines and protein therapies such as human insulin, interferon and human growth hormone. Overall, this course aims to highlight and provide students with a comprehensive understanding of the techniques, principles, and applications of recombinant DNA technology in genetic engineering to address societal needs, improve human health, enhance food and energy production, and promote sustainable development. By the end of the course, students will be able to:

- Explain the fundamental concepts of recombinant DNA technology, including gene cloning, DNA manipulation, and genetic engineering techniques.
- Understand the significance of recombinant DNA technology in various fields such as medicine, agriculture, industry, and environmental science.
- Analyze and interpret results data generated through recombinant DNA techniques, and critically evaluate the validity and significance of research findings.
- Apply knowledge of recombinant DNA technology to solve practical problems and design experiments in biotechnology research and development.
- Assess current advancements and emerging trends in biotechnological applications of recombinant DNA.
- Communicate effectively about recombinant DNA technology concepts, research findings, and applications through written reports, oral presentations, and scientific discussions.

Introduction to the course

What do you think of when you hear the word “biotechnology”? Maybe things you’ve seen in the news, such as Dolly the cloned sheep, genetically modified organisms, or gene therapy. If that's what you think of, you’re absolutely right: these are all examples of biotechnology. But what about crop breeding and the antibiotic penicillin? These processes and products – some of which have been around for thousands of years – are also examples of biotechnology.

What is biotechnology?

Biotechnology is the use of an organism, or a component of an organism or other biological system, to make a product or process for a specific use. This is a very broad definition, and as mentioned above, it can include both cutting-edge laboratory techniques and traditional agricultural and culinary techniques that have been practiced for hundreds of years (Gupta et al. 2016). Let’s look at three examples of biotechnology and see how they fit the definition:

- Genetically Modified Crops (GMOs), also known as genetically engineered crops, are plants that have had their genetic material altered in a way that does not occur naturally through mating or

natural recombination. This genetic modification is typically done in a laboratory to introduce specific traits or characteristics into the crops. These traits can be derived from other plants, animals, bacteria, or even synthetic sources.

- Penicillin: The antibiotic penicillin is generated by certain molds. To make small amounts of penicillin for use in early clinical trials, researchers had to grow up to 500 liters of "mold juice" a week. The process has since been improved for industrial production, with use of higher-producing mold strains and better culture conditions to increase yield. Here, we see an organism (mold) being used to make a product for human use – in this case, an antibiotic to treat bacterial infections.
- Gene therapy: Gene therapy is an emerging technique used to treat genetic disorders that are caused by a nonfunctional gene. It works by delivering the "missing" gene's DNA to the cells of the body. For instance, in the genetic disorder cystic fibrosis, people lack function of a gene for a chloride channel produced in the lungs. In a recent gene therapy clinical trial, a copy of the functional gene was inserted into a circular DNA molecule called a plasmid and delivered to patients' lung cells in spheres of membrane (in the form of a spray).

What is DNA technology?

Many examples of modern biotechnology depend on the ability to analyze, manipulate, and cut and paste pieces of DNA. Approaches for the sequencing and manipulation of DNA are sometimes referred to as DNA technology. For example, for the cystic fibrosis gene therapy trial, researchers used DNA manipulation techniques to insert the chloride channel gene into a piece of carrier DNA (a vector) that allowed it to be expressed in human lung cells.

DNA technology is important to both basic and applied (practical) biology. For instance, a technique used to make many copies of a DNA sequence, called2 polymerase chain reaction (PCR) (Morrow, 1979), is used in many medical diagnostic tests and forensics applications as well as in basic laboratory research.

What is recombinant DNA?

Recombinant DNA is a form of artificial DNA that is made through the combination or insertion of one or more DNA strands, therefore combining DNA sequences as per your requirement, within different species, as shown in the Figure 1.

Recombinant DNA, also known as genetic engineering, is the process of combining DNA from different organisms to create a new and potentially useful organism (Wilson et al. 2017). Here are the general steps for recombinant DNA in Figure 2:

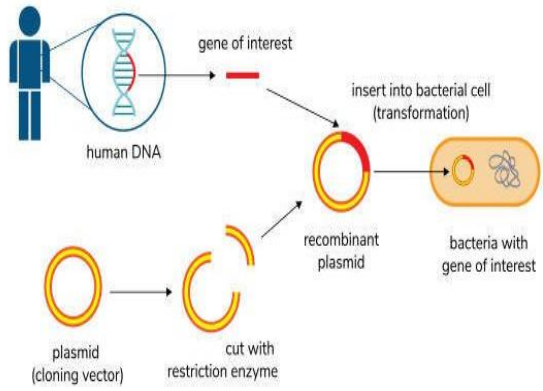

Figure 1. Illustrative image of Recombinant DNA technology.(Nivedita Dar, 2024).

Advantages of recombinant technology:

- Provide substantial quantity;
- No need for natural or organic factors;
- Tailor made product that you can control;
- Unlimited utilizations;
- Cheap;
- Resistant to natural inhibitors.

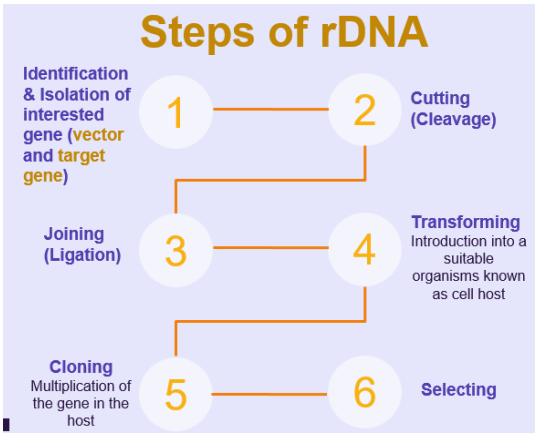

Figure 2. Overall steps for recombinant DNA. Figure created by the author.

Figure 3 displays all over the graphical structure pathway for our course and visualize the different chapters that we will be handled.

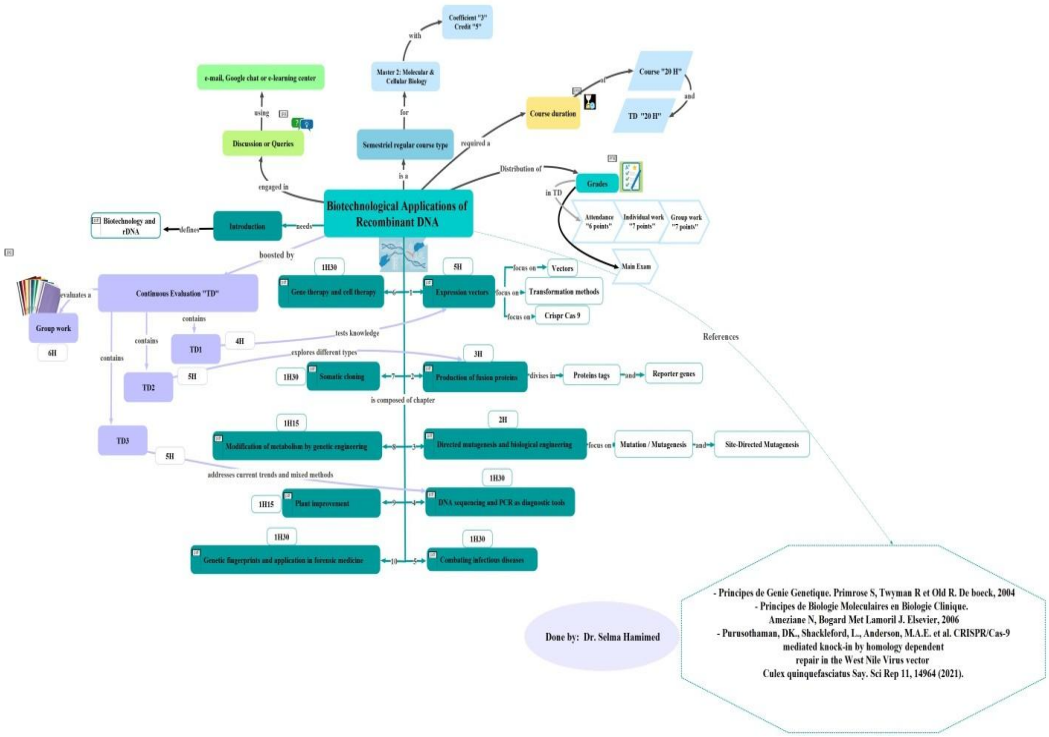

Figure 3. Graphical mental map for the structure of the course. Figure created by the author.

Chapter I: Expression vectors

1. Introduction

Researchers typically use gene expression vectors to purify the resulting protein. Expression vectors are designed with specific properties to control protein production from the cloned gene. Without this control, the gene product could severely damage the host cell, even leading to cell death.

What is a vector?

A vector is a small piece of DNA molecule, which is used as a vehicle to artificially carry foreign genetic material into host cell, where it can be replicated and/or expressed. A vector containing foreign DNA is termed as recombinant DNA. The purpose of a vector is typically to isolate, multiply, or express the insert in the target cell. Useful vectors have three main features: (i) An origin of replication (Ori), which is a specific DNA sequence at which DNA replication is initiated. The essential feature of a vector is that it can replicate autonomously in a host species, usually bacteria. The origin is therefore absolutely essential for the amplification of the vector inside a bacterial host. (ii) Presence of a multiple cloning site (MCS), a region with multiple useful restriction enzyme sites to make a compatible digest of the vector and DNA fragment. (iii) A selectable marker, which is a method of allowing hosts (usually bacteria) containing the vector to be readily identified and purified. The insert contains a selectable marker which allows for identification of recombinant molecules. An antibiotic marker is often used so that a host cell without a vector dies when exposed to a certain antibiotic, and the host with the vector will live because it is resistant. Cloning vector and expression vector are two types of vectors, used in recombinant DNA technology. A cloning vector is a small piece of DNA which can be stably maintained within a host cell, and used to introduce genes into cells while obtaining numerous copies of the insert. A vector designed specifically for the transcription and protein expression of the transgene in the target cell is called expression vector, and generally have a promoter sequence that drives expression of the transgene (Bertero et al. 2017).

2. Definition of expression vectors

Cloning vectors provide a backbone for the DNA insert to be reproduced and propagated in bacteria; however, these vectors are only useful for storing a genetic sequence. By themselves, they are incapable of allowing for transcription and translation of the gene into a functional protein product. The expression vectors are vectors, which act as vehicles for DNA insert and also allow the DNA insert to be expressed efficiently. They are special types of cloning vectors containing the regulatory sequences necessary to allow the transcription and translation of a cloned gene or genes. These vectors contain several key elements (Chiew et al. 2018), including:

- Promoter: This is a DNA sequence that initiates the transcription of the inserted gene

into messenger RNA (mRNA). Different promoters can be used to control the level and timing of gene expression.

- Transcription Stop Site: This is where transcription of the GOI ends. It is signaled by the polyadenylation signal (PolyA), which ensures proper mRNA termination.
- Tags: Help in identifying and purifying the protein of interest once it has been expressed and isolated. Common tags include His-tags (polyhistidine tags), FLAG tags, Myc tags, and GFP (Green Fluorescent Protein) tags.
- Polyadenylation Signal (PolyA): This sequence signals the termination of mRNA transcription. It ensures that the mRNA is properly processed and stable.

According to (Parmar, 2020), vectors are crucial tools in biotechnology for introducing foreign DNA into host cells for various purposes, primarily gene cloning or gene expression. They are classified based on three main criteria:

1. On the basis of our aim with gene of interest

This classification depends on whether the goal is to obtain multiple copies of a gene or its protein product.

- Cloning Vectors:
  - To generate numerous copies (clones) of a gene of interest.
  - Primarily in constructing gene libraries.
  - PUC cloning vectors, pBR322 cloning vectors, bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs).
  - Genetically modified to accommodate foreign DNA, with an insertion site.
- Expression Vectors (or Expression Constructs):
  - To obtain the protein product of a gene of interest by allowing its transcription and translation.
  - Possess additional sequences like bacterial promoters (e.g., lac promoter), ribosome binding sites, and transcription initiation/termination sequences. They may also include sequences for regulating transcription (e.g., regulator genes, operators).
  - Can produce fusion proteins, where the gene of interest is fused with a bacterial polypeptide for stabilization, transport, or purification.

2. On the basis of host cell used

Vectors are designed to be functionally compatible with their specific host cells.

- Vectors for bacteria:
  - Special bacterial origin of replication, antibiotic resistance selectable markers.

- Plasmid vectors (e.g., pBR322, pUC vectors), bacteriophages (e.g., M13, Lambda), cosmids, phasmids, phagemids, fosmids.

- Vectors for yeast:
  - Autonomously replicating sequences (ARS).
  - Yeast replicative plasmid vectors (YRp), Yeast Episomal Plasmids (YEps), Yeast Integrative Plasmids (YIps), Yeast Artificial Chromosomes (YACs).
- Vectors for animals:
  - Synthesize recombinant proteins from genes not expressed correctly in E. coli or yeast, or for gene therapy.
  - Baculovirus vectors, Bovine Papilloma Virus (BPV) vectors, SV40 virus- based vectors.
- Vectors for plants:
  - Used for producing genetically modified plants.
  - Ti-plasmid, Ri-plasmid (from Agrobacterium), plant virus vectors (e.g., *Caulimo virus* vectors, *Gemini* virus vectors).

3. On the basis of cellular nature of host cell

This classification broadly categorizes vectors based on whether the host is prokaryotic or eukaryotic.

- Prokaryotic vectors:
  - Primarily bacterial cells (e.g., *E. coli*).
  - Plasmid-based and bacteriophage-based vectors, including plasmid-derived, bacteriophage-derived, phagemid, plasmid, and cosmid vectors.
- Eukaryotic vectors:
  - Yeast, animal, and plant cells.
  - All vectors designed for yeast, animal, and plant hosts fall under this category.

Shuttle vectors

- Vectors capable of replicating in two different, often unrelated, species (e.g., prokaryotic and eukaryotic hosts).
- Contain two origins of replication, one for each species, and necessary genes for replication not supplied by the host.
- YEp vectors (for yeast and *E. coli*).

Artificial chromosomes

- Synthetically designed DNA molecules acting like natural chromosomes, capable

of stable maintenance in cells. They can clone very large DNA segments.

- Centromere (for cell division distribution), telomeres (for replication and protection of ends), and origin of replication.

  - Types:
    - Yeast Artificial Chromosomes (YACs): Functional artificial chromosomes for yeast, can carry DNA inserts up to 1400 kb. Useful for gene libraries and studying chromosome structure.
    - Bacterial Artificial Chromosomes (BACs): Cloning vectors based on E. coli F factor plasmids, useful for cloning DNA fragments up to 350 kb. Employed in sequencing large chromosomal DNA stretches and modeling genetic diseases.
    - Human Artificial Chromosomes (HACs): Mini-chromosomes artificially constructed in human cells, designed to carry new genes and behave as stable, independent chromosomes.
    - Pl-Derived Artificial Chromosome (PAC): DNA constructs from P1 bacteriophage, carrying 100-300 kilobases of other sequences for bio-engineering.

3. Types of expression vector

The choice of classic cloning vectors depends on size of the insert and application (Chiew et al. 2018).

3.1. Plasmid

Plasmids are extra chromosomal double-stranded DNA sequences found in the cytoplasm of microbes and capable of replication using the host cell's replication machinery. They are generally circular and stable genetic entity that can replicate itself autonomously, independent of the chromosomal DNA of the host organism. They are found widely in many bacteria, for example in *E. coli*, but may also be found in a few eukaryotes, for example in yeast such as *Saccharomyces cerevisiae*. Plasmids usually carry at least one gene, and many of the genes that plasmids carry are beneficial to their host organisms. Although they have separate genes from their hosts, they are not considered to be independent life. Plasmids contain genes that enhance the survival of an organism, either by killing other organisms or by defending the host cell by producing toxins. Multiple plasmids can coexist in the same cell, each with different functions. In other words, plasmids can only co-occur in a bacterium if they are compatible with each other. They are incompatible if they have the same reproduction strategy in the cell. An incompatible plasmid will be expelled from the bacterial cell (Mergeay et al. 2018).

Moreover, plasmids can be eliminated from bacterial cells by a process called curing, which

may take place spontaneously or by various curing treatments, such as acridine dyes, ultraviolet (UV) and ionizing radiation, thymine starvation and growth above optimal temperatures that inhibit plasmid replication but do not affect bacterial chromosome replication and cell reproduction. They have three key points: The origin of replication, which is used to indicate where DNA replication is to begin; The selection marker gene, which is used to distinguish cells containing the plasmid from cells that do not contain it; and the cloning site, a site in the plasmid where the DNA is inserted (Mergeay et al. 2018).

The use of plasmids as an expression vector is ones most commonly used. Most general plasmids may be used to clone DNA insert from 5 to 15 kb in size. One of the earliest commonly used expression vectors is the pBR322 plasmid. Many plasmids have high copy number, for example pUC19 which has a copy number of 500-700 copies per cell and high copy number is useful as it produces greater yield of recombinant plasmid for subsequent manipulation. However low-copy number plasmids may be preferably used in certain circumstances, for example, when the protein from the cloned gene is toxic to the cells.

A plasmid is isolated from the bacterial cell at one site by restriction enzyme. The cleavage converts the circular plasmid DNA into a linear DNA molecule. Then the two open ends of linear plasmid are joined to the ends of the foreign DNA to be inserted with the help of enzyme DNA ligase. This regenerates a circular hybrid or chimeric plasmid, which is transferred to a bacterium wherein it replicates and perpetuates indefinitely. Plasmids often confer antibiotic resistance to the bacteria, so only bacteria containing the vector will survive treatment with antibiotics (Mergeay et al. 2018). Antibiotic resistance is often used as marker, an example is the beta-lactamase gene which confers resistance to the penicillin group of beta-lactam antibiotics like ampicillin. The plasmid vectors also possess promoter sequences near the MCS for efficient expression of gene of interest. The most widely used plasmid in E. coli expression systems are presented in the table 1.

Table 1. *E. coli* expression plasmids. Table created by the author.

Plasmid Name	Promoter	Origin of Replication	Replicon Type	Copy Number
pUC19	Lac promoter (lacUV5)	pMB1 (ColE1-like) origin	Theta replication	High (10-100s)
pBR322	Tet promoter (Tn5)	pMB1 (ColE1-like) origin	Theta replication	Medium (10-20)
pET28a	T7 promoter	p15A origin	Rolling-circle	Low (1-2)
pBAD24	AraC-PBAD promoter	pSC101 origin	Rolling-circle	Low (1-2)
pACYC-Duet1	T5 promoter	p15A origin	Rolling-circle	Low (1-2)

pGEX-6P-1	Tac promoter	pMB1 (ColE1-like) origin	Theta replication	Medium (10-20)
pET21a	T7 promoter	pBR322 origin	Theta replication	High (10-100s)
pRSFDuet-1	T7lac promoter (T7/lacO)	p15A origin	Rolling-circle	Low (1-2)
pMAL-c2X	MalE promoter	pBR322 origin	Theta replication	High (10-100s)

Remarque

The reason for *E. coli* as a promising host cell for its rapid growth rate. Under optimal conditions, it can double in number every 20 to 30 minutes. Its fast growth and well-characterized makes it suitable for large-scale protein production.

3.2. Yeast expression vectors

Yeasts, eukaryotic unicellular fungi, contribute a great deal to the study of molecular genetics. They are popular organisms to clone and express DNA in because they are eukaryotes, and can therefore splice out introns. A species of yeast, Saccharomyces cerevisiae has been an important model system for biological research because its entire genome has been base sequenced. It is used as a reference to human and other higher eukaryotic genes. This is because the basic cellular mechanics of replication, recombination, cell division and metabolism are generally conserved between yeast and larger eukaryotes, including mammals. Three expression vectors used for the production of intra-extra proteins: Yeast episomal plasmids (YEps), Yeast integrating plasmids (YIps), Yeast artificial chromosomes (YACs), Yeast replicative plasmids (YRps), Yeast centromere plasmids (YCps), and Yeast linear plasmid (YLps).

Example of *Pichia pastoris* Expression Systems

The expression vectors for Pichia pastoris contain the following elements, which is presented in Figure 4:

- Strong promoter: such as methanol inducible AOX1 promoter, or strong constitutive promoters (such as GAP1).
- Signal peptide to direct protein secretion: such as α-factor signal peptide.
- Homologous sequences for chromosomal integration at the AOX1 locus.
- A double recombination happened between the AOX1p and AOX1 regions of the vector where the homologous segments DNA results in the insertion of GOI.

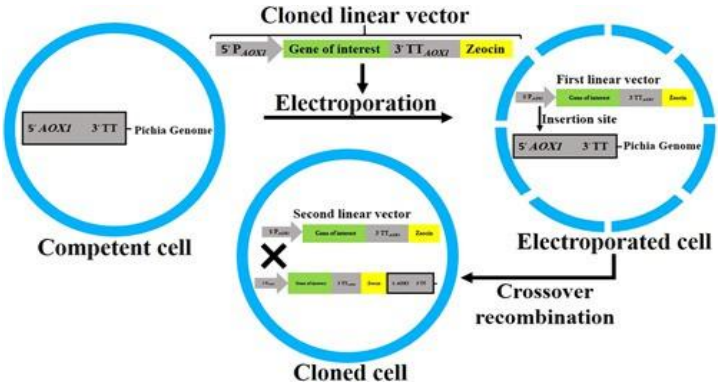

Figure 4. *Pichia pastoris* expression vector (Karbalaei et al. 2020).

Though *S. cerevisae* is successfully used to produce recombinant proteins for human, it has major drawbacks.

- The level of protein production is low.
- There is the tendency for hyper glycosylation resulting in change of protein function.
- Proteins are often retained in periplasm, increasing time and cost for purification.
- It produces ethanol at high cell densities, which is toxic to cells.
- *P. pastoris* is a methylotrophic yeast that is able to utilize methanol as a source of carbon and energy.
- Glycosylation occurs to a lesser extent and the linkages between sugar residues are of the α-1,2type.
- *P. pastoris* strain was extensively engineered with the aim of developing a "humanized" strain that glycosylate proteins in a manner identical to that of human cells.

Take home message

- A YAC is designed to clone a large segment of DNA (100kb), which is then maintained as
- separate chromosome in the host yeast cell.
- Yeps are small (2µm), circular pieces of DNA that exist independently of the yeast cell's chromosomal DNA.
- A Yip cannot replicate as a plasmid as it does not contain the 2 µm of plasmid.

- YRps are able to multiply as independent plasmids due to the presence of an origin of replication.

4. Transformation methods

How the transformation works?

The process of transferring the recombinant vector into cells usually referred by three different terminologies namely transformation, transfection and transduction. From these words, transformation commonly used.

- Transformation: The introduction of any DNA molecule into any living cell.
- Transfection: The introduction of purified phage DNA molecules into a bacterial cell. Transduction: The movement of genes from a bacterial donor to a bacterial recipient with the use of a phage as the vector.

Recombinant DNA technology involves the manipulation of DNA molecules to introduce foreign DNA into host organisms, enabling the expression of specific genes or the production of desired proteins.

Genetic transformation is the process of introducing new genes (foreign DNA) from different species (heterologous sources like other plants, microbes, or animals) into crop plants. This allows for the selective addition of desirable traits that might not be achievable through traditional breeding methods, thereby expanding the plant's genetic pool. The goal is to enhance crop quality, quantity, and resilience to stresses.

The specific method used will depend on the nature of the DNA being introduced and the properties of the host cell. Herein, we describe the most used methods as follow:

4.1. Physical methods

- Electroporation
- Microinjection
- Liposome mediated gene transfer
- Silicon carbide fibre mediated gene transfer
- Ultrasound mediated gene transfer

4.2. Chemical methods

- PEG mediated gene transfer / Calcium chloride
- Dextran mediated gene transfer / Lithium acetate

4.3. Biological methods

- Lentivirus

- Adenovirus
- Liposome

➢ **Key steps of genetic transformation (Experimental procedure)**

The development of transgenic plants is a multi-step process (Sharma et al. 2005):

A. Essential groundwork:

1. Reliable tissue culture regeneration systems:
    - The ability to regenerate a whole plant from transformed cells is crucial. This relies on the totipotency of plant cells (their ability to develop into a whole organism).
    - Methods include organogenesis (shoot formation) or somatic embryogenesis from cultured tissues.
2. Preparation of gene constructs (transformation vectors):
    - Vectors: These are DNA molecules (often plasmids) used to carry the foreign gene(s) into the plant cell.
    - Marker genes: Vectors usually include:
        - Selectable markers: Genes that confer resistance to antibiotics (e.g., kanamycin, hygromycin) or herbicides. These allow researchers to identify and select cells that have successfully incorporated the new DNA, as only transformed cells will survive on a medium containing the selective agent. Examples include nptII (kanamycin resistance) and hpt (hygromycin resistance).
        - Screenable/reporter markers: Genes that produce an easily detectable product (e.g., color change or fluorescence), like β-glucuronidase (GUS) or green fluorescent protein (GFP). These helps visualize and confirm transformation.
    - Promoters: DNA sequences that control the expression (timing, location, and level) of the introduced gene.

B. Introducing genes into plant cells: Transformation techniques

Several methods are used to deliver the gene construct into plant cells:

1. Agrobacterium-mediated gene transfer:
    - Utilizes the natural ability of the soil bacterium Agrobacterium tumefaciens (or A*. rhizogenes*) to transfer a segment of its DNA (T-DNA) into the plant genome. Scientists replace the tumor-inducing genes in the T-DNA with the gene(s) of interest.

- Plant tissues (explants) are co-cultivated with *Agrobacterium*. The bacteria infect the plant cells and transfer the T-DNA containing the desired gene.
- Production of inducing compounds (like acetosyringone) by the plant explant to activate Agrobacterium's virulence (vir) genes, and access of the bacteria to competent, regenerable plant cells.
- Often results in stable integration of a low copy number of genes, leading to simple inheritance patterns.
- Variations to enhance efficiency:
    - Sonication-assisted agrobacterium-mediated transformation (SAAT): Brief ultrasound treatment creates micro-wounds, improving Agrobacterium access and transformation efficiency, especially in recalcitrant species.
    - Floral-dip method: Agrobacterium is applied directly to floral tissues, bypassing tissue culture and potentially reducing somaclonal variation. Primarily used in Arabidopsis.
    - Vacuum infiltration: Plant tissues are submerged in an Agrobacterium suspension under vacuum, facilitating bacterial penetration.

2. Biolistics (microprojectile bombardment / Gene Gun):

- Tiny heavy metal particles (gold or tungsten) are coated with the DNA construct. These microprojectiles are then "shot" at high velocity into plant cells or tissues.
- The particles penetrate the cell walls and membranes, delivering the DNA into the cells.
- Can be used for a wide range of species, including those difficult to transform with Agrobacterium (e.g., many monocots). Can target various tissues, including nuclear, mitochondrial, and chloroplast genomes.
- Can lead to multiple or fragmented gene insertions, complex integration patterns, and potential damage to tissues.
- A hybrid approach combining biolistic delivery of DNA that includes T- DNA border sequences and genes for VirD1/VirD2 proteins (from Agrobacterium). This aims for more precise T-DNA-like integration.

3. Direct DNA transfer methods into protoplasts (Plant cells without cell walls):

- Polyethylene Glycol (PEG)-mediated transformation: PEG alters the cell membrane, making it permeable to DNA.

- Electroporation: Short electrical pulses create temporary pores in the protoplast membrane, allowing DNA uptake.
- Microinjection: DNA is directly injected into individual cells or protoplasts using a fine glass micropipette.
- Silicon carbide whiskers (Microfiber Whiskers): Cells are agitated with DNA and sharp, microscopic silicon carbide whiskers, which pierce the cells and facilitate DNA entry.
- Challenge: Regenerating whole plants from protoplasts can be difficult for many species.

4. Chloroplast transformation:

- Foreign genes are integrated into the chloroplast genome instead of the nuclear genome.
- Often uses biolistics or PEG-mediated transformation of protoplasts, with vectors containing sequences homologous to the chloroplast genome for targeted integration.
- Potential for high levels of protein expression, expression of multiple genes as a single unit (polycistronic expression), and reduced gene flow via pollen (as chloroplasts are usually maternally inherited).

C. Post-transformation steps

1. Selection of transformants:

- Transformed cells/tissues are cultured on a selective medium containing the antibiotic or herbicide for which resistance is conferred by the selectable marker gene. Only transformed cells survive and proliferate.

2. Recovery and multiplication of transgenic plants:

- Selected transgenic cells are regenerated into whole plants through tissue culture.

3. Molecular and genetic characterization:

- Techniques like Southern blotting is used to confirm the presence and number of copies of the inserted gene in the plant's genome.
- Northern blotting (for mRNA) and Western blotting (for protein) are used to verify that the introduced gene is being transcribed and translated.
- Transgenic plants are grown, and their progeny are analyzed to study the stability and inheritance pattern (Mendelian segregation) of the transgene.

5. Advanced technology in recombinant DNA Crispr (Cas9)

CRISPR, or Clustered Regularly Interspaced Short Palindromic Repeats, is an integral part of a bacterial defense system that was first identified in E. coli in the late 1980s. By comparative genomic analysis among different species Jensen et al. in 2002 named the repeat sequences as CRISPR and also defined the common characteristics of CRISPR:

i. its location in intergenic region,

ii. contain multiple short direct repeats (21–37 bp) with little sequence variation,

iii. repeats are interspaced with similar sized non-repetitive nonconserved sequences,

iv. a leader sequences of 300–500 bp were located one side of the repeat clusters (Figure 5).

The repeat sequences and leader sequences are conserved among the species. It is also the basis of the CRISPR-Cas9 system. The CRISPR molecule is made up of short palindromic DNA sequences that are repeated along the molecule and are regularly-spaced.

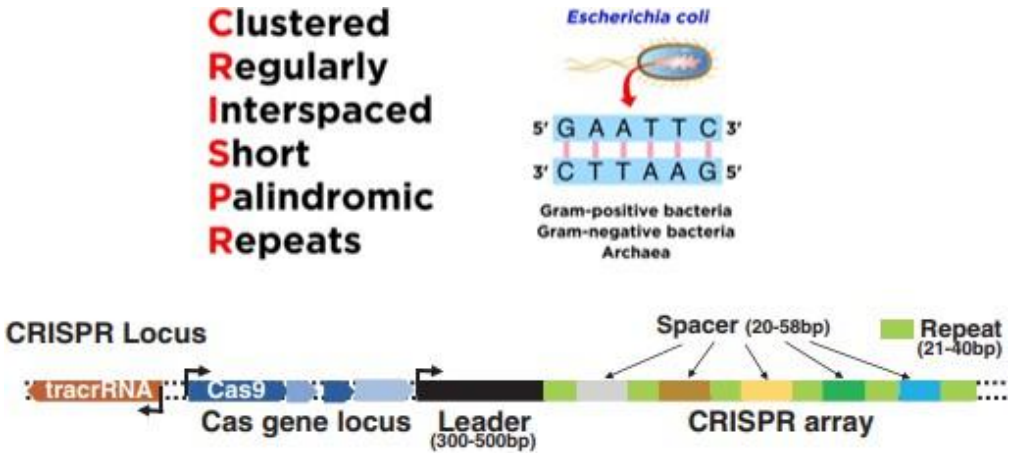

Figure 5. Structure of Crispr region in *E. coli*. Created by the author via BioRender.com.

Between these sequences are “spacers”, foreign DNA sequences from organisms that have previously attacked the bacteria. The CRISPR molecule also includes CRISPR-associated genes, or Cas genes. These encode proteins that unwind DNA, and cut DNA, called helicases and nucleases, respectively. The CRISPR immune system protects the bacteria from repeated virus attacks through three steps (Redman et al. 2016):

5.1. Production of Crispr RNA

When a bacterium survives a viral infection, it captures a small piece of the viral DNA and incorporates it into its own genome, which called adaptation. This process is known as adaptation. The captured viral DNA is stored in the bacterium's genome in the form of short, unique sequences known as CRISPR arrays.

The CRISPR arrays undergoes transcription, including spacers and Cas genes creating precursor CRISPR RNA (pre-crRNA). The resulting single-stranded RNA is called pre CRISPR RNA, which contains copies of the invading viral DNA sequence in its spacers. Then the protein Cas 9 gets involved. Cas refers to Crispr-associated nuclease proteins, and as we know, nucleases are enzymes that are capable of cleaving DNA at specific nucleotide linkages.

In particular, Cas 9 is one of the nucleases found in Streptococcus pyogenes, which is one of the most used Crispr-associated nuclease proteins. Along with Cas 9, Tracr RNA, which have a section complementary to the palindromic repeats and can anneal with it. So far, each spacer and palindromic repeats end up with a complex containing a segment of pre-Crispr RNA, tracer RNA, and Cas 9 protein. Then another enzyme called ribonucleases (RNAIII) cleave the strand in between these complexes leaving complexes called effector complexes making the bacteria ready for defending against the invaders. If these complexes encounter with a section of viral DNA that has a sequence, which is complementary to this crRNA, the nuclease enzyme will coordinate, and it recognizes a short sequence unique to the viral genome called a protospacer adjacent motif (PAM), where it will snip a few base pairs upstream from the PAM and cut where it will neutralize the virus and preserve a small part for another invention.

5.2. Targeting The Crispr RNAs

Scientists have adapted its components into a biotechnology tool for editing DNA. Jennifer Doudna and Emmanuelle Charpentier proposed genome editing using the CRISPR-Cas9 system in 2012 and were awarded the Nobel prize in chemistry in 2020. Single guide RNA (sgRNA) can be synthesized in the lab and used to cleave DNA in humans and other animal species. Cas9 protein along with sgRNA can be used to edit DNA at precise locations. Synthesize gRNA with complementary sequence and insert into cell along with Cas9 protein. Cas9 will read DNA until it finds appropriate sequence with PAM sequence, bind, and cleave at desired location.

PAM: PAM is short for proto-spacer adjacent motif. It is usually a three-nucleotide sequence consisting of 5'NGG-3' in which the N represents any nucleotide (A, C, G, or T) followed by two guanine (G) nucleotide bases. In humans, PAM motifs occur approximately every 50 bases or less.

When the guide RNA perfectly aligns with the target DNA, the RNA and DNA will form a DNA-RNA helix. This binding event activates Cas9's nuclease, or DNA-cutting, activity. It makes specific cuts in the DNA at a position three nucleotides upstream from the PAM site. Two active sites (regions where molecules bind to undergo chemical reactions) on the nuclease domain of Cas9 generate the cuts and cleave both strands of the DNA double helix, resulting in a double-stranded DNA break.

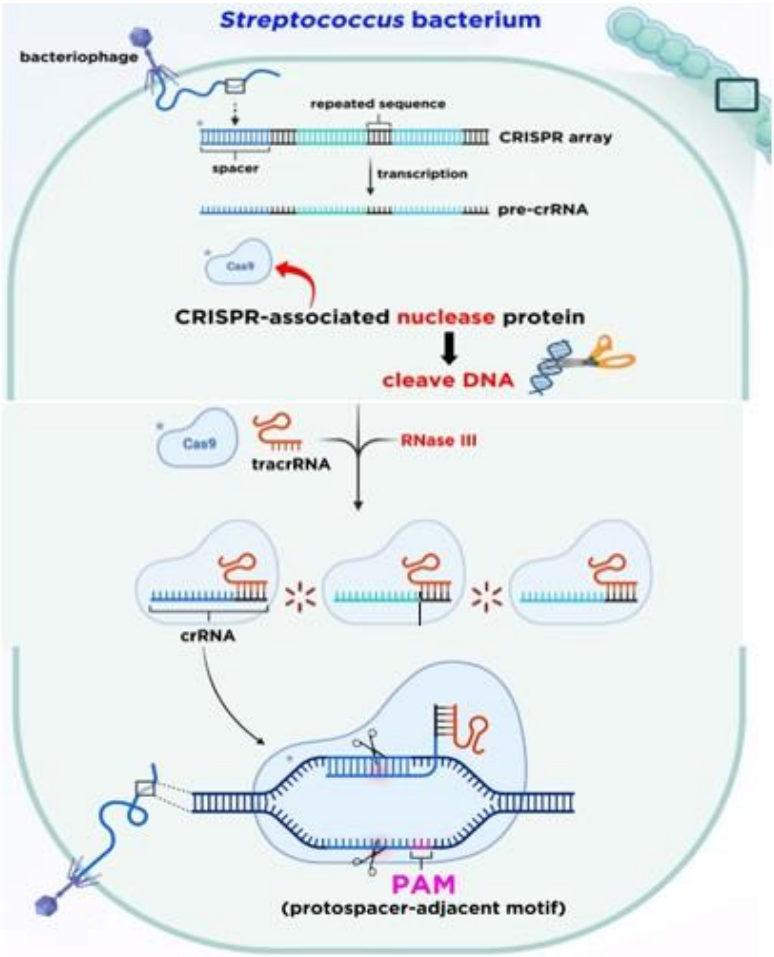

Figure 6. Crispr cas 9 mechanism. Created by the author via BioRender.com.

5.3. Restoration a er cleavage

After cleavage, CRISPR-induced double-stranded DNA breaks can be repaired by either non- homologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ is the more frequently used, faster repair mechanism, because the cell does not use a template to join broken DNA ends together. It is, however, an error-prone process that can introduce mutations in the target sequence. Errors are rare, but when the break is repaired correctly, Cas9 will once again recognize the target sequence and cleave it. Repeated cycles of cleavage and repair eventually result in a mutation. The type of mutation is random, but it will occur precisely within the desired target sequence. If the target sequence is within a gene's coding region, the mutation will likely inactivate that gene. The second type of repair mechanism is homology-directed repair (HDR), which is less error-prone and uses a homologous DNA template to accurately repair the break (for example, from a sister chromatid). Scientists can manipulate this repair system by introducing into the cell an excess of a DNA repair template along with the Cas9guide RNA complex. The cell's repair machinery will be "tricked" into using the repair template to fix the break by HDR. By designing different repair templates, scientists can change the target DNA sequence into a new sequence. These templates could also correct an existing mutation by replacing it with a non-mutated sequence of DNA.

CRISPR-Cas9: How It Edits Genes

The gene-editing mechanism of the CRISPR-Cas9 system focuses on the creation of a double- strand break and the two primary cellular repair pathways that lead to different editing outcomes.

I. The Core mechanism: Targeting and cutting DNA

The CRISPR-Cas9 system is composed of two essential components that work together to find and cut a specific DNA sequence.

- Cas9 Protein: This is an enzyme that acts as "molecular scissors," capable of cutting through both strands of a DNA molecule.
- Guide RNA (gRNA): This is a small, pre-designed RNA sequence that acts as a guide. It binds to the Cas9 protein and directs it to a specific target location in the genome. The gRNA is engineered to have a sequence that is complementary to the target DNA, ensuring the cut is made at the correct spot.

The Process implicates:

1. The gRNA and Cas9 protein are introduced into a cell.
2. The gRNA leads the Cas9 protein to the precise target sequence in the cell's DNA.
3. Once the target is identified, the Cas9 protein cuts the DNA, creating a double-strand break (DSB).

II. The Cell's response: Two types of editing

After the DNA is cut, the cell's natural repair mechanisms are activated to fix the break. The outcome of the gene edit depends on which of the two main repair pathways the cell uses (Guo et al. 2018; Chu et al. 2015).

Type 1: Non-Homologous End Joining (NHEJ) - Gene Disruption

- This is the cell's most common, fast, and error-prone repair pathway. It essentially tries to "glue" the two broken ends of the DNA back together as quickly as possible.
- During the repair process, small insertions or deletions of DNA bases (called "indels") are often accidentally introduced at the cut site.
- These small errors disrupt the original DNA sequence, which can scramble the genetic code and prevent the gene from producing a functional protein. This effectively "knocks out" or disables the gene.

Type 2: Homology Directed Repair (HDR) - Gene Correction or Insertion

- This is a less frequent but much more precise repair pathway. It uses a template to guide the repair process.
- To use this pathway, scientists supply an extra piece of DNA that serves as a repair

template. This template contains the desired sequence and has regions that match the DNA on either side of the break (homology arms). The cell's repair machinery uses this template to fill in the gap, incorporating the new sequence.

- This allows for precise edits. Scientists can use HDR to correct a disease- causing mutation by providing a template with the correct sequence, or to insert a new gene by including it in the template.

As a summary note

In essence, CRISPR-Cas9 works by using a guide RNA to direct the Cas9 enzyme to make a precise double-strand break in DNA. The cell then takes over, and its repair process determines the final edit: the error-prone NHEJ pathway is used to disable a gene, while the precise HDR pathway can be used with a supplied template to correct or insert genetic information.

6. Selectable and screening markers

Selectable marker genes are a vital part of most transformation protocols. They are delivered alongside the gene of interest, either on the same plasmid or on a separate plasmid. As shown in table2, a wide range of selectable marker regimes is available and is particularly important in species where transformation efficiencies are low. Selectable marker genes can be categorized into those based on resistance genes that confer the ability to grow in the presence of toxic compounds such as antibiotics or herbicides which kill or otherwise compromise untransformed tissue (negative selection). Alternatively, a range of positive selection systems are available which provide transformed tissues with an enhanced ability to utilize, for example, an unusual carbohydrate or amino acid supply and thus enrich the culture for transformed tissue expressing the marker gene.

Table 2. Different selectable markers used in host. Table created by the author

Host	**Selection Markers**
Prokaryotic cells	Ampicillin, Kanamycin, tetracycline, chloramphenicol
Yeast	HIS4, Zeocin, Auxotrophy
Mammalian cells	Neomycin, Puromycin, Hygromycin, Zeocin

6.1. The lac operon

The lac operon is an operon, or group of genes with a single promoter (transcribed as a single mRNA) (Figure 7). The genes in the operon encode proteins that allow the bacteria to use

lactose as an energy source. *E. coli* bacteria can break down lactose, but it's not their favorite fuel. If glucose is around, they would much rather use that. Glucose requires fewer steps and less energy to break down than lactose. However, if lactose is the only sugar available, the E. coli will go right ahead and use it as an energy source.

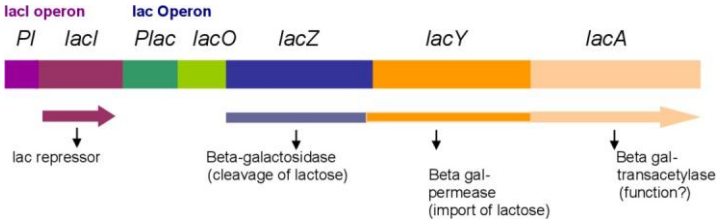

Figure 7. Lac operon's genes. Created by the author via BioRender.com.

The lac operon contains three genes: lacZ, lacY, and lacA. These genes are transcribed as a single mRNA, under control of one promoter. Genes in the lac operon specify proteins that help the cell utilize lactose. lacZ encodes an enzyme that splits lactose into monosaccharides (single-unit sugars) that can be fed into glycolysis. Similarly, lacY encodes a membrane-embedded transporter (permease) that helps bring lactose into the cell. lacA - Transacetylase encodes for Acetyltransferase that not directly involved in lactose metabolism or the regulation of the lac operon, lacA encodes for transacetylase, which acetylates some β-galactosides. In addition to the three genes, the lac operon also contains a number of regulatory DNA sequences (Ghalayini et al. 2019).

In the following points an overview of its action mechanism, as it is illustrated in Figure 8 (Tankeshwa 2025):

- Understanding operons
  - ✓ An operon is a functional DNA unit containing a cluster of genes controlled by a single promoter.
  - ✓ The lac operon consists of three main genes: lacZ, lacY, and lacA.
    - lacZ: Encodes beta-galactosidase, an enzyme that breaks down lactose into glucose and galactose.
    - lacY: Encodes lactose permease, an enzyme responsible for transporting lactose into the cell.
    - lacA: Encodes thiogalactoside transacetylase, which helps remove toxic thiogalactosides from the cell.
  - ✓ The lacI gene encodes a repressor protein that binds to the operator region of the lac operon, inhibiting gene expression.

- Functioning of Lac Operon in the absence of lactose
  - ✓ When lactose is absent, the lac repressor occupies the operator, which effectively shuts down the transcription of the lac operon genes.
  - ✓ Even without lactose, there's a "basal level transcription" where the repressor might temporarily detach from the operator, allowing some transcription of lac genes. This ensures a small amount of lactose permease, beta-galactosidase, and thiogalactoside transacetylase enzymes are always present in the cell.
- Functioning of Lac Operon in the presence of lactose
  - ✓ When lactose is present, it enters the cell via lactose permease (due to basal level transcription).
  - ✓ Lactose then binds to the lac repressor, causing the repressor to detach from the operator.
  - ✓ With the operator free, RNA polymerase can bind to the promoter and initiate transcription of the lac genes.
  - ✓ Continuous transcription leads to the production of more lactose permease, beta-galactosidase, and thiogalactoside transacetylase.
  - ✓ Lactose permease allows more lactose to enter the cell, which is then hydrolyzed into glucose and galactose by beta-galactosidase for energy.
- Role of CAP protein (Catabolite Activator Protein)
  - ✓ CAP binds with cyclic AMP (cAMP) and is also known as cAMP receptor protein.
  - ✓ When glucose levels decrease, cAMP levels in the cell increase.
  - ✓ The cAMP-CAP complex then binds to the CAP site, interacting with the RNA polymerase enzyme.
  - ✓ This interaction significantly increases the rate of transcription of the lac genes.

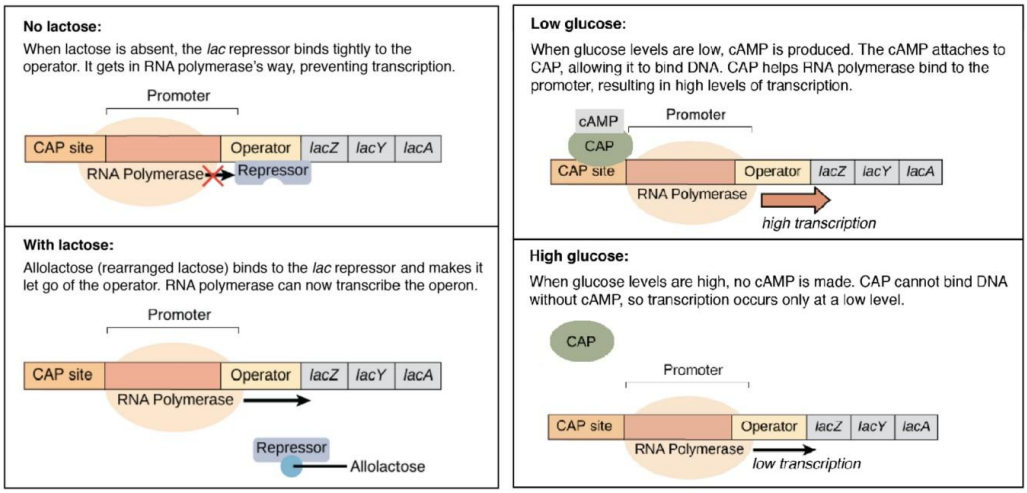

Figure 8. Lac operon's action mechanism. Image modified from "Prokaryotic gene regulation: Figure 3," by OpenStax College, Biology (CC BY 4.0).

6.2. The tryptophan operon

Bacteria such as *Escherichia coli* (a friendly inhabitant of our gut) need amino acids to survive because, like us, they need to build proteins. One of the amino acids they need is tryptophan. The trp operon includes five genes that encode enzymes needed for tryptophan biosynthesis, along with a promoter (RNA polymerase binding site) and an operator (binding site for a repressor protein). The genes of the trp operon are transcribed as a single mRNA.

- The synthesis of enzymes involved in the biosynthesis of tryptophan is controlled by the tryptophan (trp) operon in *E. coli*.
- Charles Yanofsky and others have studied in detail the working of the five structural genes and the close regulatory elements of the trp operon.
- The precursor for biosynthesis of the amino acid tryptophan is "chorismic acid".
- The five structural genes (Figure 9) of the trp operon code for enzymes required for conversion of chorismic acid to tryptophan.
- There are two levels of regulation of this operon: Repression (at the level of transcription initiation) and Attenuation (at the level of transcription termination).

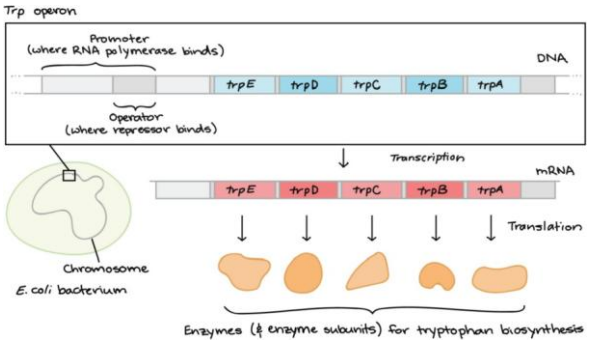

Figure 9. Tryptophan genes. Created by the author via BioRender.com.

According to (Rathore 2020), the trp mechanism is described as follow and illustrated in Figure 10:

- Repression of the trp Operon
  - ✓ The repression regulation works negatively.
  - ✓ The trpR gene coding for repressor of trp operon is distantly located from the operon itself. P1 is for primary promoter region and it contains the operator (O) region.
  - ✓ P2 is a weak promoter found at that end of the trpD gene which is more far from the
  - ✓ operator. Its function is to increase the basal transcription level of the trpC, trpB, and trpA genes.
  - ✓ There are two termination regions, namely t and t' placed downstream to the trpA gene. The trpL denotes a region for a mRNA leader sequence (162 nucleotides in length).

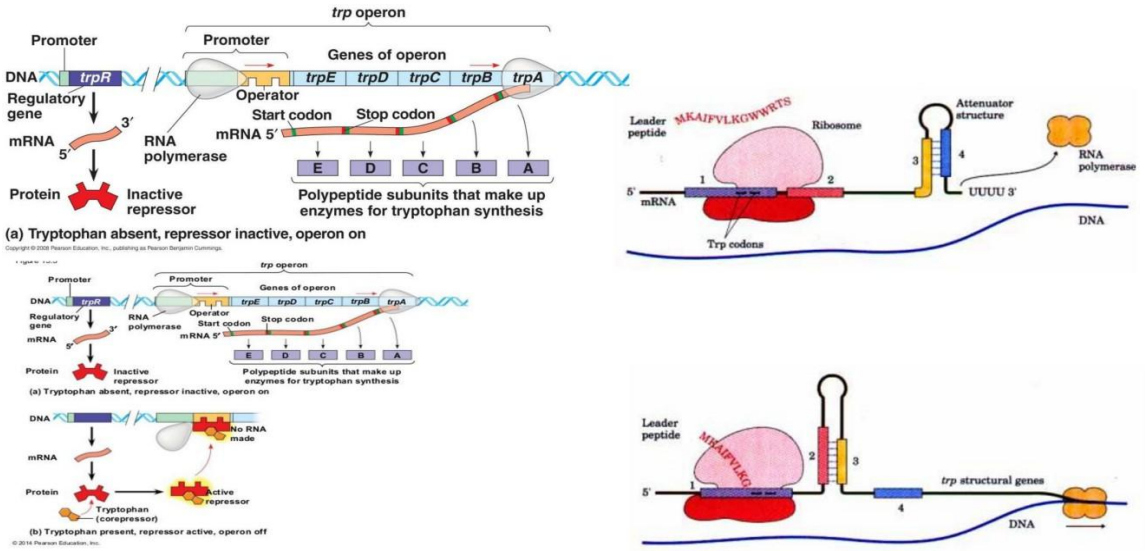

Figure 10. Operon trp mechanism. Copyright by Pearson education, Inc., publishing as Pearson Benjamin

Cummings (2008, 2004).

- *E. Coli* trp operon organization:
  - ✓ The biosynthesis of tryptophan is shown at the bottom.
  - ✓ The five structural genes coding for enzymes needed for tryptophan biosynthesis are: trpE, trpD, trpC, trpB & trpA.
  - ✓ trpL is the regulatory segment.
  - ✓ The gene lengths and the intergenic distances are shown as bp's. PRA: phosphoribosyl anthranilate.
  - ✓ CDRP: carboxyphenylamino-deoxyribulose phosphate.
  - ✓ InGP: indole-glycerol phosphate.
  - ✓ Tryptophan acts as a co-repressor. In its absence RNA polymerase's binding to the promoter initiates the transcription of the structural genes.
  - ✓ When tryptophan is present, it forms complex with the repressor (co-repressor/repressor complex).
  - ✓ This complex pairs with the operator sequence and stops RNA polymerase from initiating the transcription.
  - ✓ The RNA polymerase activity is around 70 times slower in the presence of tryptophan (repressed state) than in the absence of tryptophan (de-repressed state).
  - ✓ In trpR mutants, there is no functional repressor present but due to attenuation, presence of tryptophan can still cause about 10 folds lowering of the transcription rate.
- Attenuation of the trp Operon
  - ✓ If a portion of the trpL region gets deleted then there is an increase in the rate of transcription of the trp operon.
  - ✓ The repressebility of the trp operon does not get affected by this deletion.
  - ✓ This implies that apart from the repression/derepression mechanism, there is some other regulation occurring from the trpL sequence.
  - ✓ "Attenuation" is this second regulatory mechanism and is governed by the trpL sequence known as "attenuator".
  - ✓ Control of transcription termination, occurring at a position near to the end of mRNA leader sequence, is the event that implies attenuation.
  - ✓ Presence of tryptophan-charged tRNATrp is a prerequisite for the early

termination of transcription of the trp operon.

- ✓ A truncated trp transcript (140 nucleotides in length) is formed by this attenuation/premature termination.
- ✓ There is a nucleotide pair sequence in the attenuator region that is indistinguishable to the termination sequences present at the end of majority of bacterial operons.
- ✓ These specific terminator sequences contain several A:T pairs preceded by a G:C rich palindrome.
- ✓ When these terminator sequences are transcribed, a nascent RNA is obtained. It has regions that form a "hairpin" structure by hydrogen bonding. It contains several uracils a er the hairpin structure.

- Leader region sequence structure in the trp mRNA:
  - ✓ trpL sequence: encodes leader peptide.
  - ✓ Two tandem trp codons: control of attenuation by tryptophan.
  - ✓ Four shaded regions: hairpin structures.
  - ✓ The two different secondary structures formed by the pairing of specific mRNA regions in trpL transcript.

    i. **Alternate 1:** Regions 1 and 2 base-paired and regions 3 and 4 base-paired, forming a transcription-termination hairpin.

    ii. **Alternate 2:** Regions 2 and 3 base-paired, where Regions 3 and 4 are prevented from pairing with each other, and the transcription-termination hairpin can NOT form.

7. Application of expression vector

Expression vectors are fundamental tools in biotechnology, enabling the efficient production of recombinant proteins. Their applications are diverse and span across various fields, including medicine, industry, and fundamental research. Key applications include the construction of expression libraries for functional genomics, the analysis of gene function at the protein level, the commercial production of valuable proteins, the generation of antibodies, and the assessment of protein function through in vivo studies.

Saccharomyces cerevisiae, commonly known as baker's yeast, serves as a prominent host for recombinant protein production due to its well-understood genetics, ease of cultivation, and capacity for post-translational modifications. In the medical field, S. cerevisiae has been instrumental in producing a range of therapeutic and diagnostic proteins:

- Hepatitis B Surface Antigen (HBsAg), utilizing the GAL1 promoter and a yeast episomal plasmid, is produced for use in vaccines against Hepatitis B and as a diagnostic marker for infection.

- Human Insulin, driven by the ADH2 promoter in a yeast episomal plasmid, is crucial for diabetes treatment and insulin measurement diagnostics.

- Human Growth Hormone (hGH), expressed via the GAL10 promoter in a yeast episomal plasmid, is applied in growth hormone therapy and as a diagnostic for growth disorders.

- Green Fluorescent Protein (GFP), with the PGK1 promoter in an integrative yeast plasmid, is vital for molecular biology research and as a diagnostic marker for gene expression.

- α-Amylase, employing the TEF1 promoter in a yeast episomal plasmid, is produced for enzyme applications and as a diagnostic for starch digestion.

- Erythropoietin (EPO), expressed using the TPI1 promoter in a yeast episomal plasmid, is used in anemia treatment and to stimulate red blood cell production.

- Interferon-alpha (IFN-α), under the control of the PGK1 promoter in a yeast episomal plasmid, serves as a therapeutic protein for antiviral and anticancer treatments.

Beyond medicine, S. cerevisiae is also extensively used for producing recombinant proteins for various biotechnological applications:

- Lipase, expressed via the SUC2 promoter in an integrative yeast plasmid, finds use as an enzyme in food processing, such as in cheese and bakery products.

- Invertase, utilizing the TEF1 promoter in a yeast episomal plasmid, is critical for the production of high-fructose corn syrup in the food industry.

- Pectinase, driven by the PGK1 promoter in a yeast episomal plasmid, is an enzyme used for fruit juice clarification in beverage production.

- Xylanase, employing the TPI1 promoter in a yeast episomal plasmid, is an enzyme for breaking down plant cell wall material in biomass processing.

- Laccase, expressed with the GAL1 promoter in a yeast episomal plasmid, is applied in the bioremediation of phenolic compounds in wastewater treatment.

- Hyaluronic Acid Synthase, under the FBA1 promoter in a yeast episomal plasmid, enables the production of hyaluronic acid for cosmetics and skincare products.

- Collagen-like Protein, utilizing the TDH3 promoter in a yeast episomal plasmid, is developed for cosmetic applications related to skin rejuvenation and anti-aging.

- β-Glucosidase, expressed via the AOX1 promoter in a yeast episomal plasmid, enhances aroma in wine production.

- Tyrosinase, with the PGK1 promoter in an integrative yeast plasmid, is used for cosmetic

applications related to skin pigmentation.

- Phytase, employing the CUP1 promoter in a yeast episomal plasmid, acts as a food and feed additive for improved phosphorus utilization.

- Keratinase, driven by the MET25 promoter in a yeast episomal plasmid, is used in cosmetic and skincare applications for hair treatment.

This comprehensive range of applications highlights the versatility and importance of Saccharomyces cerevisiae as an expression system in modern biotechnology for producing diverse recombinant proteins with significant therapeutic, industrial, and research value.

Chapter II: Production of fusion proteins

Introduction: What is a fusion protein?

A fusion protein is a protein made by rDNA technology that consists of at least two domains that are encoded by separate genes that have been joined so that they are transcribed and translated as a single unit, producing a single polypeptide. Fusion proteins can be created in vivo, for example, as the result of a chromosomal rearrangement. Fusion proteins can also be created in vitro using recombinant DNA techniques. The fusion often consists of a protein that is being studied joined to one of a small number of proteins that have useful properties to aid in the study.

1. Definition of tag protein

The production of recombinant proteins in a highly purified and well-characterized form has become a major task for the protein chemist working in the pharmaceutical industry. In recent years, several epitope peptides and proteins have been developed to over-produce recombinant proteins. These affinity-tag systems share the following features: (a) one-step adsorption purification; (b) a minimal effect on tertiary structure and biological activity; (c) easy and specific removal to produce the native protein; (d) simple and accurate assay of the recombinant protein during purification; (e) applicability to a number of different proteins. Nevertheless, each affinity tag is purified under its specific buffer conditions, which could affect the protein of interest.

Thus, several different strategies have been developed to produce recombinant proteins on a large scale. One approach is to use a very small peptide tag that should not interfere with the fused protein. The most commonly used small peptide tags are poly-Arg-, FLAG-, poly-His-, c-myc-, S-, and Strep II-tag. For some applications, small tags may not need to be removed. The tags are not as immunogenic as large tags and can o en be used directly as an antigen in antibody production. The effect on tertiary structure and biological activity of fusion proteins with small tags depends on the location and on the amino acids composition of the tag. Another approach is to use large peptides or proteins as the fusion partner. The use of a large partner can increase the solubility of the target protein. The disadvantage is that the tag must be removed for several applications e.g. crystallization or antibody production. In general, it is difficult to decide on the best fusion system for a specific protein of interest. This depends on the target protein itself (e.g. stability, hydrophobicity), the expression system, and the application of the purified protein (Kimple et al. 2013).

- Translation fusion of sequences coding a recombinant protein and a) short peptides [ex. (His)n, (Asp)n, (Arg)n...]. b) protein domains, entire proteins[ex. MBP, GST, thioredoxin …].
- Engineering a tagged protein requires adding the DNA encoding the tag to either the 5’ or 3’ end of the gene encoding the protein of interest to generate a single, recombinant protein with a tag at the N-or C-terminus. The stretch of amino acids containing a target cleavage sequence (CS) is included to allow selective removal of the tag (Figure 11).

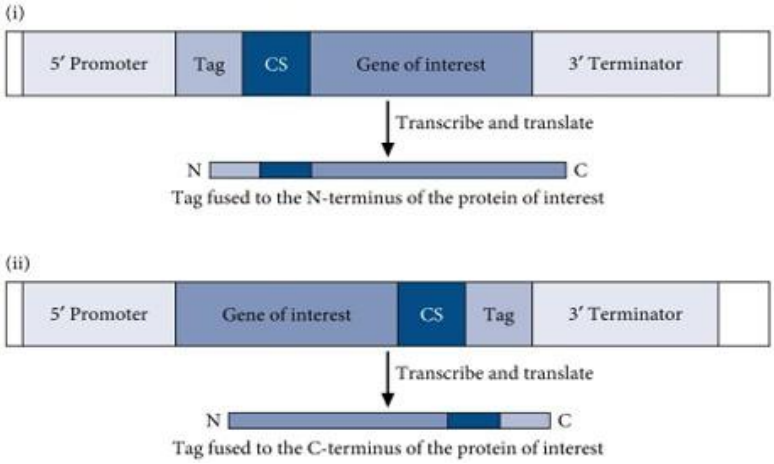

Figure 11. Tagged protein sites. Created by the author via BioRender.com.

What is a Protein Tag?

Protein tag refers to the sub-domain or the peptide sequence of a fusion protein or molecules that are attached to a protein to modify its properties or function.

Many kinds of protein tags available for recombinant protein production that can be used to identify, purify, or track proteins, and they have numerous applications in protein research and biotechnology.

Protein Tagging

Protein tagging is a technique used in biomedical research where peptide sequences are genetically added to recombinant proteins. This method allows for the study and manipulation of proteins in various ways.

- Studying Subcellular Localization: By tagging a protein with GFP (Green Fluorescent Protein), its real-time location within live cells can be observed using fluorescence microscopy. This technique provides a significant advantage over immunohistochemistry, which cannot show real-time localization of endogenous proteins.
- Protein Purification: Tags like the His-tag (histidine tag) facilitate the purification of a protein of interest from a complex mixture. The His-tagged protein binds to a specific matrix, allowing for separation through affinity purification, often employing column chromatography.
- Studying Protein-Protein Interaction: Tagging proteins with fluorophores such as Cyan Fluorescent Protein (CFP) or Yellow Fluorescent Protein (YFP) allows for the use of Förster Resonance Energy Transfer (FRET). The presence of FRET indicates a direct interaction between the two tagged proteins.

Important note

Almost all recombinant proteins are prepared using fusion domains, also known as “tags” or protein tags.

2. Types of tag

There are several types of protein tags commonly used in molecular biology and protein research. These tags are attached to the target protein of interest to facilitate purification, detection, localization, and/or functional studies. Some of the commonly used protein tags include (Terpe, 2005):

2.1. Polyarginine-tag (Arg-tag)

The Arg-tag was first described in 1984 and usually consists of five or six arginines. It has been successfully applied as C-terminal tag in bacteria, resulting in recombinant protein with up to 95% purity and a 44% yield. Arginine is the most basic amino acid. Arg5-tagged proteins can be purified by cation exchange resin SP-Sephadex, and most of the contaminating proteins do not bind. A er binding, the tagged proteins are eluted with a linear NaCl gradient at alkaline pH. Polyarginine might affect the tertiary structure of proteins whose C-terminal region is hydrophobic. The Arg-tagged maltodextrin- binding protein of *Pyrococcus furiosus* has been crystalized. This enzymatic process has been successfully used in several instances, but often has been limited by poor cleavage yields or by unwanted cleavage occurred within the desired protein sequence. The Arg-tag can be used to immobilize functional proteins on flat surfaces; this is important for studying interactions with ligands. The Arg-tag is not used very often, in combination with a second tag it can be an interesting tool for protein purification.

2.2. Polyhistidine-tag (His-tag)

A widely employed method utilizes immobilized metal- affinity chromatography to purify recombinant proteins containing a short affinity-tag consisting of polyhistidine residues. Immobilized metal-affinity chromatography. Histidine is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices, as electron donor groups on the histidine imidazole ring readily form coordination bonds with the immobilized transition metal. Peptides containing sequences of consecutive histidine residues are efficiently retained on IMAC. Following washing of the matrix material, peptides containing polyhistidine sequences can be easily eluted by either adjusting the pH of the column buffer or by adding free imidazole.

2.3. FLAG-tag

The FLAG-tag system utilizes a short, hydrophilic 8- amino-acid peptide that is fused to the protein of interest. The FLAG-tag can be located at the C- or N-terminus of the protein. The purification condition of the system is non-denaturing and thus allows active fusion proteins to be purified. The complex can be dissociated by chelating agents such as EDTA or by transiently reducing the pH.

2.4. Strep-tag

The Strep-tag is an amino acid peptide that was developed as an affinity tool for the purification of corresponding fusion proteins on streptavidin columns. Strep-tagged proteins are bound under physiological buffer conditions in the biotin binding pocket, and can be eluted gently with biotin derivatives. The tag can be engineered to either the C- or N-terminus of a protein. Recombinant Strep-tag-hybrids are produced in bacteria. The Strep-tag system is of relevance for studies on protein-protein interaction and special applications in which large or charged tags are not functional.

2.5. S-tag

The S-tag sequence is a fusion-peptide tag that allows detection by a rapid, sensitive homogeneous assay or by colorimetric detection in Western blots. The system is based on the strong interaction between the 15-amino- acid S-tag and the 103-amino-acid S-protein, both of which are derived from RNaseA. The tag is composed of four cationic, three anionic, three uncharged polar, and five non-polar residues. This composition makes the S-tag soluble. The S-tag rapid assay is based on the reconstitution of ribonucleolytic activity. Tagged proteins can be bound on S-protein matrices. The elution conditions are very harsh, e.g. buffer with pH 2. However, it is recommended to cleave the tag with protease to get functional proteins. The system is functional to purify recombinant proteins from bacteria.

2.6. Cellulose-binding domain

More than 13 different families of proteins with cellulose- binding domains (CBDs) have been classified. CBDs can vary in size from 4 to 20 kDa; they occur at different positions within polypeptides: N-terminal, C-terminal and internal. Some CBDs bind irreversibly to cellulose and can be used for immobilization of active enzymes. Hydrogen bond formation and van der Waals interaction are the main driving forces for binding. The advantage of cellulose is that it is inert, has low non-specific affinity, is available in many different forms, and has been approved for many pharmaceutical and human uses. CBDs bind to cellulose at a moderately wide pH range, from 3.5 to 9.5. The tag can be placed at the N- or C-terminus of the target protein. The affinity of the tag is so strong that an immobilized fusion protein can only be released with buffers containing urea or guanidine hydrochloride. This de-naturating elution conditions make refolding of the fused target protein necessary. Fused proteins with CBDs can be eluted gently from cellulose with ethylene glycol. This low-polarity solvent presumably disrupts the hydrophobic interaction at the binding site.

2.7. Maltose-binding protein

The 40-kDa maltose-binding protein (MBP) is encoded by the malE gene of *E. coli* K12. Vectors that facilitate the expression and purification of foreign peptides in E. coli by fusion to MPB were first described in 1988. Fused proteins can be purified by one-step affinity chromatography on cross-linked amylose. Bound fusion proteins can be eluted with 10 mM

maltose in physiological buffer. Binding affinity is in the micro-molar range. The MBP-tag can be easily detected using an immuno- assay. It is necessary to cleave the tag with a site-specific protease. The MBP can be fused at the N- or C-terminus of the protein if the proteins are expressed in bacteria. N-terminal location can reduce the efficiency of translation. The MBP system is widely used in combination with a small affinity tag.

3. Uses of protein tags

Protein tags are widely used in molecular biology and protein research for various purposes (Zhao et al. 2013). Some common uses of protein tags include:

3.1. Purification of recombinant proteins

Purification of recombinant proteins is a crucial step in the production process to eliminate impurities and obtain a high-quality final product. Common methods of purification include chromatography, dialysis, centrifugation, and filtration. The choice of method depends on the nature of the protein, its impurities, and the desired purity level. In addition, it presents these properties:

- The recombinant protein to be produced is attached with a fusion tag.
- The tagged protein can be easily and conveniently purified by affinity chromatography.

As example: Fusion tags available with affinity, such as glutathione S-transferase, ß-galactosidase, Maltose- binding protein, Polyhistidine tag, Cellulose-binding protein, and Staphylococcus protein A.

3.2. Enhancing the solubility of recombinant proteins

Enhancing the solubility of recombinant proteins is crucial for their successful expression, purification, and subsequent functional studies, as shown in Figure 12 from (Gilroy et al. 2018). Depending in the soluble molecule, different mechanism of action can be defined into:

- Maltose binding protein (MBP) might bind reversibly to exposed hydrophobic regions of nascent target polypeptide, steering the polypeptides towards their native conformation by a chaperone like –mechanism.
- NusA decreased translation rates by mediating transtriptional pausing, that might enable critical folding events to occur.
- Negative charged tags (highly acidic peptide) inhibit aggregation by increasing electrostatic repulsion between nascent polypepdides.

Taking in example of solubility-enhancing tags: *In vitro* using short peptide tag like Poly-Lys tag, poly-Arg tag= one, three and five lysine or arginine residues fused to the C-or N- terminus of the target protein. The solubilization effect of poly-Lys tags is lower than that of poly-Arg tags (lysines are less hydrophilic than arginines) of tagged BPTI-22 (bovine pancreatic trypsin inhibitor).

- The solubilization factor of all C-terminal tags was slightly higher than that of the respective N- terminal tags. Table 3 demonstrates the characteristic between using peptide or protein tags for solubility enhancement of the recombinant protein.

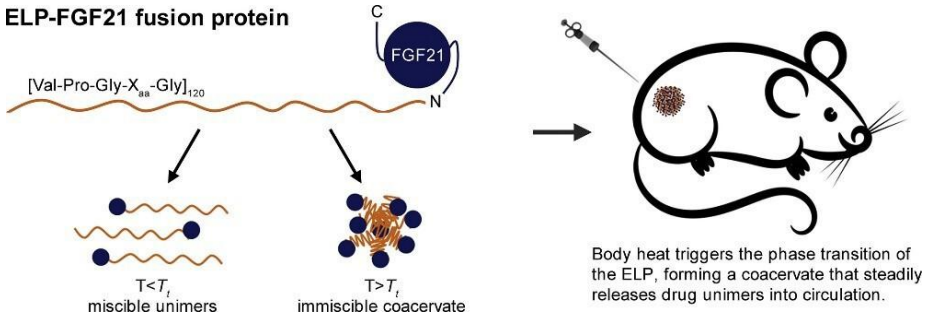

Figure 12. Enhancing the Solubility of Recombinant Proteins (Gilroy et al. 2018).

Table3. Solubility-enhancing tags –comparison of peptide and protein tags

Characteristic	**Peptide Tags**	**Protein Tags**
Structure	Short peptide sequences, typically 10-20 amino acids long.	Full-length proteins, often larger in size.
Solubility enhancement	Peptide tags are often used to improve the solubility of the tagged protein. They may include hydrophilic residues.	Protein tags can enhance solubility, but they may also introduce additional structural features that affect solubility.
Examples	- His6 tag (6x Histidine) - FLAG tag - HA tag (Hemagglutinin)	- GST tag (Glutathione S-Transferase) - MBP tag (Maltose-Binding Protein) - Thioredoxin tag
Affinity-based purification	Often used for purification through immobilized metal ion affinity chromatography (IMAC)	Can be used for various affinity purification methods, including affinity chromatography using specific ligands.
Structural impact on tagged protein	Minimal structural impact, and the tag is often easily removable.	The addition of a full-length protein tag can have a more significant structural impact, which may or may not be desirable.
Flexibility in tag placement	Typically, peptide tags can be placed at either the N- or C-terminus of the protein.	Protein tags may have more specific requirements for tag placement due to their structural characteristics.
Field of application	Widely used in recombinant protein expression and purification to enhance solubility and simplify purification.	Often used when solubility enhancement and affinity purification are both desired, or when the tag has additional functions (e.g., enzymatic activity).
Tag removal	Peptide tags are often removable using proteases or other specific cleavage methods.	Protein tags may or may not be easily removable, depending on the tag and its function.

3.3. Membrane or lipid-binding fusion proteins

These proteins facilitate the fusion of membranes or lipids, enabling the transport of molecules across cellular boundaries. Examples of membrane or lipid-binding fusion proteins include viral fusion proteins, which mediate viral entry into host cells, and membrane fusion proteins found in various neurological disorders. By understanding the structure and function of these proteins can provide insights into the development of therapeutic strategies for a range of diseases.

3.4. Toxin fusion proteins

Some fusion proteins include toxin fragments, allowing for targeted cell killing, such as immunotoxins used in cancer therapy. Interestingly, cost effective full-length IgG and IgG-fusion protein production in *E. coli* of made an attractive option for antibody production for research and hopefully for clinical applications.

3.5. Enzyme fusion proteins

Enzyme fusions, such as β-lactamase or alkaline phosphatase, are used for enzyme-linked assays or to monitor protein localization and activity.

3.6. Targeted delivery fusion proteins

These fusions incorporate protein domains that specifically target them to particular cell types or subcellular compartments, improving precision in drug delivery and gene therapy.

3.7. Fusion tags as reporters

Enable direct observation of dynamic intracellular processes. Reporter genes are used to observe expression of a specific gene, using GFP, luciferases, and Lac Z. To track the expression level of the specific protein. These tags are typically small and can be easily detected, purified, or localized within cells.

3.8. Combinatorial tagging

- No single tag is ideally suited for all purposes. Therefore, combinatorial tagging might be the only way to harness the full potential of tags in a high-throughput setting.
- Solubility-enhancing tag + purification tag: MBP + His6 tag
- 2xpurification tag:IgG-binding domain + streptavidin-binding domain
- Localization tag + purification tag: GFP + His6tag
- Localization tag + 2x purification tag + immunodetection: GFP + SBP domain + His8tag + c-Myc.

4. Reporter genes

Reporter genes are genes that are used in molecular biology and genetics to study the expression patterns, regulation, and localization of other genes or genetic elements. These genes encode proteins or enzymes that produce easily detectable signals, allowing researchers to track the activity of the reporter gene and indirectly monitor the activity of the target gene or genetic element. The video bellow explains the importance of these genes

Reporter genes are indispensable tools in molecular biology, primarily employed to visualize the spatio-temporal dynamics of gene expression. Their fundamental utility lies in investigating the activity and regulatory mechanisms of specific gene promoters. By genetically fusing a reporter gene, such as Green Fluorescent Protein (GFP), to a promoter of interest, researchers can quantitatively assess the transcriptional output from that promoter.

The output of reporter genes can be visualized through various detection methods:

- Fluorescence-based detection is a prevalent method, commonly utilizing GFP. The GFP gene sequence produces a protein that inherently fluoresces, enabling straightforward detection. This approach offers high cellular and molecular resolution, making it suitable for tracking gene expression dynamics in live specimens over time.
- Enzymatic detection frequently employs the LacZ gene, which encodes for beta-galactosidase. Upon substrate addition, this enzyme yields colored products, providing a colorimetric indication of gene expression. While historically significant, LacZ reporters typically provide tissue-level information rather than fine cellular resolution.
- Bioluminescence-based detection leverages luciferase enzymes, which catalyze reactions that produce bioluminescence, allowing for sensitive detection.

Reporter gene assays are pivotal for addressing a range of scientific inquiries:

- Tissue or cell type specificity: These assays can elucidate whether a gene of interest is expressed in particular tissues or cell types. By transfecting different cell lines with a reporter construct, the presence of reporter expression (e.g., GFP fluorescence) indicates transcriptional activity in those specific cells (Grierson et al. 1992).
- Impact of treatments or pathological conditions: Reporter assays can quantify changes in gene expression (increase or decrease) under specific experimental conditions, such as drug treatment or pathological states. By comparing reporter activity in treated versus control cells, researchers can deduce the effect of the condition on transcription (McLean et al. 2009).
- The transcriptional activity or "strength" of a promoter can be quantitatively evaluated by cloning it upstream of a reporter gene. The intensity of the reporter's signal (e.g., GFP fluorescence) directly correlates with the promoter's activity, with stronger promoters leading to higher reporter protein production. Beyond fundamental gene

expression studies, reporter genes are increasingly applied to strategically placing a reporter gene under the control of a promoter responsive to a specific cell signaling pathway, researchers can ascertain whether the pathway is active ("on") or inactive ("off"). This facilitates real-time monitoring of complex cellular signaling dynamics.

- Specialized activity-based reporters can provide insights into physiological changes within neurons, such as fluctuations in calcium levels or membrane voltage, which are indicative of neuronal activation (Barth, 2007).

4.1. Transcriptional fusion

Transcriptional fusion involves joining a promoter region (the regulatory sequence that controls gene expression) to a reporter gene (tag gene). It allows researchers to monitor the transcriptional activity of a gene by observing the expression of the reporter gene. This method helps in understanding when and where a gene is turned on (but not where the translated protein ends up). In a typical transcriptional fusion, the promoter region of a gene of interest is linked to a reporter gene, such as GFP (Green Fluorescent Protein). When the gene of interest is transcriptionally active, the reporter gene produces a visible signal (fluorescence in the case of GFP), indicating gene expression.

4.2. Translational fusion

Translational fusion involves combining a GOI with a reporter gene in a way that their coding sequences are fused together, often resulting in a chimeric protein. It allows researchers to investigate the expression and localization of the chimeric protein, formed by the fusion of the target gene and the reporter gene. This method is used to study protein expression and function (and not necessarily where the RNA was transcribed). GOI is fused with a reporter gene, such as GFP, at the DNA level. The resulting protein, which contains a portion of the gene of interest and the reporter, is expressed and localized within the cell. By observing the fluorescence of GFP, researchers can track the protein's presence and movement.

5. Removal tags

Removal protein tags refer to the process of removing biological tags or epitopes from proteins. These tags are often added to proteins for identification or labelling purposes, but their presence can interfere with protein interaction studies or structure determination. There are several methods for removing protein tags, including enzymatic cleavage, chemical modification, or genetic engineering strategies. We can cite them as follow (Goh et al. 2017):

5.1. Chemical cleavage

Chemical cleavage is a harsh method, efficient, but rather non-specific and may lead to unnecesary denaturation or modification of the target protein.

- Cyanogen bromide (CNBr) cleaves at C- terminal of methionine (Met) residues

- Hydroxylamine (NH2OH) cleaves at Asn-Gly (NG) sites in proteins

5.2. Self -cleaving

Self-cleavage is a process in which a protein releases itself from its binding site or frees itself from its active form a er performing its function. This can be achieved through different mechanisms:

a. Intervening proteins

Inteins (intervening proteins) are protein segments that can excise themselves from protein precursors in which they are inserted and rejoin the flanking regions.

b. Self-cleaving peptide sequences

System based on the catalytic domain of *Staphylococcus aureus* sortase A(SrtA). SrtA cleaves the Thr-Gly bond at the conserved LPXTG motif in the substrates. Cleavage is inducible by adding calcium (cofactor of Srt A).

c. FrpC modul

FrpC protein undergoes calcium–inducible autocatalytic proccesing at the peptide bond between residues Asp and Pro.

d. N-terminal protease (Npro)

The first protein of the pestivirus polyprotein. It posesses auto proteolytic activity and catalyzes the cleavage by switching from chaotropic to cosmotropic conditions.

e. Proteolytic cleavage

Vibrio cholerae secretes a large multifunctional auto processing repeats-in-toxin (MARTX) toxin that under goes proteolytic cleavage during translocation into host cells. Proteolysis of the toxin is mediated by a conserved internal cystein protease domain (CPD), which is activated upon binding of inositol hexakisphosphate.

5.3. Enzymatic cleavage

- Unspecific cleavage (optimization of protein cleavage conditions or using re-engineered proteases with increased specificity such as ProTEV and AcTEV proteases).
- Optimization of protein cleavage conditions (mainly enzyme-to-substrate ratio, temperature, pH, salt concentration, length of exposure).
- Precipitation of the target protein when the fusion partner is removed (approach for protein sollubilization has to be found, figure 13).

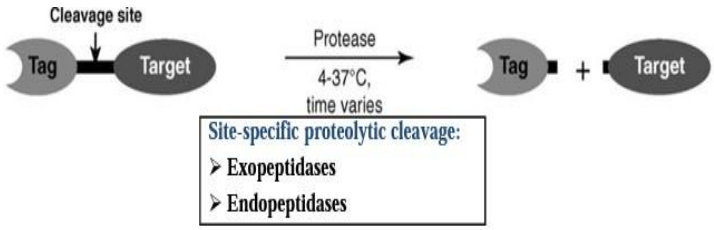

Figure 13. Enzymatic cleavage. (Goh et al. 2017).

6. Purification

In order to resume the principal technique that are broadly used in purification of the fusion protein, the following Table 4 presents the key advantages of both co-immunoprecipitation and pull-down assay.

Table 4. Difference between co-immunoprecipitation and pull-down assay in protein fusion purification. Table created by the author.

Characteristic	Pull-down assay	Co-immunoprecipitation (co-ip)
Aim	To isolate and identify proteins that interact with a specific bait protein.	To study protein-protein interactions, including those occurring in complexes.
Bait protein	The bait protein is immobilized onto a solid support (e.g., beads or resin).	Typically, the bait protein is in its native form within the cellular context.
Support material	Solid support (e.g., beads or resin) with immobilized bait protein.	Antibodies specific to the bait protein or an associated tag.
Protein of interest (poi)	Bait protein is the primary focus, and interactors are pulled down with it.	Bait protein, along with any interaction partners (co-immunoprecipitated).
Immobilization method	Bait protein is typically tagged and immobilized using affinity tags or antibodies.	Antibodies specific to the bait protein are used to capture the bait and its interactors.
Sample complexity	Often used with purified samples or overexpressed proteins for controlled conditions.	Typically used with complex cellular lysates or extracts.
Detection method	Proteins are eluted and detected using methods like gel electrophoresis, Western blotting, mass spectrometry, or enzymatic assays.	Proteins are analyzed by Western blotting or mass spectrometry, often using antibodies specific to the bait protein.
Specificity	Relies on the affinity of the immobilized bait protein for its interacting partners.	Utilizes specific antibodies to pull down the bait protein and its associated partners.
In Vivo vs. *In Vitro*	Can be used in both in vitro (purified proteins) and in vivo (cellular context) settings.	Typically performed in the in vivo context of cell lysates or tissues.
Applications	Primarily used for the identification of proteins that interact with a specific bait protein.	Used to study protein-protein interactions and protein complexes within cells.

Chapter III: Biological engineering via directed mutagenesis and agricultural applications

1. Overview

What is mutation?

Mutation is defined as random and permanent change in a DNA sequence which caused by some mutagens that can be viruses, mutagenic chemicals (ethyl methane sulfonate, methyl methane sulfonate, etc.), transposons, and radiation as well as errors that occur during meiosis or DNA replication. Many mutations can be neutral in effect that can cause a gene or a chromosome (Johnston 2006).

Thus, broadly mutation maybe:

- Gene mutation where the allele of a gene changes.
- Chromosome mutation where segments of chromosomes, whole chromosomes, or entire sets of chromosomes change.

There are various schemes for classification of different kind of mutations. Depending on :

A. The type of cell involved and mode of origin

i. **Somatic mutations**

- Mutations that are in the somatic tissues of the body.
- Mutations are not transmitted to progeny.
- The extent of the phenotypic effect depends upon whether the mutation is dominant or recessive (dominant mutations generally have a greater effect).
- The extent of the phenotypic effect depends upon whether it occurs early or late in development (early arising mutations have a greater effect).

ii. **Germinal mutations**

- Mutations that are in the germ tissues of the body.
- Mutations may be transmitted to progeny.
- Dominant mutations are seen in first generation a er the mutation occurs.
- If a female gamete containing an X-linked mutation is fertilized, the males will show the mutant phenotype.
- Recessive mutations will only be seen upon the chance mating with an individual

carrying the recessive allele too; thus, the recessive mutation may remain hidden for many generations.

iii. Spontaneous mutations

- The spontaneous mutations occur suddenly in the nature and their origin is unknown. They are also called “background mutation” and have been reported in many organisms such as, Oenothera, maize, bread molds, microorganisms (bacteria and viruses), Drosophila, mice, man, etc.

iv. Induced mutations

- Besides naturally occurring spontaneous mutations, the mutations can be induced artificially in the living organisms by exposing them to abnormal environment such as radiation, certain physical conditions (i.e., temperature) and chemicals.

Example (case of sickle cell anemia)

- Hemoglobin is a protein in your red blood cells that helps you carry oxygen but in the disorder sickle cell anemia, the gene that codes for hemoglobin is mutated (figure 14).
- If you inherit two copies of this gene (one from each parent), you can have this disorder.
- This disorder can make it difficult for your red blood cells to carry oxygen because the shape of red blood cell is affected from this mutation.
- It is worth to mention that if an individual only inherits one copy of the mutated gene from one parent, they are a carrier and do not officially have the disease.

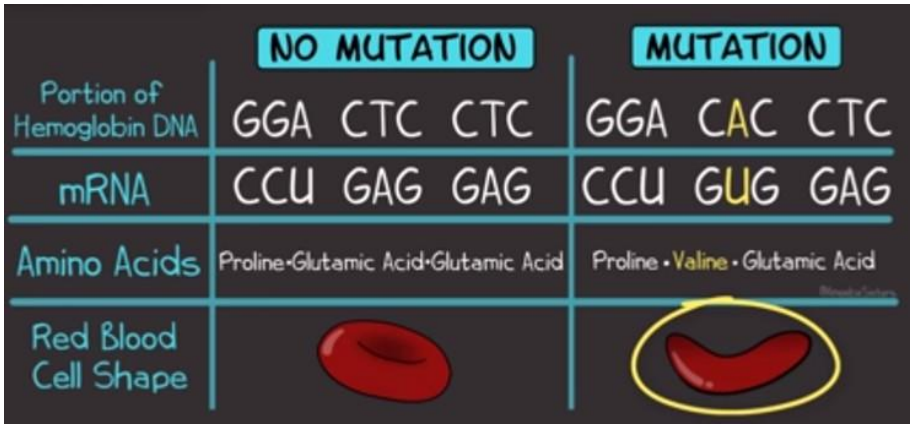

Figure 14. Sickle cell anemia gene mutation. Created by the author via BioRender.com.

2. What is mutagenesis?

Mutagenesis is defined as “to generate mutation” in a DNA for a particular cause. It gives us

the capability of testing the role of any amino acid in a protein by replacing it with any of the other naturally occurring amino acids. We can do it in the lab through chemical irradiation or insertional methods.

Mutagenesis is the process of an organism's deoxyribonucleic acids (DNA) change, resulting in a gene mutation. A mutation is a permanent and heritable change in genetic material, which can result in altered protein function and phenotypic changes. DNA consists of nucleotides that contain a phosphate backbone, a deoxyribose sugar, and1 of 4 nitrogen-containing bases (adenine [A], guanine [G], cytosine [C], and thymine [T]). DNA mutagenesis occurs spontaneously in nature or as a result of mutagens (agents predisposing to alter DNA). Furthermore, molecular genetic techniques, such as polymerase chain reaction (PCR), have revolutionized how mutations are obtained and studied. Mutagenesis is the driving force of evolution; however, it can also lead to cancers and heritable diseases.

Mutagenesis is the driving force behind evolution and genetic variation results from mutations. Any change in genetic information may result in advantageous or disadvantageous phenotypic characteristics that impact an organism's fitness. When a mutation results in a higher fitness, natural selection favors these phenotypes, and these traits are more likely to be passed to offspring (Durland and Ahmadian-Moghadam 2022).

Case study of mutagenesis of Arabidopsis

Arabidopsis is one of the most popular model organisms in the modern lab. It is characterized by a small size, rapid generation time, grows well in the lab, and very small plant genome that has been entirely sequenced. Mutagenesis of Arabidopsis involves the use of various mutagenic agents, such as ethyl methanesulfonate to the seeds, which alkylated all guanine base leading to not link with cytosine base. Then, screen the desired plants that contain the desired mutation. Grow the selected candidate to maturity and retest phenotype in the next generation (figure 15).

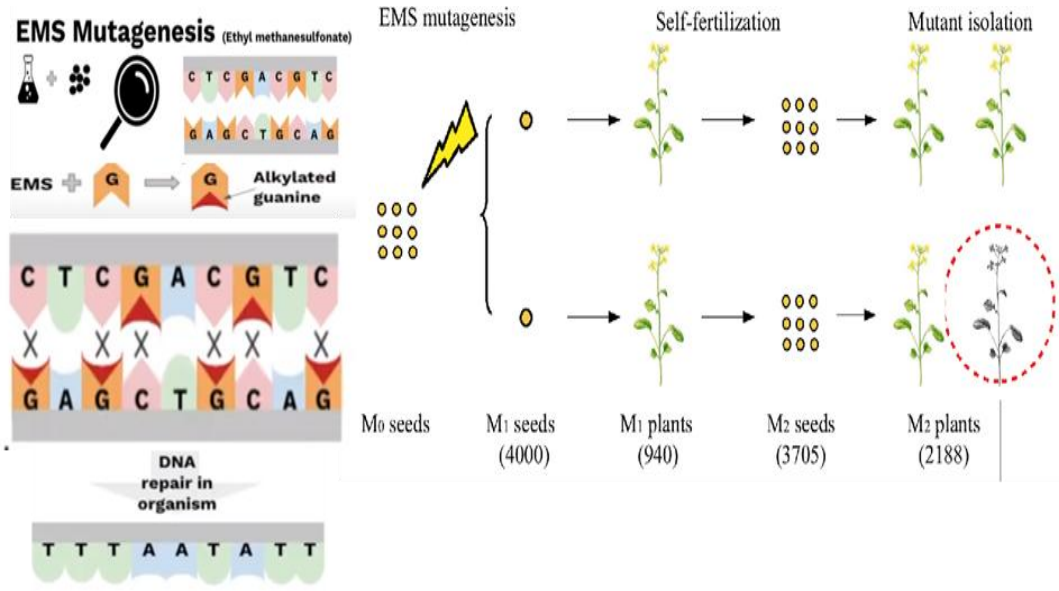

Figure 15. Mutagenesis of Arabidopsis (Mizuno et al. 2014).

3. Mechanism of mutagenesis

Mutagenesis occurs due to DNA replication errors, damage, and lab techniques. Here, we break mutagenesis down into endogenous and exogenous causes (Durland and Ahmadian-Moghadam 2022).

3.1. Endogenous mutagenesis

Endogenous mutagenesis refers to the process by which mutations occur naturally within an organism's own cells, without the influence of external factors like radiation or chemicals. These mutations arise due to internal cellular processes that can damage DNA or lead to errors during DNA replication or repair. Here are some of the key sources of endogenous mutagenesis:

a. Errors in DNA replication

Our body possesses high and low-fidelity DNA polymerases. Although the error rate is minimal, high-fidelity polymerases and mismatch repair (MMR) mechanisms still make 1 in 10^6 to 10^8 base substitution per cell per generation. Furthermore, some errors occur due to replication slippage at repetitive sequences, which can lead to insertions and deletions. A mutation results if these errors are not repaired before the next round of DNA replication.

b. Errors in DNA repair mechanisms

The DNA damage response (DDR) is a group of mechanisms that sense DNA errors and promote repair. Errors in repair or mutations affecting the DDR network cause different cancers. There are multiple DNA repair mechanisms, including MMR, base excision repair (BER), nucleotide excision repair (NER), translesion synthesis (TLS), homologous recombination (HR), and nonhomologous end-joining (NHEJ) pathways. A cell is predisposed to DNA damage if these repair mechanisms disappear. For example, xeroderma pigmentosum (a rare autosomal recessive skin disorder that makes a person highly prone to developing skin cancer) is caused by a mutation in the NER pathway, resulting in a build-up of UV-associated damage. The TLS repair system and NHEJ are of interest to endogenous mutagenesis. During DNA replication, high-fidelity polymerases have difficulty passing damaged bases (eg, pyrimidine dimers or crosslinked DNA), which stalls DNA replication. Failure to restart replication can result in double-stranded breaks, chromosomal rearrangements, and cell death. Therefore, it is o en beneficial to circumvent these replicative arrests to promote cell survival. One mechanism to accomplish this is the TLS system. TLS involves DNA polymerases with larger active sites that allow them to tolerate and bypass DNA lesions. However, it comes at the expense of lower replication fidelity and high error rates, resulting in a greater likelihood of base substitutions.Otherwise, NHEJ is a repair mechanism for double-stranded DNA breaks. It brings 2 ends of DNA fragments together and does not require homologous sequences. Therefore, this can result in deletions and insertions.

c. Spontaneous base deamination

Base deamination is when a nucleotide base loses an amine group, effectively changing the nucleotide. The following are major deamination reactions: cytosine to uracil (U), adenine to hypoxanthine, guanine to xanthine, and 5-methyl cytosine (5mC) to thymine. If these alterations are not repaired, there can be a change in the DNA sequence. For example, if cytosine deaminates to uracils, there are new A:T mutations in 2 replication events. This G:C to A:T transition accounts for 33% of the single-site mutations resulting in human hereditary diseases.

d. Oxidative DNA damage

Reactive oxygen species (ROS) are a byproduct of the electron transport chain (ETC) and other cellular processes in normal cell physiology. ROS serves essential cellular roles, including redox signaling and immune defense. However, in high quantities, ROS damages a cell and its DNA. There have been 100 different oxidative base lesions and 2-deoxyribose modifications described. One example is the oxidation of carbon #8 of guanine, forming 8-oxoguanine (8-OG). 8-OG incorrectly pairs with adenine (instead of cytosine), resulting in a G:C to A:T transition. Excess ROS are known to be associated with many diseases, including Alzheimer's disease, cancer, and heart failure.

e. Base methylation

S-adenosylmethionine (SAM) is used as a methyl donor during physiologic DNA methylation. At a concentration of $4x10^{-5}$M, SAM can generate over 4,000 methylated base changes per cell per day. The significance of this can be depicted in the methylated products O6-methylguanine and O4-methylamine. These highly mutagenic products result in G:C to A:T and T:A to C:G transition mutations.

f. Abasic sites (ie, apurinic and apyrimidinic sites)

Approximately 10,000 abasic sites are made daily by spontaneous hydrolysis or DNA glycosylase cleavage. Abasic sites are unstable and commonly removed by endonucleases. In other cases, they are repaired by TLS polymerases. If these damage sites are not corrected, they may result in mutagenesis.

3.2. Exogenous mutagenesis

Exogenous mutagenesis refers to the process by which mutations are induced in an organism's DNA by external environmental factors, rather than arising from internal cellular mechanisms. These external agents are referred to as mutagens and can include physical, chemical, or biological factors that cause DNA damage, leading to mutations. Here are the major categories and mechanisms of exogenous mutagenesis:

a. Ionizing radiation (IR)

IR comes from the soil, radon, medical devices, and cosmic radiation, among others. It can damage DNA directly (eg, DNA strand breaks) or indirectly (eg, radiolysis of water molecules

producing ROS). Ionizing radiation can generate a range of nucleotide base lesions, similar to that discussed for ROS, resulting in mutagenesis.

b. Ultraviolet (UV) radiation

UV light falls between 100 to 400 nm, with the most harmful radiation at lower wavelengths- UV light damages DNA through direct and indirect energy transfer (to nearby molecules. The 2 main products of UV damage are pyrimidine dimers and pyrimidine pyrimidone photoproducts. These byproducts distort the DNA helix, requiring the NER system or the error-prone TLS polymerases to bypass them. Pyrimidine dimers cause C:G to T:A, T:A to C:G, and tandem CC to TT transition mutations.

c. Alkylating agents & Aromatic amines

Alkylating agents (eg, nitrogen mustard gas, methyl methanesulfonate [MMS], ethyl methanesulfonate [EMS], N-ethyl-N-nitrosourea [ENU]) have a high affinity for nitrogens on nucleotide bases, mainly N3 of adenine and N7 of guanine. MMS, for example, reacts with adenine and guanine to produce N3-methyladenine and N7-methylguanine, respectively. These methyl products are susceptible to N-glycosidic bond cleavage that can create abasic sites. Aromatic amines (eg, 2-aminofluorene, previously used in insecticides) are metabolized by the CYP450 system and converted into alkylating agents. These products primarily cause lesions to the C8 position of guanine. C8-guanine lesions are known to give rise to base substitutions and frameshi mutations.

d. Polycyclic aromatic hydrocarbon (PAH)

AHs (eg, dibenzo[a,l]pyrene, naphthalene, anthracene, and pyrene) are carbon compounds with 2 or more aromatic rings. They are commonly present in tobacco smoke, automobile exhaust, charred food, and combustion products of fossil fuels and organic matter. The CYP450 enzymes convert PAHs into reactive DNA intermediates that intercalate into DNA, ultimately forming a DNA adduct (a segment of DNA bound to a cancer-causing chemical). This results in DNA damage and, thus, can cause mutagenesis.

e. Crosslinking

Crosslinking occurs when 2 nucleotides form a covalent link. Agents commonly associated with crosslinking include cyclophosphamide, cisplatin, and psoralens. Interstrand crosslinking blocks DNA replication, and this requires repair or bypassing. TLS is one of these repair mechanisms and is associated with high substitution rates.

f. Insertional mutagenesis

This is the process by which exogenous DNA integrates into host DNA. Insertional mutagenesis can be natural, mediated by transposons or viruses, or accomplished in a laboratory. Given there is an addition of nucleotides, insertional mutagenesis commonly results in frameshi mutations.

g. Other toxins

Aflatoxin is a naturally occurring toxin from Aspergillus. The CYP450 system metabolizes aflatoxin into an active form that adducts with N7 of guanine, which results in depurination. Aflatoxin is a well-established liver carcinogen that is associated with hepatocellular carcinoma. N- nitrosamines are organic compounds commonly found in tobacco smoke, preserved meats, and the environment. The CYP450 system metabolizes them to form DNA alkylating agents. N-nitrosamines have been implicated in nasopharyngeal, esophageal, and gastric.

h. Laboratory techniques

Different laboratory techniques, including PCR, non-PCR, and gene-editing tools, induce mutagenesis.

4. Mutagen and their types

A mutagen is defined as an agent capable of inducing heritable changes in the DNA sequence, known as mutations, through a process termed mutagenesis. Mutagens are broadly classified into three principal categories based on their nature: physical, chemical, and biological agents (Figure 16).

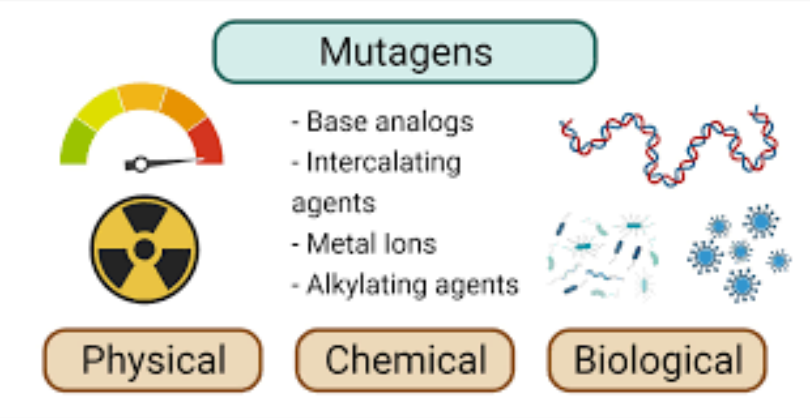

Figure 16. Mutagen and their types. (Sagar Aryal, 2023) Microbe Notes, Accessed from: https://microbenotes.com/mutagens-definition-types-examples/

a. Physical mutagens

Physical mutagens primarily involve forms of energy that can damage DNA:

- Radiation encompasses both ionizing radiations (e.g., Alpha and beta rays, X-rays, gamma rays, electrons, protons, neutrons) and non-ionizing radiations (e.g., ultraviolet light, visible light). Ionizing radiations are highly energetic and can cause severe DNA damage, including deletions, duplications, inversions, and translocations, by breaking the DNA's sugar-phosphate backbone. Non-ionizing radiations, particularly UV light, lead to the formation of nucleotide dimers (e.g., thymine dimers), base deletions, strand breakages, and cross-

linking.

- Heat is another physical mutagen; an increase in temperature can significantly elevate mutation rates by destabilizing both hydrogen bonds and phosphodiester bonds within the DNA molecule.

b. Chemical mutagens

Chemical mutagens are compounds that interact directly with DNA to cause alterations:

- Base Analogues are chemicals structurally similar to natural DNA bases. They can be incorporated into the DNA during replication, leading to incorrect base pairing and subsequent base pair substitutions. Notable examples include 5-bromouracil and aminopurine.
- Intercalating Agents are molecules with heterocyclic ring structures that can insert themselves between adjacent base pairs in the DNA helix. This intercalation distorts the DNA structure, interfering with processes like replication, transcription, and translation, o en resulting in frameshi mutations. Examples include ethidium bromide, proflavin, and acridine orange.
- Metal Ions, such as nickel, chromium, and cobalt, can act as mutagens by generating reactive oxygen species. These species can induce DNA damage and promote DNA hypermethylation, thereby hindering DNA repair mechanisms.
- Alkylating Agents are chemicals that add alkyl groups to DNA bases, leading to mispairing during replication and subsequent base pair substitutions. Examples include nitrous acid and formaldehyde.

c. Biological mutagens

Biological mutagens are living agents or genetic elements that can induce mutations:

- Transposons, or "jumping genes," are mobile genetic elements capable of moving from one genomic location to another. Their insertion into a gene can disrupt its normal functionality. Transposons are categorized by their mechanism of movement: replicative transposons (copy-paste), conservative transposons (cut-paste), and retrotransposons (transposed via an RNA intermediate).
- Viruses can act as mutagens when their genetic material integrates into the host genome, potentially disrupting host gene function. The Rous sarcoma virus, for instance, is known for its ability to induce cancer through such integration.
- Bacteria can also contribute to mutagenesis. Certain bacterial species, like Helicobacter pylori, produce reactive oxygen species that inflict direct damage

to host DNA.

Understanding the various types of mutagens is crucial for comprehending the origins of genetic variation, the mechanisms of carcinogenesis, and the development of strategies for genetic engineering and disease prevention. Types of mutagenesis can be divided into:

- Random Mutagenesis: Mutation can be generated randomly at any region of DNA sequence using a variety of mutagens like UV-irradiation, X-ray, chemicals.
- Sited-directed Mutagenesis: induces site specific changes in DNA. It is a method that target a specific gene or specific nucleotide.

5. Difference between mutation and mutagenesis

Table 5 describes difference between mutation and mutagenesis.

Table 5. Difference between mutation and mutagenesis. Table created by the author.

Mutation	**Mutagenesis**
A mutation is a permanent change in the DNA sequence of an organism.	Mutagenesis is the process of deliberately inducing mutations in an organism's DNA for research or practical purposes.
Mutations can occur naturally as a result of DNA replication errors, environmental factors, or genetic recombination.	Mutagenesis is typically induced artificially, although it can also occur naturally due to exposure to mutagenic agents.
Mutations can result from various factors, including spontaneous errors in DNA replication, radiation, and chemical agents.	Mutagenesis is primarily caused by intentionally introducing mutagenic agents, such as chemicals or radiation.
Mutations can occur without human intervention, and they are not necessarily deliberate or controlled.	Mutagenesis is a controlled and deliberate process, often used in research or breeding programs.
Mutations can lead to genetic diversity and evolution; they may also be responsible for genetic diseases.	Mutagenesis is used to create specific mutations for scientific studies, to study gene function, or to develop organisms with desired traits.
Mutations can be beneficial, harmful, or neutral in their effects on an organism's traits or survival.	Mutagenesis is carried out with specific goals, such as studying gene function or creating organisms with desired characteristics.

6. Site-directed mutagenesis

Also known as Site specific Mutagenesis. SDM is a molecular biology technique in which a mutation is created at a defined site in a DNA molecule known as Plasmid.

Or

Powerful technique where site specific changes in DNA sequences are produced in vitro for instance to change an amino acid residue into another by changing the codon sequence within the gene sequence (Madhavan et al. 2017).

There are many reasons to make specific DNA alterations (insertions, deletions and substitutions), including:

- To study changes in protein activity that occur as a result of the DNA manipulation.
- To select or screen for mutations (at the DNA, RNA or protein level) that have a desired property.
- To introduce or remove restriction endonuclease sites or tags.

During experimental procedure

- Require knowledge of the DNA sequence.
- Require ability to synthesize Oligonucleotides (primer).
  - Conditions require for primer
    - Primer must contain the mutation.
    - Mutation should be in middle of the primer.
    - Primer should be 24-25 nucleotides long and have GC contents of at least 40%.
    - The melting temperature should be >78°C.
    - The 3'-end of the primer has to end on a G or C.

As powerful tool for protein engineering

Site-directed mutagenesis, also known as oligonucleotide-directed mutagenesis, is a fundamental technique in recombinant DNA technology that allows for the precise modification of proteins by changing a single amino acid within their sequence. This technique is crucial for understanding protein function and creating proteins with altered properties. The process involves three key steps:

i. A double-stranded DNA plasmid containing the gene for the target protein is obtained. The DNA strands are separated, and the specific template strand is isolated.

ii. A short, synthetic DNA primer is designed to be complementary to the target region of the DNA template but includes a deliberate mismatched base at the site where the amino acid change is desired. Despite the mismatch, the primer anneals to the template. DNA polymerase then extends this primer, synthesizing a new DNA strand that incorporates the desired point mutation. For example, a

codon for glutamate (GAA) can be changed to a codon for aspartate (GAC) by altering a single nucleotide (A to C) in the primer.

iii. The newly synthesized strand, now containing the point mutation, is used as a template to create a complementary strand. This results in a double-stranded DNA plasmid that carries the intended mutation. This modified plasmid can then be used to express the protein with the altered amino acid.

7. Types of Site-Directed Mutagenesis

Various types of site-directed mutagenesis exist, each suited for specific applications. Here are the main types (Carter 1986):

7.1. Oligonucleotide SDM

The synthetic primer contains the desired mutation and is complementary to the template DNA around the mutation site so it can hybridize with the DNA in the gene of interest. The single strand primer is them extended using the DNA polymerase, which copies rest of gene. Taking in note that the primer is a chemically synthesized oligonucleotide (7-20 nucleotides long). It is complementary to a position of a gene around the site to be mutated; however, it contains mismatch of the base to be mutated. The gene thus copied contains the mutated site, and is then introduce into a host cell as a vector (plasmid or phage) and cloned.

Hybridization (despite a single base mismatch) is possible by mixing at low temperature with excess of primer, and in the presence of high salt concentration. The addition of 4-deoxyribonucleoside triphosphates and DNA polymerase to occur replication followed by sealed the ends by DNA ligase. The double-stranded DNA containing the mismatched introduced by nucleotide into host cell transformation. It is expected that half of the plasmid or M13 phage should carry wild type sequence while the other half mutant sequence. Finally mutant are selected by Screening and selection method (Figure 17).

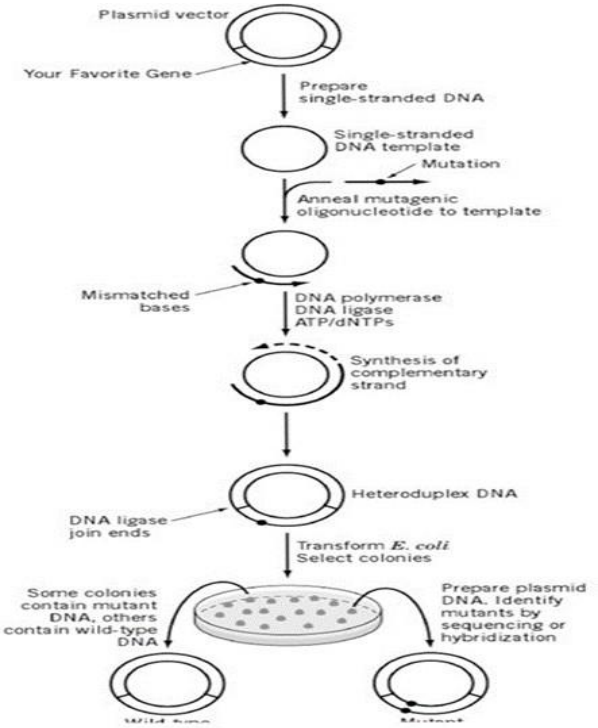

Figure 16. Oligonucleotide SDM (Carter, 1986).

Drawbacks of using the M13 DNA for carrying out SDM

- There is a need to subclone a target gene from a plasmid into M13 and then, a er mutagenesis, clone it back into a plasmid.
- Additional step of transforming enzyme defective E. coli necessary to enrich the yield.
- Lengthy process involving multiple steps.

7.2. Cassette mutagenesis

Cassette mutagenesis that uses a short double stranded oligonucleotide (gene cassette) to replace the fragment of target DNA. Technique for altering a protein sequence at DNA level by replacing section of genetic information with alternative sequence. It differs from the single oligonucleotide in that a single gene cassette can contain multiple mutations and also does not involve Primer extension by DNA polymerase. Cassette mutagenesis is possible if the fragment of the gene to be mutated lies between two restriction enzyme cleavage sites (The target DNA is cut with restriction enzymes, such as EcorRI and Hind111). It is complementary to a position of a gene around the site to be mutated; however, it contains mismatch of the base to be mutated (Figure 18).

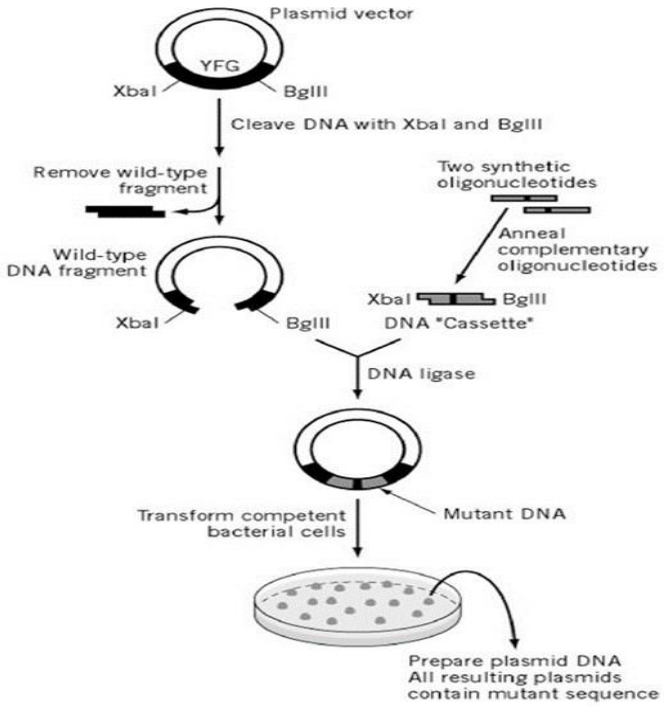

Figure 18. Cassette mutagenesis (Carter, 1986).

7.3. PCR-SDM

PCR and non-PCR techniques are in vitro methods for performing SDM. In vitro synthesis has 4 essential components: DNA template, modified primers, deoxyribonucleic nucleoside triphosphates (dNTPs [ie, dATP, dCTP, dGTP, dTTP]), and DNA polymerase (thermostable or thermolabile). The polymerase chain reaction technique utilizes thermostable DNA polymerases (eg, Taq, Pfu, and Vent), at least 2 primers, and multiple heating and cooling cycles (average 30 cycles). Each cycle has 3 phases: denaturation (approximately 95 degrees Celsius), annealing (about 55 degrees Celsius), and extension (about 72 degrees Celsius). The denaturation phase separates the template DNA molecule into 2 single-stranded molecules. The annealing phase is when the modified primer base pairs at the sequence of interest.

For PCR-based mutagenesis point mutations, nucleotide changes are introduced in the middle of the primer sequence.

- To create deletion mutations, primers must border the region of target DNA to be deleted on both sides and be perfectly matched to their annealing (or template) sequences.
- To create mutations with long insertions, a stretch of mismatched nucleotides is added to the 5′ end of one or both primers, while for mutations with short insertions, a stretch of nucleotides is designed in the middle of one of the primers.

Lastly, the extension phase extends the annealed primers according to the template strand. The advantage of PCR is that it works more favorably with different DNA templates (ie, single or

double-stranded DNA and GC-rich regions) and produces millions of copies of a target gene. The disadvantage is there is less sequence fidelity (ie, thermostable polymerases are more error-prone than thermolabile polymerases). The non-polymerase chain reaction technique utilizes thermolabile DNA polymerases, 1 primer, and a constant reaction temperature (eg, 37 degrees Celsius). The DNA template is denatured with an alkali solution or heat in this strategy. The template is annealed with the modified primer, and DNA synthesis occurs at 37 degrees Celsius. Though single or double-stranded DNA templates can be used, more favorable outcomes occur if a single-stranded DNA template is used (this increases the success of annealing). This process produces a hybrid DNA molecule (ie,1 template strand and 1 newly synthesized mutant strand), which can be transformed or infected into *E. coli*. Once in *E.coli*, the mutant DNA and wild-type DNA can be segregated. Non-PCR has greater sequence fidelity than the PCR method; however, only 1 DNA strand is generated (rather than millions) (Ling and Robinson 1997) (Figure 19).

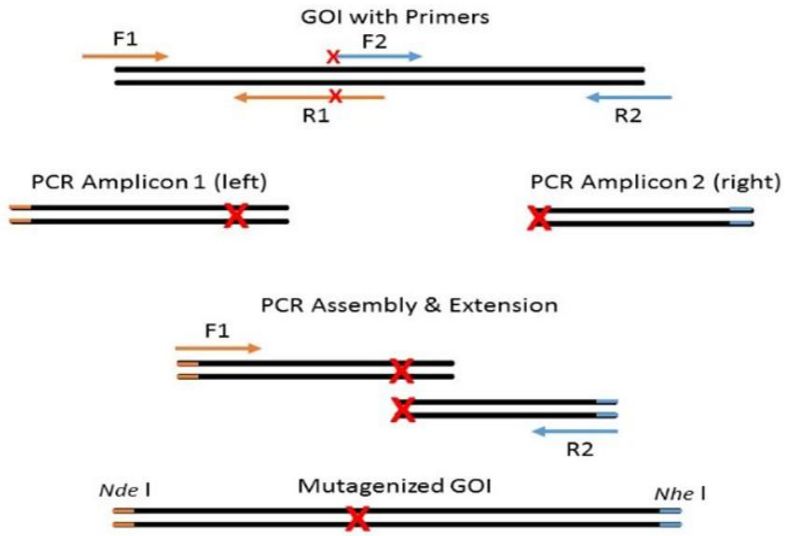

Figure 19. PCR-SDM (Ling and Robinson, 1997).

7.4. TALENS vs. Zinc Finger Nucleases

Various tools have emerged to facilitate precise genetic editing, including TALENs (Transcription Activator-Like Effector Nucleases) and ZFNs (Zinc Finger Nucleases), which operate using similar principles but differ in design and specificity. Both TALENs and ZFNs are engineered proteins designed to target specific DNA sequences, creating double-stranded DNA breaks (DSBs) at precise locations in the genome. These breaks are then repaired by the cell's natural repair mechanisms, which can result in either non-homologous end joining (NHEJ) or homologous recombination (HR). NHEJ may lead to small insertions or deletions (indels), potentially disrupting a gene's function, while HR allows for the introduction of new DNA sequences.

Both TALENs and ZFNs consist of two major parts:

1. DNA binding domain: This part is responsible for recognizing and binding to the specific target DNA sequence.

2. DNA cleavage domain: This part, usually the FokI nuclease, cuts the DNA. FokI needs to dimerize (pair with another FokI molecule) to create a break in the DNA, meaning that two TALEN or ZFN proteins must bind on either side of the target sequence for DNA cleavage to occur (Jabalameli et al. 2015)(Jabalameli et al. 2015) p.125.

Background differences in DNA Binding

The key difference between TALENs and ZFNs lies in how they recognize their DNA target:

- TALENs: The DNA-binding domain in TALENs consists of TALE (Transcription Activator- Like Effector) proteins. Each TALE protein recognizes a single base pair in the DNA. This single-base recognition gives TALENs a high level of specificity because each protein can be designed to bind precisely to its target sequence.
- ZFNs: Each zinc-finger domain recognizes a specific DNA triplet, allowing ZFNs to bind sequences ranging from 6 to 18 base pairs (Upta et al. 2012). The DNA-binding domain is composed of zinc finger proteins, with each zinc finger recognizing 3-4 base pairs. Because zinc fingers recognize multiple bases at once, ZFNs can be less specific than TALENs, especially when recognizing sequences in which zinc fingers target overlapping base pairs.

8. Plant genetic improvement technologies

8.1. Introduction

Plants are a vital part of our ecosystem, providing us with oxygen, food, and resources. Over the years, scientists and farmers have been working on various methods to improve the growth and development of plants. Plant improvement involves a range of techniques aimed at enhancing desirable traits in plants, such as increased yield, disease resistance, and improved nutritional value. One common method of plant improvement is selective breeding. This process involves carefully selecting plants with desirable traits and cross-pollinating them to create offspring with a combination of these traits. Through selective breeding, farmers have been able to develop crop varieties that are more resistant to pests, drought, and other environmental stresses.

Another technique used in plant improvement is genetic engineering. This cutting-edge technology allows scientists to directly modify the DNA of plants, introducing specific genes to enhance desired characteristics. Genetic engineering has led to the development of genetically modified organisms (GMOs), which can have traits such as increased yield, pest resistance, and improved nutritional content.

In addition to selective breeding and genetic engineering, plant improvement also involves other methods such as tissue culture and marker-assisted selection. Tissue culture is a technique

where small pieces of plant tissue, such as cells or embryos, are grown in a laboratory under controlled conditions. This method enables the mass production of plants with desirable traits, such as disease resistance. Marker-assisted selection, on the other hand, involves using molecular markers to identify plants with specific genes or traits of interest, allowing for more precise and efficient breeding. About 500 out of over one million plant species have been domesticated or brought under cultivation as crop plants through centuries of human selection (Philippe et al. 2017).

It turns out it is super common for plants to play around with polyploidy, the technical name for having more than two full sets of chromosomes. Even in the food we eat. Maybe especially in the food we eat.

Examples:

- Potatoes can have four sets of chromosomes= 4n
- Kiwis can have six sets of chromosomes= 6n
- Strawberries can have eight sets of chromosomes= 8n
- Extra copies of genes can be kind of a blank slate for evolution to play around with over time. With other copies of those genes still doing their thing, the extra copies of genes are then free to change things up and acquire new functions.

The publication details the Alliance for a Green Revolution in Africa's (AGRA) contributions to transforming African agriculture, particularly through its Program for Africa's Seed Systems (PASS) (Tongoona et al. 2019). This program has significantly increased the number of active plant breeders within National Agricultural Research Systems (NARS), facilitated the development and release of numerous improved crop varieties, and fostered local seed companies and distribution networks. AGRA's investment in training plant breeders at MSc and PhD levels in African universities has been instrumental, resulting in the training of 495 crop improvement scientists focusing on 15 priority staple crops, including maize, rice, sorghum, wheat, and various legumes and root crops. A key development in this educational framework is the Improved MSc for Cultivar Development in Africa (IMCDA) e-curriculum, which integrates modern breeding techniques, robust data analysis, and internships to cultivate product development skills.

Case studies from West, East, and Southern Africa are presented, illustrating breeding activities for key crops such as maize, sorghum, pearl millet, rice, cowpea, groundnut, sweet potato, and cassava. These case studies highlight practical breeding methodologies, challenges encountered by breeders, and successful outcomes in crop improvement. For instance, the development of the Bondofa maize hybrid in Burkina Faso, noted for its drought tolerance and disease resistance, exemplifies a successful outcome that led to a national maize sufficiency program. Similarly, efforts in groundnut breeding in Ghana focused on introgressing rosette virus

resistance genes into local landraces, while cassava breeding programs aim to develop high-yielding, CMD-resistant varieties. The application of male sterility systems, such as Cytoplasmic-genetic male sterility (CMS) in sorghum and the A4 system in pearl millet, is also discussed as a method for large-scale hybridization.

8.2. Genetically modified organisms (GMOs)

Genetically modified organisms (GMOs) have become a significant topic of discussion in the field of plant improvement. Scientists have been using genetic engineering techniques to modify the DNA of plants to enhance their traits and improve their overall performance. This process involves introducing specific genes from other organisms into the plants, giving them new characteristics that they wouldn't naturally possess.

One of the main objectives of genetically modifying plants is to increase their resistance to pests and diseases. By inserting genes that produce toxins harmful to pests or make the plant less appealing to insects, scientists can reduce the need for chemical pesticides. This not only benefits the environment but also helps farmers protect their crops more effectively.

GMOs also offer the potential for increased crop yields. Through genetic modifications, scientists can enhance traits such as drought tolerance, nutrient uptake efficiency, and resistance to harsh environmental conditions. These modifications allow plants to grow better in challenging climates and produce higher yields, which can contribute to global food security (Hug 2008).

Another area where GMOs have shown promise is in improving the nutritional content of crops. Scientists can introduce genes that enhance the levels of essential nutrients in plants, such as vitamins and minerals. This has the potential to address nutritional deficiencies and improve the health outcomes of populations that rely on these crops as their primary food source. These advancements in plant improvement techniques are crucial for ensuring food security and addressing the needs of a growing global population.

Despite the potential benefits, GMOs also raise concerns. Some worry about the potential environmental impacts of genetically modified plants, such as the spread of modified genes to wild plant populations. Additionally, there are concerns about the long-term effects on human health, though extensive scientific research has shown no evidence of harm to date. To address these concerns, many countries have established regulations for the cultivation and labeling of GMOs. These regulations aim to ensure the safety of genetically modified crops and provide consumers with the information they need to make informed choices.

Examples:

Golden Rice

- Nutrient content of a crop plant can be improved by metabolic engineering (Rice, maize, wheat, tomatoes) in order to enhance the level of certain nutrients of beta carotene to have higher levels of Vitamin A.

- GM rice was accomplished by inserting phytoene synthase gene from daffodils and phytoene desaturase gene from *Erwinia* bacteria.

BT crops

- Bt Crops are transgenic crops that produce the same toxin as the bacterium Bacillus thuringiensis in the plant cell, thereby, protecting the crops from pests. The bacterium secretes specific proteins known as “cry proteins” that are toxic to insects. A few of the Bt crops include cotton, brinjal, corn, etc.

Glucosinolate Broccoli

- Plant improvement led to some beneficial effects to the broccoli by exhibited higher levels of glucosinolate glucoraphanin resulting in improving the human metabolism and reducing the fatty acids associated with inflammation.

Enhanced strawberries

- Strawberries have been modified to provide 3 times more of vitamin C.
- GalUR gene codes for an enzyme that converts a protein in the plant to vitamin C. Similarly, another gene was found in the Arabidopsis thaliana responsible for the same function.
- The incorporation A. thaliana gene into strawberry plant produces 3 times as much vitamin C.

New compounds in plants

- Polyhydroxyalkanoates produced in plants, a way to synthesize a biodegradable polymer. Provitamin A carotenoids in food such as corn, rice, and wheat.
- Folate or vitamin B produced in tomato fruits. Salt tolerant tomatoes.
- Enhanced carrots to produce a vaccine against hepatitis B.

8.3. Application of mutation breeding

To understand the application of mutation breeding for improving plant products, it is essential to first grasp the concept of mutation. A mutation refers to a change in the genetic composition of an organism, which can occur naturally or be induced deliberately, as in plant breeding. When mutations are introduced, they can alter a plant's chromosomal structure, potentially leading to desirable traits. However, mutation breeding is a delicate process, as small errors in manipulation can result in crop failure or unintended characteristics that may compromise the plant's genetic integrity (Wang et al. 2006). Despite these challenges, mutation breeding has been successful, particularly in enhancing cereals and rice. For instance, dwarf varieties have been developed to shorten the growing period.

Biotechnology plays a crucial role in plant breeding by facilitating genetic transfers to improve plant species. Techniques like vectors and pathogens are now used to introduce genetic material into host cells. Common methods include (Dwivedi et al. 2016):

i. Selection: Genetically distinct progeny rows are evaluated, and their performance and yield are compared with commercial varieties.

ii. Hybridization: Cross-pollination is employed to combine desirable traits from different plants.

iii. Polyploidy: Chemicals are used to increase chromosome numbers in plants.

iv. Induced Mutation: Radiation or chemicals are applied to induce mutations.

With advancements in biotechnology, essential minerals and components can be introduced into plants through biological methods. These techniques, when applied properly, can yield the desired plant traits while preserving the plant's natural characteristics.

Chemical mutagenesis, unlike genetic engineering, induces random mutations throughout the genome. While this approach does not classify resulting plants as genetically modified organisms (GMOs), it can lead to both advantageous and disadvantageous alterations in desired traits. In contrast, genetic engineering involves the precise insertion of specific genes to confer targeted characteristics.

A demonstration of chemical mutagenesis involves the application of ethane methane sulfonate (EMS) to an F1 hybrid tomato variety, resulting in a mutant population displaying a range of phenotypic variations. These variations include modifications in flower morphology, fruit ripening, fruit shape and size, and chlorophyll pigmentation. Key considerations in experimental design for chemical mutagenesis include: (i) Assessment of seed germination rates are crucial for determining the impact of the mutagen and establishing baseline viability; (ii) The concentration of the mutagen must be carefully calibrated to achieve an optimal mutation rate without causing excessive mortality. The ideal dosage typically results in 50% seed survival.; (iii) Selection for specific traits of interest is performed based on phenotypic or genotypic characteristics, such as tolerance to salinity or herbicides, or alterations in sugar or oil content.

8.4. Application of genome editing technologies in the improvement of plant

In the context of genetically modified crops, agriculture is a continuously evolving field that requires constant scientific advancements to meet societal needs. As part of this ongoing progress, genome editing (GE) technologies have emerged as crucial tools for crop improvement and are under continuous refinement.

Meganucleases, also known as homing endonucleases, were the first class of site-specific enzymes applied in genome editing (GE). These enzymes have dual functional domains that enable them to bind to DNA and induce double-stranded breaks (DSBs) at specific target sequences. Meganucleases are typically restricted to sequences of 10–40 base pairs (bp),

limiting their range of target sites (Carroll 2017).

One of the earliest examples is the intron-encoded endonuclease I-SceI, isolated from Saccharomyces cerevisiae, which cleaves an 18 bp DNA sequence to induce DSBs (Stoddard, 2005; Takeuchi et al., 2011). However, most meganucleases are species-specific and do not function effectively across diverse biological kingdoms. Thus, in plant GE, meganucleases are often derived from plant species. For instance, the endonuclease I-CreI from *Chlamydomonas reinhardtii* has been utilized to introduce DSBs in various plant species (Antunes et al. 2012).

Despite their precision, meganucleases have a major limitation: they are restricted to a narrow range of target sequences. Protein engineering could potentially broaden their applicability, but this approach is labor-intensive and time-consuming. Therefore, more versatile GE technologies like Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) are o en preferred due to their flexibility and efficiency.

However, ZFNs require a customized nuclease for each target, making the process expensive and time-intensive. This limitation restricts their use to single-gene modifications, which complicates efforts to edit multiple genes or complex traits.

Despite these challenges, ZFNs have been successfully employed to create transgenic crops like tobacco, maize, rice, and soybean. ZFNs have enabled precise gene insertions, facilitating the stacking of multiple desirable traits in crops, as demonstrated in maize. They have also been utilized to improve traits such as yield and stress resistance in crops like rice and maize, although more advanced GE tools have largely surpassed ZFNs in terms of efficiency and ease of use (Martínez-Fortún et al. 2017).

For instance, TALENs were used to disrupt the Ossweet14 gene in rice, creating bacterial blight-resistant plants. Additionally, TALENs have been employed to improve the nutritional profiles of crops, such as developing high-oleic acid soybeans by targeting fatty acid desaturase (fad) genes (Du et al. 2016). Despite their flexibility, TALENs are limited by their large size, which complicates their use in editing multiple genes simultaneously. Furthermore, the production of TALE repeats can be labor- intensive and has shown inconsistencies in targeting efficiency across species.

Furthermore, CRISPR/Cas9 system has been widely applied in agriculture, generating transgenic crops such as rice, maize, wheat, and tomato (Ricroch et al. 2017). CRISPR/Cas9 has been used to develop herbicide-resistant rice, drought-tolerant maize, and pathogen-resistant citrus (Shi et al. 2017). Additionally, the technology has been employed to improve crop yield, nutritional quality, and stress resistance (Li et al. 2016).

The adoption of transgenic crops, particularly those developed using genome-editing (GE) technologies, has brought both economic and environmental advantages. These benefits include higher crop yields, lower carbon dioxide emissions, increased income for farmers, and improved consumer health . Herbicide-tolerant crops have acquired additional agronomic traits beyond general pest resistance, such as the ability to boost net yield while maintaining pest resistance.

The use of stacked traits, made possible through GE technologies, has significantly enhanced the agronomic qualities of herbicide-tolerant and other transgenic crops, transforming the market for both producers and consumers. Insect-resistant crops, in particular, have introduced valuable agronomic traits by improving yield potential and offering an alternative to insecticides, which can otherwise have harmful environmental impacts (Brookes and Barfoot 2020).

Chapter IV: Genetic engineering in medical applications

1. Introduction

Cells are the fundamental building blocks of all living organisms, with genes located deep within them. Genes are segments of DNA that carry genetic information and provide instructions for protein synthesis, which is crucial for body structure and function. Each individual possesses approximately 20,000 genes, with two copies of each, one inherited from each parent. Minor genetic variations contribute to differences in individual appearance and health.

Scientists apply genetic engineering in the biomedical field via manipulating an organism's DNA to diagnose, treat, or prevent diseases, revolutionizing medicine through gene therapy, personalized medicine, and the production of therapeutic proteins.

2. Diagnostic tools

2.1. Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) is a revolutionary technique in modern molecular biology, developed by Nobel laureate Kary Mullis in 1984. This in vitro method allows for the precise amplification of a target DNA molecule, making it easier to handle and examine using routine molecular biological methods. PCR has significantly contributed to the advancement of biological sciences since its inception, with the first PCR machine introduced to the market in 1988. Its wide-ranging applications have led to the development of numerous PCR variants over the past few decades, playing a crucial role in projects such as the Human Genome Project.

The PCR process involves three main steps: denaturation, annealing, and polymerization. First, the target DNA strands are separated through denaturation. Then, oligonucleotide primers hybridize (anneal) to the single-stranded templates. Finally, a thermostable DNA polymerase typically derived from thermophilic microorganisms, extends these primers by adding nucleotides (dNTPs) complementary to the template sequence. This cycle is repeated multiple times, with each cycle potentially doubling the amount of target DNA.

PCR's ability to make millions or billions of copies of a specific DNA region has made it an invaluable tool in various fields, including biotechnology, genomics, diagnostics, and systematics. Researchers can use PCR to amplify genes for functional studies, forensic scientists can analyze genetic markers from crime scenes, and the amplified DNA can be used for sequencing, gel electrophoresis, or cloning into plasmids for further experiments. The versatility and power of PCR have cemented its position as a cornerstone technique in modern molecular biology and biotechnology (Shahzad et al. 2020).

a. Steps of PCR

As shown in the figure 19 below, the key ingredients of a PCR reaction are Taq polymerase, primers, template DNA, and nucleotides (DNA building blocks). The ingredients are assembled in a tube, along with cofactors needed by the enzyme, and are put through repeated cycles of heating and cooling that allow DNA to be synthesized. The basic steps are:

- Denaturation (96 °C): Heat the reaction strongly to separate, or denature, the DNA strands. This provides single-stranded template for the next step.
- Annealing (55-65°C): Cool the reaction so the primers can bind to their complementary sequences on the single-stranded template DNA.
- Extension (72 °C): Raise the reaction temperatures so Taq polymerase extends the primers, synthesizing new strands of DNA.
- This cycle repeats [25-35] times in a typical PCR reaction, which generally takes [2-4] hours, depending on the length of the DNA region being copied. If the reaction is efficient (works well), the target region can go from just one or a few copies to billions.

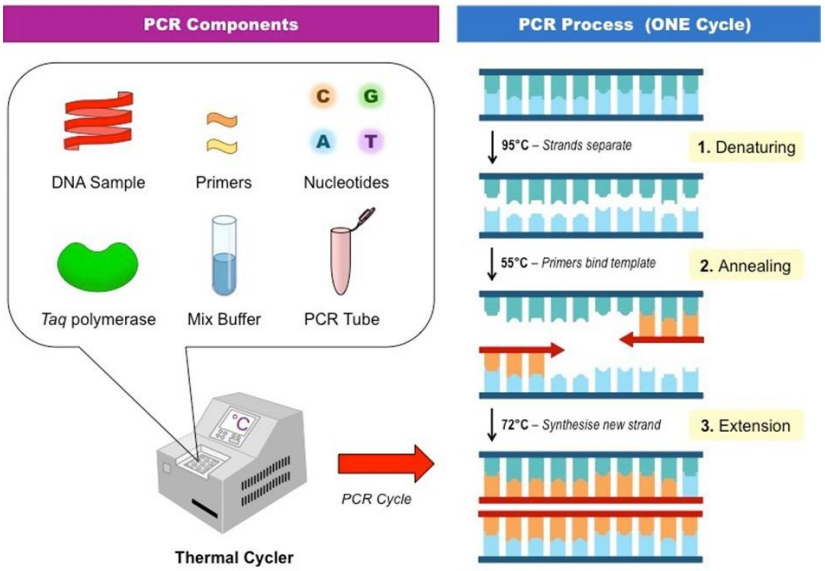

Figure 19. Steps of PCR. Created by the author via BioRender.com.

b. Quantitative PCR

The power of qPCR lies in its ability to calculate the initial template concentration, making it a versatile analytical tool for various applications. These include evaluating DNA copy number, measuring viral load, detecting single nucleotide polymorphisms (SNPs), and performing allelic discrimination. When combined with reverse transcription (RT-qPCR), it becomes an invaluable method for quantifying mRNA expression, serving as the gold standard for confirming microarray

gene expression data, as presented in figure 20 by (Stratagene 2007).

qPCR can be performed in two main ways: endpoint QPCR and real-time QPCR. In endpoint QPCR, fluorescence data are collected a er the amplification reaction is complete, typically a er 30-40 cycles, and this final fluorescence is used to back-calculate the initial template quantity. In contrast, real-time QPCR measures the fluorescent signal while the amplification is still in progress, providing a more dynamic view of the reaction kinetics (Dymond 2013). The qPCR is mainly performed using one of these two essays SYBR Green assay and TaqMan probe.

i. SYBR Green assay

DNA binding dyes such as SYBR Green are cost effective and easy to use, especially for researchers who are new to using QPCR techniques. These same factors make SYBR Green a common choice for optimizing QPCR reactions. When free in solution, SYBR Green displays relatively low fluorescence, but when bound to double-stranded DNA its fluorescence increases by over 1000- fold. The more double-stranded DNA that is present, the more binding sites there are for the dye, so fluorescence increases proportionately to DNA concentration. This property of the dye provides the mechanism that allows it to be used to track the accumulation of PCR product. As the target is amplified, the increasing concentration of double stranded DNA in the solution can be directly measured by the increase in fluorescence signal (Figure below). Compared to probe- based methods, SYBR Green assays are relatively easy to design and optimize. All that is necessary is to design a set of primers, optimize the amplification efficiency and specificity, and then run the PCR reaction in the presence of the dye. One limitation of assays based on DNA- binding dye chemistry is the inherent non-specificity of this method. SYBR Green will increase in fluorescence when bound to any double-stranded DNA (dsDNA). Therefore, the reaction specificity is determined solely by the primers (Figure 20).

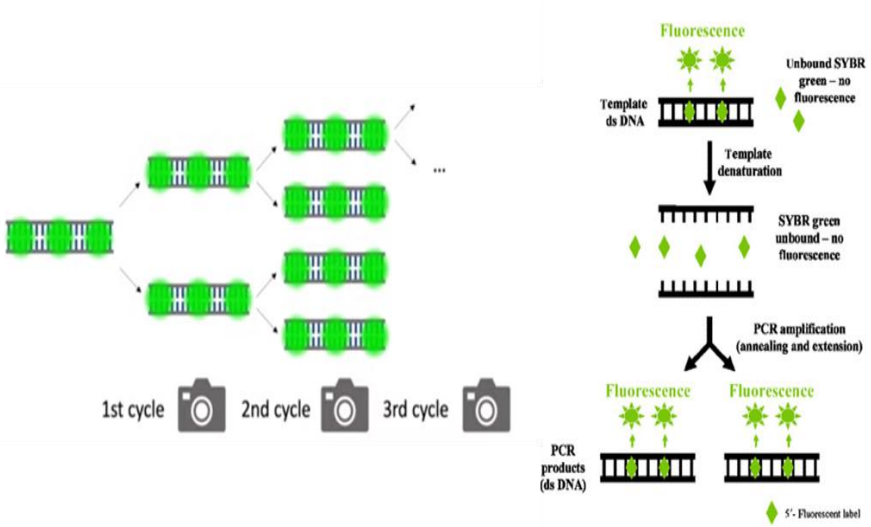

Figure 20. Illustrative image of SYBR Green assay. Created by the author via BioRender.com.

ii. TaqMan qPCR probes

TaqMan qPCR probes are the most widely used and published detection chemistry for QPCR applications. In addition to the PCR primers, this chemistry includes a third oligonucleotide in the reaction known as the probe. A fluorescent dye, typically FAM, is attached to the 5′ end of the probe and a quencher, historically TAMRA, is attached at the 3′ end. As long as the two molecules (reporter and quencher) are maintained in close proximity, the fluorescence from the reporter is quenched and no fluorescence is detected at the reporter dye's emission wavelength. The probe is designed to anneal to one strand of the target sequence just slightly downstream of one of the primers. As the polymerase extends that primer, it will encounter the 5′ end of the probe. Taq DNA polymerase has 5′–3′ nuclease activity, so when Taq DNA polymerase encounters the probe, it displaces and degrades the 5′ end, releasing free reporter dye into solution. Following the separation of reporter dye and quencher, fluorescence can be detected from the reporter dye (Figure 21).

During gene expression analysis multiple genes can compared and the number of cycles a er which the signal intensity reaches the threshold is shown as Ct value. Where Low Ct values go along with high DNA concentrations and high Ct values mean that signals are detected late indicating low DNA concentrations.

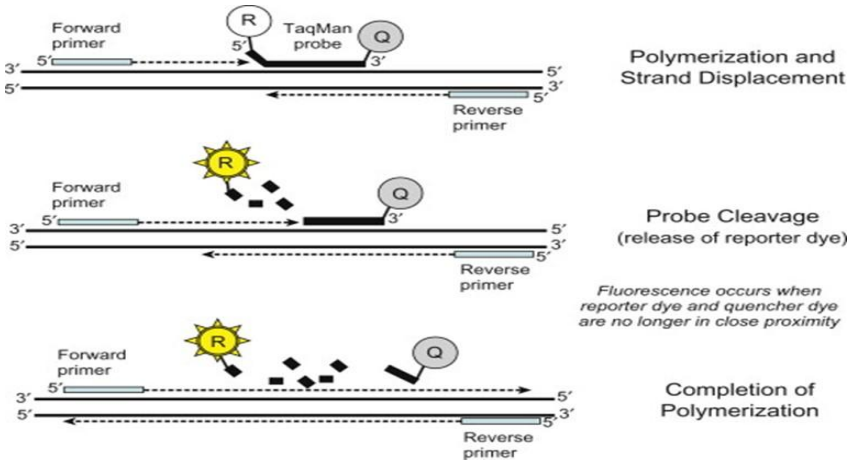

Figure 21. Illustrative image of TaqMan qPCR probes. (Butler, 2012).

c. Reverse transcriptase PCR (RT-PCR)

The isolation of messenger RNA (mRNA) and its subsequent conversion to complementary DNA (cDNA) is a pivotal technique in molecular biology, particularly for gene expression studies. This process involves several precise steps that ensure the accurate analysis of gene activity within cells (Joyce 2002; Lin et al. 2000).

- RNA extraction and mRNA isolation:

The process begins with the extraction of total RNA from cells or tissues, followed by the isolation of mRNA. Affinity chromatography, which utilizes the mRNA's 3' poly-A tail, is

commonly employed for this purpose. The poly-A tail binds selectively to oligo-dT sequences, allowing for the enrichment of mRNA from the total RNA population.

- Reverse transcription:

The isolated mRNA then acts as a template for the synthesis of complementary DNA (cDNA). This reaction is catalyzed by reverse transcriptase, an enzyme derived from retroviruses. The reverse transcriptase synthesizes a cDNA strand that corresponds to the mRNA sequence, converting the single-stranded mRNA into a double-stranded cDNA molecule.

- mRNA elimination:

Once the cDNA is synthesized, the original mRNA template is no longer needed and is subsequently removed. This is typically achieved through methods such as RNase treatment, alkaline hydrolysis, or thermal denaturation, ensuring that only the newly synthesized cDNA remains for further analysis.

- cDNA amplification:

The single-stranded cDNA is then amplified through polymerase chain reaction (PCR). During PCR, the cDNA is subjected to repeated cycles of heating and cooling in a thermal cycler, allowing for the exponential replication of the target sequences. This amplification step is critical for generating sufficient quantities of cDNA for downstream analysis.

2.2. DNA sequencing

DNA sequencing is the process of determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a piece of DNA. Today, with the right equipment and materials, sequencing a short piece of DNA is relatively straightforward.

Sequencing an entire genome (all of an organism's DNA) remains a complex task. It requires breaking the DNA of the genome into many smaller pieces, sequencing the pieces, and assembling the sequences into a single long "consensus." However, thanks to new methods that have been developed over the past two decades, genome sequencing is now much faster and less expensive than it was during the Human Genome Project.

Next-generation sequencing (NGS) is a highly effective genetic diagnostic tool used in disease diagnosis. The use of NGS methods has increased with technological advancements, although the Sanger method remains a traditional approach in genome studies. The foundation of NGS was laid by methods developed by Allan Maxam-Walter Gilbert and Nobel laureate Frederick Sanger. Early first-generation sequencing methods required several days for partial DNA sequencing, whereas current technology allows for sequencing of entire genomes of complex organisms in a single day. Second- and third-generation sequencing methods have

seen improvements in terms of cost, time, and accuracy. Bioinformatics plays a crucial role in interpreting the data obtained from these methods, contributing to the advancement of NGS technology. These developments have heightened interest in studies exploring the relationship between NGS and DNA or RNA in the context of diseases (Eren et al. 2022).

Next-generation sequencing is kind of like running a very large number of tiny Sanger sequencing reactions in parallel. Thanks to this is possible to sequence an entire human genome in one day.

Experimental procedure of NGS

1. Generate smaller DNA fragments, which will ligate with a certain oligonucleotide called adaptors. (Figure 21)
2. The double stranded fragments are denatured.
3. All fragments have attached to different types of binding sites on each end (complementary sequences). All DNA fragments are now applied on ship called flow cell.
4. DNA fragments will at this end hybridize with the flow cell as applied to the plate.
5. The first step in flow cell is DNA amplification by PCR, the complementary sequence to our S-D DNA fragment is synthesized. This process seeing here is done simultaneously for all DNA fragments on the plate.
6. DNA is denatured and the single-strand which not attached to the flow cell oligonucleotides get washed away.
7. The bridge building where the second type oligonucleotide is hybridized to one of this sticky oligonucleotides on the plate.
8. The bridge amplification is a process where DNA is amplified again where a polymerase synthesize the complementary sequence again.
9. DNA is denatured again but both fragments stick to the plate.
10. The DNA fragments will build a bridge again and will have bridge amplification.
11. The whole process will repeat it for several rounds, so one DNA fragment we generate multiple copies by PCR. Reverse and forwards strands are presented from our copy.
12. The reversed strands are cleaved from the plate and then the sequencing begin.
13. • The primer can bind to the oligonucleotide, which is the sequencing binding sites and specific nucleotides fluorescents
14. The fluorescent excite by a laser and the fluorescent signal is obtained due to the

color code.

15. NGS generated billions of reads which are overlaid and compare to reference genome.

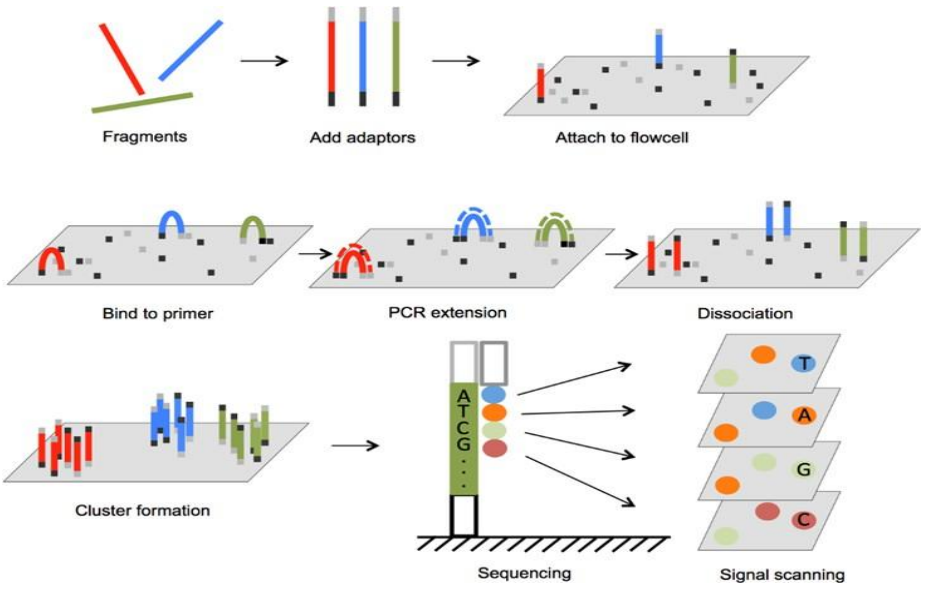

Figure 21. Illustrative image of NGS. (Lu et al., 2016).

3. New drugs and new therapies for genetic diseases: case of gene therapy

Genetic diseases arise from genetic mutations, which are changes such as substitutions, deletions, or duplications of a critical piece or entire section of DNA. Some severe genetic diseases caused by these mutations can be inherited by future generations. Gene Therapy aims to treat diseases by replacing, inactivating, or introducing genes into cells. This can be performed either inside the body (in vivo) or outside the body (ex vivo). Some therapies are considered both cell and gene therapies if they involve altering genes within specific cell types that are then introduced into the body (Dunbar et al. 2018).

3.1. Types of gene therapy

There are several approaches for correcting faulty genes; the most common being the

insertion of a normal gene into a specific location within the genome to replace a non functional gene. Gene therapy is classified into the following two types (Figure 22):

a. Somatic gene therapy

Somatic gene therapy involves the introduction of foreign genes into the somatic (non-reproductive) cells of a patient to treat or mitigate a genetic disorder (figure below). This therapeutic approach targets specific cells in the body, such as liver, muscle, or blood cells, to correct genetic defects, introduce beneficial functions, or suppress harmful genes. Unlike germline gene therapy, which affects reproductive cells and can be passed on to future generations, somatic gene therapy affects only the treated individual. As a result, the genetic modifications made are not heritable and do not influence the patient's offspring (Ledley 1996) (Ledley 1996). Nonetheless, somatic gene therapy has demonstrated significant potential, with several therapies already approved for clinical use. The non-heritable nature of this approach makes it a safer option from an ethical standpoint, as it does not raise concerns about passing genetic changes to future generations, making it a key tool in personalized medicine (Munung et al. 2024).

b. Germ line gene therapy

The therapeutic approach of germline gene therapy aims to target genetic disorders at their root by altering the genome of reproductive cells. Once the modified germ cells are used in reproduction, the corrected or added gene becomes part of the genetic makeup of the offspring, potentially preventing the inheritance of genetic diseases such as cystic fibrosis, Huntington's disease, or muscular dystrophy. The key advantage of this approach is its long-term impact on genetic disease prevention across multiple generations (Wolf et al. 2019).

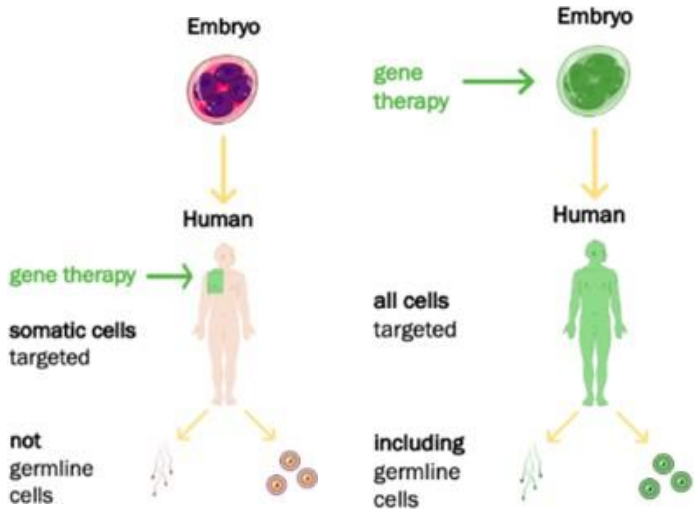

Figure 22. Illustrative image of types of gene therapy. Created by the author via BioRender.com.

3.2. *In vivo* gene therapy

In vivo gene therapy involves the direct transfer of genes into cells within a patient's body, bypassing the need for cell culture outside the body. This approach is particularly useful when dealing with tissues whose individual cells cannot be easily cultured *in vitro*, such as brain cells, or in cases where re-implanting cultured cells back into the patient is inefficient or impractical.

Therapeutic genes are delivered directly to the target tissue or organ, typically using viral vectors or non-viral delivery systems such as liposomes. Viral vectors, such as adenoviruses or adeno-associated viruses, are engineered to carry the desired gene and deliver it to the patient's cells. These vectors are designed to efficiently enter the target cells and incorporate the therapeutic gene into the patient's genome or maintain it as an extrachromosomal element, depending on the vector type and therapeutic goal.

Liposomes, on the other hand, are lipid-based nanoparticles that can encapsulate and protect DNA or RNA molecules during delivery. They serve as non-viral carriers for gene transfer, avoiding some of the immunogenic risks associated with viral vectors.

This approach has shown promise in treating genetic disorders, certain cancers, and diseases affecting the central nervous system, where direct access to the affected tissue is necessary. Clinical trials using in vivo gene therapy have demonstrated its potential for diseases such as cystic fibrosis, hemophilia, and neurological disorders, though challenges such as delivery efficiency, vector safety, and long-term gene expression remain areas of active research (Nishikawa and Hashida 2002).

3.3. *Ex vivo* gene therapy

Ex vivo gene therapy involves transferring genes to cells that are cultured outside the patient's body. In this approach, cells are initially sourced from the patient, allowing for autologous transplantation, which minimizes the risk of immune rejection. Once isolated, the cells, such as hematopoietic or skin cells, are genetically modified to correct specific genetic defects. A er successful transformation, the modified cells are multiplied and reintroduced into the patient's body. This method is particularly advantageous for tissues that can be easily harvested and manipulated outside the body. For example, hematopoietic stem cells can be genetically corrected in culture and then reinfused into the patient, where they can function over an extended period, potentially offering a durable therapeutic effect. This technique has shown promise in treating various genetic disorders and is a key strategy in the development of personalized medicine (Figure 23) (Gowing et al. 2017).

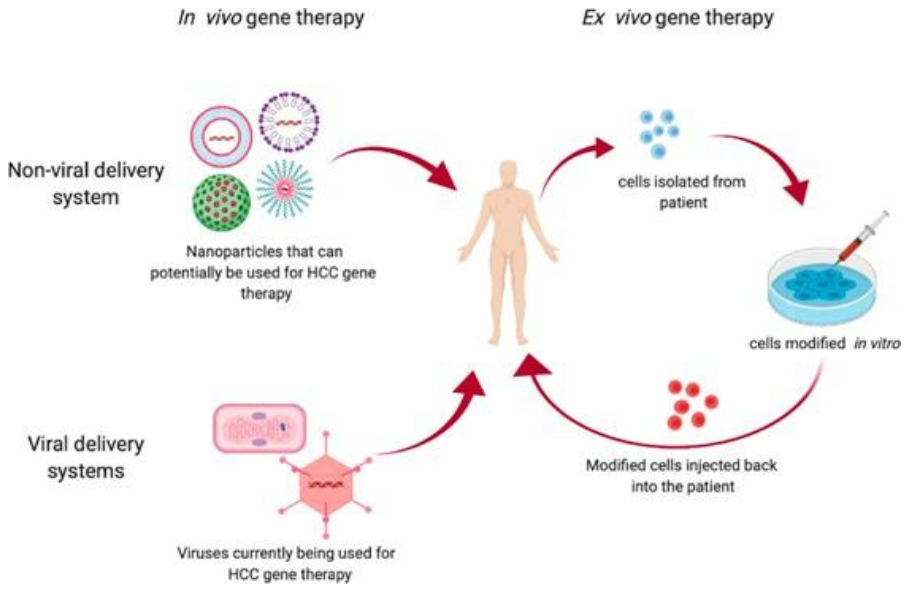

Figure 23. Illustrative image of ex vivo vs in vivo gene therapy. Adapted and modified from (Sarkar, 2019).

3.4. Gene therapy strategies

Gene therapy strategies encompass a variety of approaches designed to treat or prevent diseases by modifying the genetic material within a patient's cells (Wang et al. 2014).

a. Gene augmentation therapy

If a disease is caused by a mutation leading to loss of function, introduction of a functional copy of the gene into the cell will restore the normal function of the gene in another term, it is used to treat inherited disorders (Figure 24). E.g. Hemophilia or Deficiency of ADA.

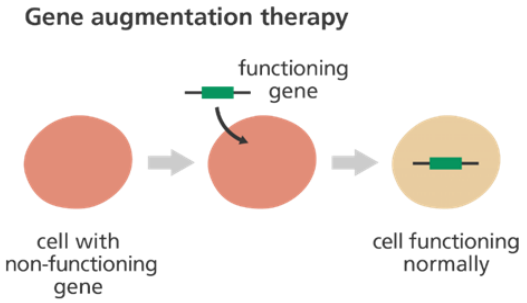

Figure 24. Illustrative image of gene augmentation therapy. Provided by yourgenome, CC BY 4.0). Acceded: https://www.labxchange.org/library/items/lb:LabXchange:a9dc2dad:html:1

b. Targeted mutation correction

It is used to correct a defective gene [by changing the mutated nucleotide] to restore its function which can be done at genetic level by homologous recombination or at mRNA level by using therapeutic ribozymes or therapeutic RNA editing (Figure 25).

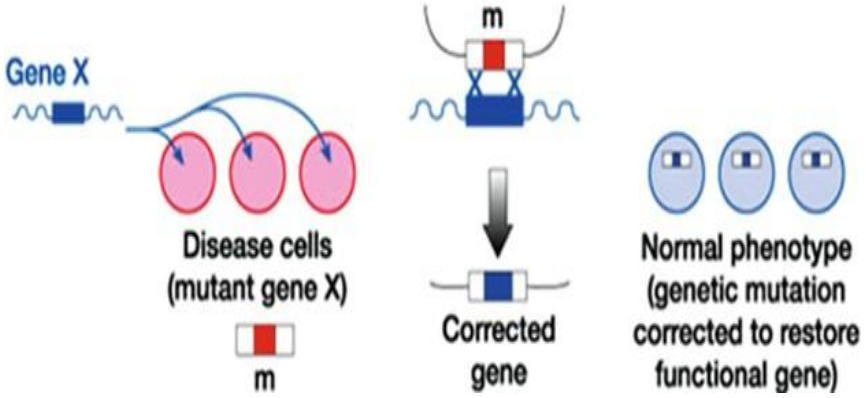

Figure 25. Illustrative image of targeted mutation correction. Provided by yourgenome, CC BY 4.0). Acceded: https://www.labxchange.org/library/items/lb:LabXchange:a9dc2dad:html:1

c. Inhibition of gene expression

The expression of the mutated gene can be blocked at the DNA/RNA/protein level. This particularly useful in cancers by inappropriate expression of a gene or treating infectious diseases (Figure 26).

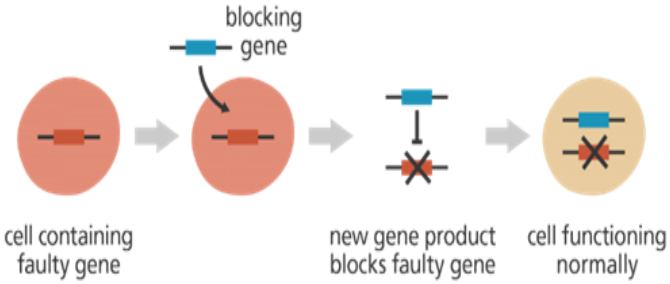

Figure 26. Illustrative image of inhibition of gene expression. Provided by yourgenome, CC BY 4.0). Acceded: https://www.labxchange.org/library/items/lb:LabXchange:a9dc2dad:html:1

d. Gene therapy to achieve pharmacological effects

Introduction of a gene that makes cancer cells susceptible to anticancer drugs. • Introduction of a toxic gene whose expression kills cancer cells.

Genes of cytokines can be introduced into cells of immune system to enhance their potential

to kill diseased cells.

4. New drugs and new therapies for genetic diseases: Case of cell therapy

Cell therapy is a promising approach within the field of regenerative medicine that involves the transplantation of human cells to replace or repair damaged tissues and cells. This therapeutic strategy capitalizes on the body's inherent capacity for self-repair and healing, aiming to restore normal function to compromised tissues or organs.

Various types of cells can be utilized in cell therapy, including stem cells, lymphocytes, dendritic cells, and pancreatic islet cells. Stem cells are particularly valuable due to their unique ability to differentiate into multiple cell types and their potential to proliferate indefinitely. They can be sourced from different tissues, including embryonic stem cells, which are pluripotent and can develop into virtually any cell type, and adult stem cells, which are more limited in their differentiation potential but are crucial for tissue repair and regeneration (El-Kadiry et al. 2021). Figure 27 shows the principal steps for cell therapy proceeding.

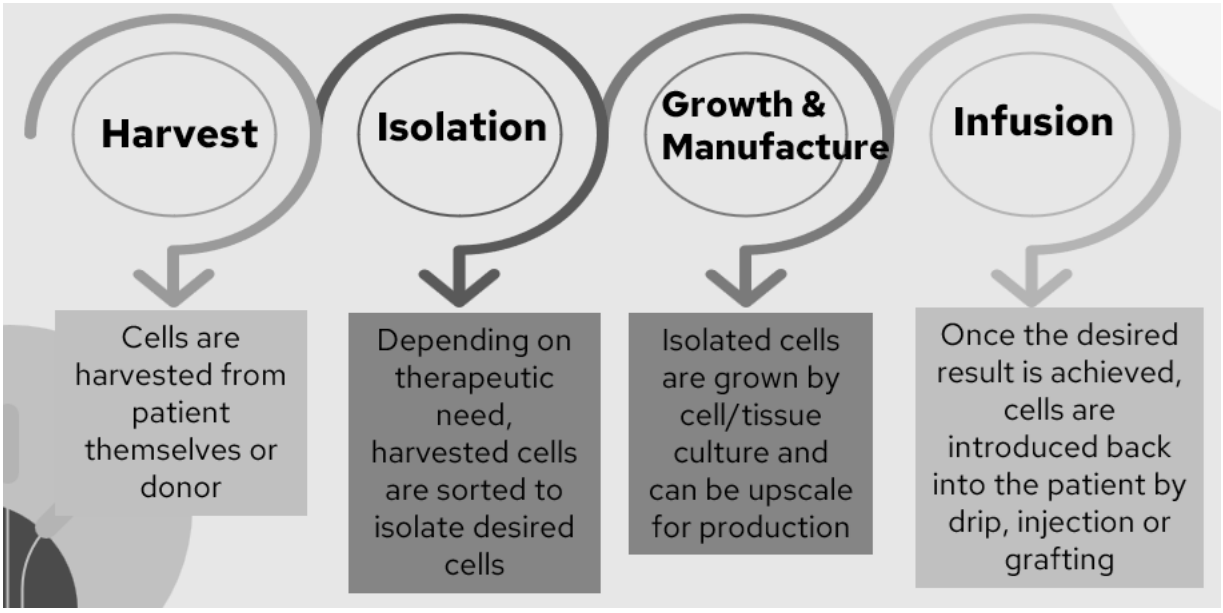

Figure 27. Illustrative image of cell therapy proceeding. Figure created by the author.

Case of CAR-T cell therapy

Autologous CAR T-cell therapy involves engineering a patient's own T cells to recognize and attack cancer cells. "Autologous" means the use of an individual's own cells or tissues in this therapy. In CAR T-cell therapy, white blood cells are taken from a patient's blood in a procedure called "apheresis" or "leukapheresis" and sent to a laboratory or manufacturing facility. There, the

T cells are separated and then modified so they have artificial receptors on their surface. The receptors direct the engineered T cell to find and attack the cancer cells. These artificial receptors are called “chimeric antigen receptors.” The number of engineered CAR T cells is multiplied in the laboratory or manufacturing facility. When there are enough of these cells, they are frozen and sent to the patient’s hospital or treatment center. When the patient is ready for treatment, the CAR T cells are thawed and given back to the patient through an intravenous (IV) infusion. The CAR T cells may not only eradicate all the cancer cells in the body, but they may remain in the body for months a er the infusion has been completed. This therapy has resulted in long-term remissions for some patients with certain types of cancer.

5. Combating infectious diseases

Despite advancements in public health, including the development and administration of vaccines and antimicrobial and antiviral agents, the efficacy of these interventions is o en compromised by the emergence of novel infectious agents and the rise of drug-resistant pathogens. Furthermore, the development of vaccines and therapeutic agents typically involves extensive clinical trials, which can prolong the timeline for effective intervention and hinder the rapid response necessary for managing infectious disease outbreaks. Rapid and accurate identification of pathogens is essential for enhancing treatment efficiency, minimizing disease transmission, and facilitating responses to critical public health events (Antabe and Ziegler 2020). In order to diagnose and combat infectious diseases, various techniques have been employed as follow:

5.1. Plasmid profiling

Plasmid profiling, introduced in the 1970s, was the first DNA-based bacterial subtyping method. This technique distinguishes bacterial strains by analyzing the number and size of plasmids they harbor. Plasmid DNA is separated from bacterial isolates using agarose gel electrophoresis (AGE), often visualized with fluorescent dyes. Plasmids play a key role in the dissemination of antibiotic resistance genes (ARGs), which is of growing concern due to the increasing prevalence of antibiotic-resistant bacteria (ARB). The emergence of multidrug-resistant (MDR) bacteria poses a significant threat to public health, limiting the effectiveness of available antibiotics (Grobelak et al. 2020). The global effort to combat antibiotic resistance now focuses on slowing the spread of ARGs and improving antibiotic stewardship (Bouki et al. 2013).

Plasmid profile analysis is a genotyping method that isolates and analyzes the plasmid DNA from bacterial cells. The technique generates unique plasmid patterns based on the size and number of plasmids present, which can be further refined by digesting the plasmid DNA with restriction enzymes for additional differentiation. This method is particularly useful for epidemiological studies, as it has been applied to typing methicillin-susceptible and methicillin-resistant Staphylococcus aureus (MSSA and MRSA). Despite its utility, plasmid profiling is limited

to strains containing plasmids, and the ability of bacteria to acquire or lose plasmids must be considered when interpreting results (Grobelak et al. 2020).

5.2. RFLP (Restriction Fragment Length Polymorphism)

Restriction Fragment Length Polymorphism (RFLP) is a molecular technique used to differentiate organisms by analyzing patterns generated from the cleavage of their DNA by specific restriction enzymes. If two organisms differ in the distance between cleavage sites for a particular restriction endonuclease, the resulting DNA fragments will vary in length when digested with the enzyme. The comparison of these fragment patterns enables the differentiation of species and even strains within species (Rohi et al. 2016). Polymorphisms are inherited differences that occur in more than 1% of the population, making them useful for genetic studies.

The method relies on restriction endonucleases, enzymes that cut DNA at specific nucleotide sequences, typically between 4 and 6 base pairs in length. Shorter recognition sequences produce more fragments, and differences in nucleotide sequences between organisms result in fragments of varying sizes. These fragments are separated using gel electrophoresis, providing a visual representation of the DNA pattern (Rohi et al. 2016).

5.3. AFLP (Amplified Fragment Length Polymorphism)

A molecular marker, such as an AFLP (Amplified Fragment Length Polymorphism) marker, is a DNA-based tool used in genetic analysis. AFLP markers are highly polymorphic, meaning they detect a wide range of genetic variation within populations, making them particularly useful in studies related to population genetics, evolution, and biodiversity. The AFLP technique, first introduced by (Paun et al. 2012; Vos et al. 1995), is a PCR-based fingerprinting method that generates large numbers of marker fragments without prior knowledge of the organism's genome, enabling its application across diverse species.

5.4. RAPD (Random Amplified Polymorphic DNA)

Random Amplified Polymorphic DNA (RAPD) is a versatile and cost-effective PCR-based method commonly used for genotyping bacterial species. RAPD markers are produced by amplifying random segments of genomic DNA using a single universal primer with an arbitrary nucleotide sequence. This technique generates DNA fragments that reveal polymorphisms between bacterial isolates, providing insights into their genetic diversity.

In RAPD analysis, two isolates are considered genetically indistinguishable if they produce the same number of DNA bands, and these bands exhibit the same apparent size on a gel electrophoresis profile. Variations in band patterns, such as a difference of three bands, can be used to differentiate closely related strains or clones. For instance, if the genetic similarity between Staphylococcus aureus isolates is 65% or greater, the isolates are likely to share a

common ancestor, indicating potential genetic relatedness (Eid et al. 2023).

RAPD relies on short primers, typically 8-12 nucleotides in length, to amplify random genomic regions. This technique is advantageous for its simplicity and affordability, as it does not require prior knowledge of the organism's genome. RAPD is widely used in epidemiological studies, bacterial typing, and population genetics due to its ability to distinguish between bacterial strains and assess genetic relatedness within species.

5.5. Pulsed field gel electrophoresis

In PFGE, the organisms are embedded in agarose, lysed in situ to extract the DNA and followed by digesting the DNA by using a restriction enzyme that cleaves infrequently. Slices of agarose are then inserted into wells of an agarose gel and electricity applied at three different angles to resolve the DNA into discrete bands. Multiple strategies have been developed to produce homogenous electrical field by different companies (Figure 28). For rapid separation of fragments 100bp to 250 kb in size, electrical field is fixed at 180o angle and is inverted in the forward and reverse directions. In another approach different voltages are applied to forward and reverse directions (De Moissac et al. 1994).

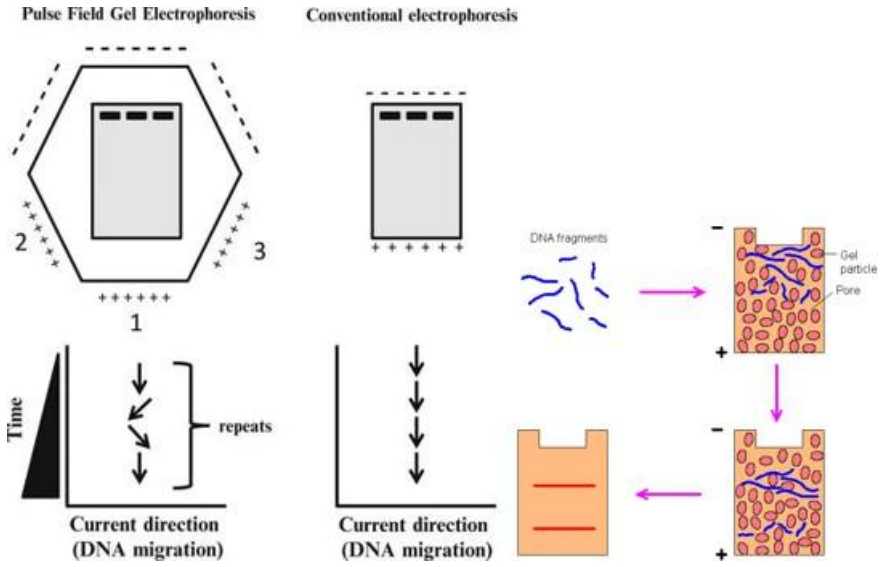

Figure 28. Illustrative image of Pulsed Field Gel Electrophoresis (Barry and Pollard, 1993).

5.6. Ribotyping

Ribotyping is a highly specific fragment-based molecular technique used for bacterial typing and strain differentiation. This method leverages the use of restriction enzymes to target and cleave conserved regions within the ribosomal RNA (rRNA) genes, specifically 5S, 16S, and 23S, as well as the intergenic spacer regions, including Glu-tRNA. By cutting these regions, ribotyping generates a unique DNA fingerprint that can be used to differentiate organisms down to the strain

level.

The restriction enzymes cleave the genomic DNA at specific recognition sites within the rRNA gene regions, producing DNA fragments of varying lengths that reflect the genetic composition of the organism. These fragments are then separated by gel electrophoresis, yielding a distinctive banding pattern or "fingerprint" that is characteristic of the organism. This pattern can be compared across different isolates to assess genetic relatedness and identify specific strains.

6. Genetic fingerprints and its application in forensic medicine

DNA fingerprinting, also known as genetic fingerprinting, is a powerful technique used for identifying individuals based on their unique DNA profiles. This method relies on detecting and analyzing variations in specific regions of the genome that are known to differ significantly between individuals. By examining these regions, scientists can determine whether two DNA samples come from the same person, related individuals, or unrelated individuals, making it a critical tool in fields such as forensic medicine, paternity testing, and genealogical research (Jeffreys 2005).

Several genetic markers are commonly used in DNA fingerprinting, each offering different advantages in terms of resolution and application. The most frequently utilized markers include Short Tandem Repeats (STRs), Restriction Fragment Length Polymorphisms (RFLPs), and Single Nucleotide Polymorphisms (SNPs). These markers differ in their underlying genetic properties and the techniques required for their analysis.

6.1. Short Tandem Repeats (STRs)

Short Tandem Repeats (STRs) are repeating sequences of 2-6 base pairs of DNA that are found at specific loci in the genome. These repeats vary in number between individuals, making them excellent candidates for genetic identification. STRs are highly polymorphic, meaning that the number of repeats at each locus can differ significantly between individuals, and even between close relatives. By analyzing multiple STR loci simultaneously, scientists can generate a unique DNA profile for each individual (Fan and Chu 2007).

The primary advantage of STRs is their high discriminatory power. Because these loci exhibit a high degree of variability, even individuals from the same family or ethnic group can be distinguished based on their STR profiles. STRs are commonly used in forensic science for criminal investigations, paternity testing, and identification of missing persons. The PCR (Polymerase Chain Reaction) amplification method is commonly employed to generate STR profiles, making this technique both sensitive and efficient.

6.2. Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) represent the most basic form of genetic variation.

SNPs are variations in a single base pair of DNA at a specific location in the genome. Unlike STRs and RFLPs, which involve longer regions of the genome, SNPs involve changes in a single nucleotide, such as an A-to-T substitution. These variations are abundant in the human genome and are used as genetic markers for identifying individuals, especially in association with certain traits or diseases (Li et al. 2023).

SNPs offer the advantage of being stable across generations, making them ideal for genealogical studies and ancestry tracing. However, their lower degree of variability compared to STRs limits their utility in some forensic applications, where greater discrimination is needed. Despite this, SNPs are increasingly being incorporated into DNA fingerprinting, particularly in applications where high-throughput sequencing or the analysis of large genomic datasets is required.

6.3. The use of mitochondrial genes

Mitochondrial DNA (mtDNA) is a unique type of DNA located within mitochondria, the membrane- bound organelles found in the cytoplasm of human cells. These organelles are primarily responsible for generating the cell's supply of adenosine triphosphate (ATP), the main energy source for cellular functions. Unlike nuclear DNA, which is inherited from both parents, mtDNA is maternally inherited, providing a distinct genetic lineage that is highly valuable in various scientific and forensic applications.

One of the key advantages of mtDNA in forensic science is its high copy number. While a cell only contains two copies of nuclear DNA, it can contain hundreds to thousands of copies of mtDNA. This abundance makes it more likely that usable mtDNA can be retrieved from forensic samples that are too degraded for nuclear DNA analysis, such as hair sha s, bones, and teeth. Moreover, mtDNA's maternal inheritance pattern allows for the reconstruction of maternal lineages over many generations, making it a valuable tool in cases of unidentified remains or missing persons (Syndercombe Court, 2021; Wiesner et al. 1992).

Advantages

- Older biological samples that lack nucleated cellular material, such as hair, bones, and teeth, cannot be analyzed with STR and RFLP, but they can be analyzed with mtDNA.
- Comparing the mtDNA profile of unidentified remains with the profile of a potential maternal relative can determine whether they share the same mtDNA profile and are related.
- Since mtDNA remains virtually the same from generation to generation, changing only about 2% to 4% every million years due to random mutation.
- Consequently relationships can be traced through the unbroken maternal line,

as shown in Figure bellow.

- It has been shown that forensic scientists can amplify the HV1 and HV2 regions of the mtDNA, and then sequence each region and compare single nucleotide differences to a reference sample.
- Also, mtDNA that is maternally inherited come from the mother and can be directly linked to maternal relatives and can be used as match references to establish family relationships through the mother's side.

6.4. Fingerprinting by Y chromosome

The male-specific region of the human Y chromosome is a crucial tool in forensic DNA analysis, particularly in scenarios where conventional autosomal DNA profiling fails to provide informative results. A Y-chromosomal gene fragment can be used to determine the biological sex of an individual linked to a crime scene. Y-chromosome short tandem repeat polymorphisms (Y-STRs) are utilized to construct haplotypes that identify paternal lineages of unknown male contributors. This is especially useful in cases where both males and females have contributed to the same sample, such as in sexual assault investigations (Kayser 2017).

Y-STR haplotyping has several key applications in forensic investigations: (i) it can exclude male suspects from involvement in a crime; (ii) it can identify the paternal lineage of male perpetrators; (iii) it can reveal the presence of multiple male contributors to a single sample, (iv) it can offer investigative leads for identifying unknown male perpetrators; (v) additionally, Y-STR analysis is applied in paternity disputes involving male offspring and other paternal kinship tests, including historical cases, missing person investigations, and disaster victim identification.

Y-chromosome polymorphisms are also useful in determining the paternal biogeographic ancestry of unknown individuals or missing persons when autosomal DNA profiling is inconclusive. This makes Y-chromosome analysis a valuable tool in both forensic and genealogical contexts, providing insights where other DNA methods fall short (Kayser 2017).

Chapter V: Somatic cloning

1. Introduction

Nuclear transfer, a method first developed in 1952 to study frog embryonic development, has played a crucial role in advancing cloning technology. By the 1980s, this technique was adapted for cloning cattle and sheep using cells from early-stage embryos. A significant breakthrough came in 1995 when Ian Wilmut, Keith Campbell, and their team successfully cloned live lambs, Megan and Morag, using cells cultured in vitro for an extended period. This was the first time live animals were cloned from cultured cells, marking a transformative moment in biotechnology. This achievement laid the foundation for the development of precise genetic modifications in livestock, with broad implications for agricultural and medical applications (Fléchon et al. 2004).

2. What is somatic cell cloning?

Somatic-cell nuclear transfer (SCNT) or Somatic Cell Cloning, is a laboratory technique for creating a clone embryo with a donor nucleus. It can be used in embryonic stem cell research, or, potentially, in regenerative medicine where it is sometimes referred to as “therapeutic cloning”. It can also be used as the first step in the process of reproductive cloning. Because somatic cell cloning is a cloning technique, it is best to outline principals underlying the process of cloning. Cloning is derived from the Greek word “ Klon” which means a twig which can replicate itself and grows eventually into a tree.

To clone is to reproduce asexually or to make a copy or a set of copies of an organism following the fusion or insertion of a diploid nucleus into an ovocyte. A true clone is an individual which has all the components that make up the individual, including nuclear genetic material (genome) and other maternally derived factors that are gained from a single unique embryo as a result of sexual reproduction.

In vivo, cloning in mammals involves replacing the genetic material of an egg with the genetic material of somatic cell from an embryo or adult which will eventually develop into a full organism or being; this is a biological clone; an organism genetically identical to another organism (Tian et al. 2003).

Steps for somatic cell cloning

Common somatic cloning protocols involve the following major technical steps (Figure 29 and 30).

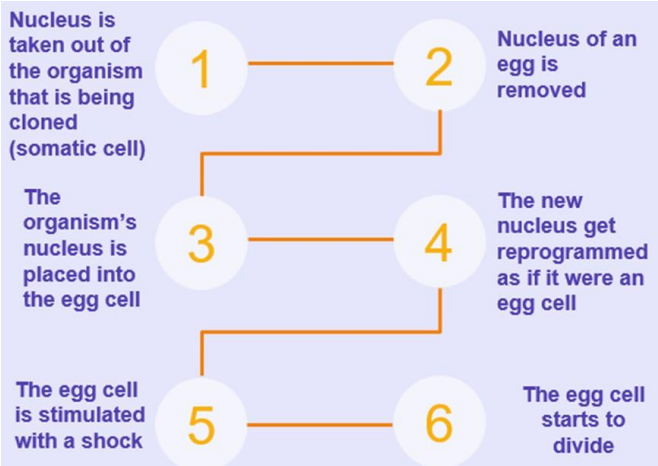

Figure 29. Steps for somatic cell cloning. Figure created by the author.

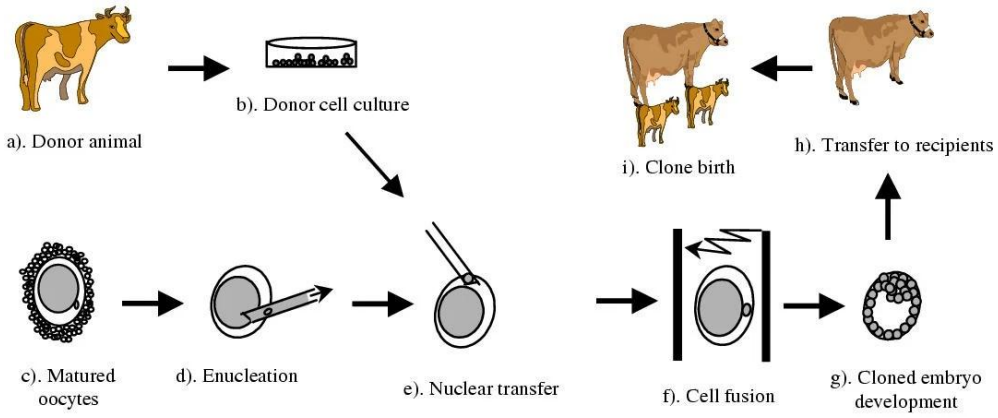

Figure 42. Schematic diagram of the somatic cloning process. Cells are collected from donor (a) and cultured in vitro (b). A matured oocyte (c) is then enucleated (d) and a donor cell is transferred into the enucleated oocyte (e). The somatic cell and the oocyte are then fused (f) and the embryos is allowed to develop to a blastocyst in vitro (g). The blastocyst can then be transferred to a recipient (h) and cloned animals are born a er completion of gestation (i) (Tian et al., 2003).

3. Factors affecting somatic cell cloning

Diverse methodologies have been implemented to enhance the efficacy of nuclear transfer, encompassing modifications to both donor cells and the nuclear transfer procedure. The predominant focus of these actions is directed towards optimizing donor cell characteristics. Noteworthy strategies in this regard involve: a) synchrony of the cell cycle stage of donor cells, as well as synchrony between donor cells and recipient oocytes; b) using somatic cells from donors of various ages, tissue origins, passages and culture conditions; c) transfer of stem cells with low levels of epigenetic marks; and d) modifying epigenetic marks of donor cells with drugs.

Despite marked advancements in enhancing the efficiency of nuclear transfer, with a notable departure from the initial success rate of one live clone per 277 embryo transfers, none of the aforementioned interventions have fully mitigated the inherent challenges associated with nuclear transfer. These outcomes underscore the imperative for further investigations into nuclear reprogramming mechanisms to comprehensively grasp the underlying processes and substantially refine the capacity of differentiated somatic nuclei to undergo successful reprogramming. The subsequent section elucidates various strategies employed to augment nuclear transfer efficiencies (Wells et al. 2003).

3.1. Effect of donor age

Several studies have found that the development rates of somatic cloned embryos remain similar regardless of the donor's age. However, it has observed that clones derived from adult cells often experience abortion during later stages of pregnancy. Moreover, calves that develop to term from adult cells show a higher number of abnormalities compared to those derived from newborn or fetal cells. In addition, results conducted a study in cattle where a large number of cloned embryos were transferred. They concluded that, in general, embryos cloned from fetal cells have higher pregnancy and calving rates than those from adult cells (Tian et al. 2003).

3.2. Effect of cell culture duration

Studies suggest that cells of higher passages were receptive to nuclear reprogramming. Additional support for this hypothesis comes from a study by (Enright et al. 2003) who showed that cells of later passages contain less epigenetic modifications, i.e., their histones are more acetylated than in earlier passages. This observation agrees with an earlier notion that in vitro culture of cells can induce expression of genes that were not expressed before culture. In summary, a greater proportion of late passage cells (passage 18), vs. earlier passage cells (passage 2), were found to be in G0/G1 whether or not they were in serum-starved culture conditions.

3.3. Effect of modification of pre-existing epigenetic marks in donor cells

Epigenetic modifications, such as histone acetylation and DNA methylation, are heritable alterations in chromatin that do not involve changes to the underlying gene sequences. These modifications play a key role in the differentiation of various cell types, despite them having identical genetic material. In natural reproduction, gametes typically exhibit low levels of DNA methylation, which further decreases during early embryonic development. However, in nuclear transfer, the somatic donor nucleus retains the specific epigenetic marks of its original tissue, and these need to be erased during nuclear reprogramming. The presence of these epigenetic modifications in donor cells can affect their ability to be reprogrammed, which in turn influences the success of cloned embryo development both in vitro and in vivo. As a result, pharmacological

treatments that reduce epigenetic marks in donor cells before nuclear transfer may enhance their reprogrammability (Tian et al. 2003).

4. Example: Dolly ‘’ same to her mother”

Dolly was the first mammal cloned from a cell from an adult animal. She was created using cells taken from the udder of a 6-year old Finn Dorset ewe. These cells were then cultured in a laboratory for several weeks. The next step involved fusing individual cells with unfertilised eggs, from which the genetic material had been removed. A total of 277 of these "reconstructed eggs" were then cultured for 6 days in temporary recipients. Out of these reconstructed eggs, 29 appeared to have developed normally to the blastocyst stage. These 29 eggs were then implanted into surrogate Scottish Blackface ewes. A er 148 days, one of these implanted eggs successfully gave rise to a live lamb - Dolly. The cloning is passed through these steps:

i. Scientists collected udder cells from Dolly's DNA mother. They allowed the cells to multiply until they reached a sufficient number.

ii. A different sheep provided an egg cell, from which the nucleus was removed.

iii. Using electricity, the scientists combined a udder cell with the egg cell that had no nucleus. As a result, the egg cell contained all the DNA from the udder cell.

The combined cell divided and developed into an embryo, which is the early stage of an animal before it is born or hatched. The embryo was then transferred into another sheep. A er five months, this sheep gave birth to Dolly.

5. Therapeutic cloning

Therapeutically cloning is almost identical to reproductive except for the cloned embryo is never implanted in uterus. Instead, the embryo is cloned for the sole purpose of extracting stem cells.

Stem cells have the incredible ability to turn to any other cell in the human body, which means it is crucial for developing new treatment for diseases and have the potential to repair or regenerate tissues or organs.

Induced pluripotent stem cell (iPSC)

PSCs are non-specialized cells able to differentiate into most cell types when they are placed in certain environment, which can form (bone, limbs, …). PSCs are harvested from embryos in the blastocyst stage (unethical).

iPSC allows to convert any mitotic cell to a stem cell. Mitotic cells are induced to revert back to a state where it can specialize in two different cell types. Take a mitotic cell from a patient body (skin cells) and treated the cells with a specific transcription factors Oct4 (Octamer-binding

transcription factor 4), Sox2 (Sex-determining region Y-box 2), Klf4 (Kruppel-like factor 4), and c-Myc (Myc proto-oncogene). These transcription factors enter in competition with the native ones in the cell and took control how the cell behave by changing the packing of DNA and the level of expression of different genes. This cause to mitotic cell to revert back to its pluripotent state.

Chapter VI: Modification of metabolism by genetic engineering

1. Introduction

Modification of metabolism using genetic engineering or Metabolic engineering is a fascinating field that combines biology, chemistry, and engineering principles to manipulate and optimize the metabolic pathways of living organisms. One of the key tools used in metabolic engineering is recombinant DNA technology. This technique allows scientists to insert or delete specific genes in an organism's genome, altering its metabolic capabilities. By introducing genes from different organisms or even synthesizing entirely new genes, researchers can create novel pathways that enable the production of valuable compounds. The success of metabolic engineering relies on a deep understanding of the biochemical pathways involved. Scientists must carefully analyze the metabolic network, identify rate-limiting steps, and select appropriate targets for modification. They also need to consider the interplay between different metabolic pathways and the impact that changes may have on overall cellular function. Computer modeling and simulation play a crucial role in metabolic engineering. By using mathematical models, scientists can predict the behavior of engineered pathways and optimize their performance. These models take into account factors such as enzyme kinetics, substrate availability, and cellular metabolism. Through iterative cycles of design, simulation, and experimental validation, researchers can fine-tune their engineered pathways to achieve desired outcomes (Benedito and Modolo 2014).

2. How to conduct modification of metabolism?

Metabolic engineering refers to the deliberate and targeted modification of an organism’s metabolic pathways to better understand and exploit cellular processes for chemical transformation, energy transduction, and supramolecular assembly (Lessard 1996). This interdisciplinary field integrates principles from chemical engineering, computational science, biochemistry, and molecular biology. At its core, metabolic engineering applies engineering principles to the design and analysis of metabolic pathways to achieve specific goals, such as enhancing productivity in the production of antibiotics, biosynthetic precursors, or polymers, or extending metabolic capabilities for chemical synthesis or degradation.

2.1. Setting up metabolic pathways

This is the first step in the process where the identification of desired goal to achieve through the improvement or modification of an organism’s metabolism. It requires a databases and scientific articles to determine the metabolic pathways of our product of interest.

2.2. Analyzing a metabolic pathway

The completed metabolic pathway is modeled by different approaches using AI or mathematic to find the theoretical yield of the product or reaction fluxes. Flux refers to the rate at which metabolites are produced, consumed, or transformed within a biological system.

2.3. Determining the optimal genetic manipulations

It is necessary to determine which reactions may be altered in order to maximize the yield of the desired product. To determine what specific genetic manipulations to perform, it is necessary to use computational algorithms, such as OptGene or OptFlux.

2.4. Experimental measurements

In order to verify the effect of genetic manipulations on the metabolic network (to ensure they align with the model), it is necessary to experimentally measure the fluxes in the network. To measure reaction fluxes, carbon flux measurements are made using carbon-13 isotopic labeling.

2.5. Resulting approaches

a. Over expressing the gene encoding the rate-limiting enzyme of the biosynthetic pathway of the desired end-product: This can be achieved by introducing additional copies of the gene into the organism or by using genetic engineering techniques to enhance the gene expression.
b. Inhibit the competing metabolic reactions which involve the same substrate leading to metabolically channel the substrate towards the desired chemical
c. The production of the desired biochemical can be carried out in the non-native organism. Gene can be isolated from the organism which naturally produces the desired biochemical and can be expressed in another organism which might be easier to cultivate than the host organism.
d. Engineered an enzyme that is not found in nature. This mode relies on creating mutations in the related gene so that the amino acid composition of the enzyme is altered.

3. Case of non-canonical amino acids

In the world of biology, amino acids are the essential building blocks of proteins. These tiny molecules play a crucial role in various biological processes, from supporting cellular functions to forming the structure of tissues and organs. While there are 20 standard amino acids commonly found in living organisms, scientists have also discovered a fascinating group of molecules known as non-canonical amino acids (ncAAs). In 1992, researchers were able to tweak cytosine to create what was essentially a 65th codon, which could then code for a non-canonical amino acid.

Non-canonical amino acids are chemically similar to their canonical counterparts but possess

unique properties and structures. These exceptional molecules offer researchers new avenues to explore in the field of metabolic engineering, where scientists manipulate cellular metabolic pathways to produce desired compounds or enhance organism capabilities.

In *E. coli* the non-canonical amino acids acids norvaline, norleucine, and β-methylnorleucine, which derive from an off-pathway of the branched-chain amino acid synthesis route are synthesized and incorporated into cellular and recombinant proteins

The study of non-canonical amino acids also sheds light on the natural processes of protein synthesis in living organisms. By investigating these unique molecules, scientists gain insights into the evolution and function of proteins. Understanding how ncAAs are incorporated into proteins provides valuable knowledge about the complexity of cellular machinery and the potential for expanding the diversity of life's building blocks.

References

Antabe R, Ziegler BR. Diseases, Emerging and Infectious. International Encyclopedia of Human Geography. 2020:389–91. doi: 10.1016/B978-0-08-102295-5.10439-1. Epub 2019 Dec 4. PMCID: PMC7153343.

Antunes M. S., Smith J. J., Jantz D., Medford J. I. (2012). Targeted DNA excision in Arabidopsis by a re-engineered homing endonuclease. BMC Biotechnol. 86, 86–12.

Bailey, J. E., Shurlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A.and Tsai, P.E. (1996). Inverse metabolic engineering a strategy for directed genetic engineering of useful phenotypes. Biotechnology & Bioengineering 52:109-121.

Barth, A. L. (2007). Visualizing circuits and systems using transgenic reporters of neural activity. Current opinion in neurobiology, 17(5), 567-571.

Benedito VA, Modolo LV. Introduction to metabolic genetic engineering for the production of valuable secondary metabolites in in vivo and in vitro plant systems. Recent Pat Biotechnol. 2014;8(1):61-75. doi: 10.2174/1872208307666131218125801. PMID: 24354528.

Berry, A. (1996). Improving production of aromatic compounds in Escherichia coli by metabolic engineering. Trends in Biotechnology 14:219-259.

Bertero, A., Brown, S., & Vallier, L. (2017). Methods of Cloning. Basic Science Methods for Clinical Researchers, 19–39. doi:10.1016/b978-0-12-803077-6.00002-3 10.1016/B978-0- 12-8030

Bouki, C., Venieri, D., Diamadopoulos, E., 2013. Detection and fate of antibiotic resistant bacteria in wastewater treatment plants: a review. Ecotoxicol. Environ. Saf. 91, 1e9.

Brookes G., Barfoot P. (2020). GM crop technology use 1996-2018: Farm income and production impacts. Gm. Crops Food 11 (4), 242–261.

Brown, G.C., Hafner, R.P. and Brand, M.D. (1990). A 'top-down' approach to the determination of control coefficients in metabolic control theory. European Journal of Biochemistry 188:321-325.

Carroll D. (2017). Genome editing: Past, present, and future. Yale J. Biol. Med. 90 (4), 653–659.

Carter, P. (1986). Site-directed mutagenesis. Biochemical Journal, 237(1), 1.

Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. Expert Systems with Applications, 106, 1–20. doi:10.1016/j.eswa.2018.03.050

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene

editing in mammalian cells. Nature biotechnology, 33(5), 543-548.

De Moissac, Y. R., Sheryl L. Ronald, and Mark S. Peppler. "Use of pulsed- field gel electrophoresis for epidemiological study of Bordetella pertussis in a whooping cough outbreak." Journal of clinical microbiology 32.2 (1994): 398-402.

Du H., Zeng X., Zhao M., Cui X., Wang Q., Yang H., et al. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. J. Biotechnol. 217, 90–97.

Dunbar, C. E., High, K. A., Joung, J. K., Kohn, D. B., Ozawa, K., & Sadelain, M. (2018). Gene therapy comes of age. Science, 359(6372), eaan4672. doi:10.1126/science.aan4672

Durland J, Ahmadian-Moghadam H. Genetics, Mutagenesis. [Updated 2022 Sep 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK560519/

Dwivedi S. L., Ceccarelli S., Blair M. W., Upadhyaya H. D., Are A. K., Ortiz R. (2016). Landrace germplasm for improving yield and abiotic stress adaptation. Trends plant Sci. 21 (1), 31–42.

Dymond, J. S. (2013). Explanatory Chapter. Laboratory Methods in Enzymology:

DNA, 279–289. doi:10.1016/b978-0-12-418687-3.00023-9

Edman-Wallér JE, Toepfer M, Karp J, Rizzardi K, Jacobsson G, Werner M. Clostridioides difficile outbreak detection: Evaluation by ribotyping and whole-genome sequencing of a surveillance algorithm based on ward-specific cutoffs. Infect Control Hosp Epidemiol. 2023 Dec;44(12):1948-1952.

Eid, Amira Ezzat, et al. "Random amplified polymorphic DNA technique (RAPD) for typing of Staphylococcus aureus causing Infection in intensive care units of Tanta University Hospitals." Microbes and Infectious Diseases 4.3 (2023): 819-832.

El-Kadiry AE-H, Rafei M andShammaa R (2021) Cell Therapy:Types, Regulation, and ClinicalBenefits. Front. Med. 8:756029.

Enright, B. P., Kubota, C., Yang, X., & Tian, X. C. (2003). Epigenetic Characteristics and Development of Embryos Cloned from Donor Cells Treated by Trichostatin A or 5-aza-2′-deoxycytidine1.Biology of Reproduction,69(3),896– 901. doi:10.1095/biolreprod.103.017954

Eren, K., Taktakoğlu, N., & Pirim, I. (2022). DNA sequencing methods: from past to present. The Eurasian journal of medicine, 54(Suppl 1), S47.

Fan, H., & Chu, J.-Y. (2007). A Brief Review of Short Tandem Repeat Mutation. Genomics, Proteomics & Bioinformatics, 5(1), 7–14. doi:10.1016/s1672-0229(07)60009-6

Fléchon, J. E., Kopecny, V., Pivko, J., Pavlok, A., & Motlik, J. (2004). Texture of the zona pellucida of the mature pig oocyte. The mammalian egg envelope revisited. Reproduction Nutrition Development, 44(3), 207–218. doi:10.1051/rnd:2004026

Follstad, B. D. and Stephanopoulos, G. (1998). Effect of reversible reactions on isotope label redistribution analysis of the pentose phosphate pathway. European Journal of Biochemistry 252: 360-371.

Fortin NY, Mulchandani A, Chen W. Use of real time polymerase chain reaction and molecular beacons for the detection of Escherichia coli O157:H7. Analytical Biochemistry. 2001;289:281-288

Ghalayini M., Magnan M., Dion S., Zatout O., Bourguignon L., Tenaillon O., et al. (2019). Long-term evolution of the natural isolate of Escherichia coli 536 in the mouse gut colonized a er maternal transmission reveals convergence in the constitutive expression of the lactose operon. Mol. Ecol. 28 4470–4485.

Gilroy, C. A., Roberts, S., & Chilkoti, A. (2018). Fusion of fibroblast growth factor 21 to a thermally responsive biopolymer forms an injectable depot with sustained anti-diabetic action. Journal of Controlled Release, 277, 154–164. doi:10.1016/j.jconrel.2018.03.015

Goh, H. C., Sobota, R. M., Ghadessy, F. J., & Nirantar, S. (2017). Going native: Complete removal of protein purification affinity tags by simple modification of existing tags and proteases. Protein Expression and Purification, 129, 18–24.

Gowing, G., Svendsen, S., & Svendsen, C. N. (2017). Ex vivo gene therapy for the treatment of neurological disorders. Progress in Brain Research, 99– 132. doi:10.1016/bs.pbr.2016.11.003

Grierson, A. W., Darnbrough, C., & MacDonald, C. (1992). A Comparison of Eukaryotic Control Elements using Tissue Plasminogen Activator (tPA) as a Reporter Gene. In Animal Cell Technology (pp. 75-77). Butterworth-Heinemann.

Grobelak, A., Sobczyk, M., & Całus, K. (2020). Drug and multidrug resistance in waterborne pathogens. Waterborne Pathogens, 279–300. doi:10.1016/b978-0-12-818783-8.00014- 1

Guo, T., Feng, Y. L., Xiao, J. J., Liu, Q., Sun, X. N., Xiang, J. F., ... & Xie, A. Y. (2018). Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9- mediated genome editing. Genome biology, 19, 1-20.

Gupta, V., Sengupta, M., Prakash, J., & Tripathy, B. C. (2016). An Introduction to Biotechnology. Basic and Applied Aspects of Biotechnology, 1–21. doi:10.1007/978-981-10-0875- 7_1

Hang C., Wohlhueter R., Zhang H. (2016a). Genetically modified foods: A critical review of their promise and problems. Food Sci. Hum. Wellness. 5, 116–121.

Heinrich, R. and Rapoport, S.M. (1997). Metabolic regulation and mathematical models. In Progress in Biophysics and Molecular Biology, Vol. 32, Butler, J. A. V., Noble, D., Ed., Pergamon Press: Oxford, UK, pp 1-82.

Hsu P. D., Lander E. S., Zhang F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell 157 (6), 1262–1278.

Hug, K. Genetically modified organisms: Do the benefits outweigh the risks? Medicina 2008, 44, 87. https://doi.org/10.3390/medicina44020012

abalameli HR, Zahednasab H, Karimi-Moghaddam A, Jabalameli MR. Zinc finger nuclease technology: advances and obstacles in modelling and treating genetic disorders. Gene. 2015 Mar 01;558(1):1-5.

effreys, A. J. (2005). Genetic fingerprinting. Nature Medicine, 11(10), 1035–

1039. doi:10.1038/nm1005-1035

Jeyaseelan K, Ma D, Armugam A. Real-time detection of gene promotor activity: Quantification of toxin gene transcription. Nucleic Acids Research. 15 June 2001;29(12):e58

Johnston, M. O. (2006). Mutations and New Variation: Overview. Encyclopedia of Life Sciences. doi:10.1038/npg.els.0004165

Joyce C. Quantitative RT-PCR. A review of current methodologies. Methods in Molecular Biology. 2002;193:83-92

Kacser, H. and Burns, J.A. (1973). The control of flux. Symposium of the Society of Experimental Biology 27:65104.

Kadri, K. (2020). Polymerase Chain Reaction (PCR): Principle and Applications. Synthetic Biology - New Interdisciplinary Science. doi:10.5772/intechopen.86491

Karbalaei, M., Rezaee, S. A., & Farsiani, H. (2020). Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. Journal of cellular physiology, 235(9), 5867-5881.

Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. Human Genetics, 136(5), 621–635. doi:10.1007/s00439-017-1776-9

Kimple, M. E., Brill, A. L., & Pasker, R. L. (2013). Overview of Affinity Tags for Protein Purification. Current Protocols in Protein Science, 9.9.1–9.9.23.

Ledley, F.D. Pharmaceutical Approach to Somatic Gene Therapy. Pharm Res 13, 1595–1614 (1996). https://doi.org/10.1023/A:1016420102549

Lessard, P. (1996). Metabolic engineering, the concept coalesces. Nature Biotechnology 14: 1654-1655.

Liao, J.C., Hou, S.Y. and Chao, Y.P. (1996). Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. Biotechnology & Bioengineering 52:129-140.

Li J., Li X., Zhou Z., Wu P., Fang M., Pan X., et al. (2016b). Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. Front. Plant Sci. 7, 377.

Li, Jianjun, et al. "Construction of SNP fingerprint and population genetic analysis of honeysuckle

germplasm resources in China." Frontiers in Plant Science 14 (2023): 1080691.

Lin MH, Chen TC, Kuo TT, Tseng C, Tseng CP. Real-time PCR for quantitative detection of Toxoplasma gondii. Journal of Clinical Microbiology. 2000;38:4121-4125

Ling MM, Robinson BH. Approaches to DNA mutagenesis: an overview. Anal Biochem. 1997 Dec 15;254(2):157-78.

Madhavan, A., Sindhu, R., Binod, P., Sukumaran, R. K., & Pandey, A. (2017). Strategies for design of improved biocatalysts for industrial applications. Bioresource Technology, 245, 1304–1313. doi:10.1016/j.biortech.2017.05

Marcela AAV, Rafael LG, Lucas ACB, Paulo RE, Alessandra ATC, Sergio C. Principles and applications of polymerase chain reaction in medical diagnostic fields: A review.Brazilian Journal of Microbiology. 2009;40:1-11

Martínez-Fortún J., Phillips D. W., Jones H. D. (2017). Potential impact of genome editing in world agriculture. Emerg. Top. Life Sci. 1 (2), 117–133.

McLean, T., Fourie, P. H., & McLeod, A. (2009). Reporter gene transformation of the trunk disease pathogen Phaeomoniella chlamydospora and biological control agent Trichoderma harzianum. Australasian Plant Pathology, 38(2), 153-167.

Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA Sequencing. NextGeneration Genome Sequencing: Towards Personalized Medicine, 1–11.https://doi.org/10.1002/9783527625130.ch1

Mergeay, M., & Van Houdt, R. (2018). Plasmids as secondary chromosomes. Molecular Life Sciences an Encyclopedic Reference; Wells, RD, Bond, JS, Klinman, J., Masters, BSS, Eds, 961-964.

Mittal, B., Chaturvedi, P., & Tulsyan, S. (2013). Restriction Fragment Length Polymorphism. Brenner's Encyclopedia of Genetics, 190–193. doi:10.1016/b978-0-12-374984- 0.01314-0

Mizuno, Y., Okamoto, S., Hara, M., & Mizoguchi, T. (2014). EMS mutagenesis and characterization of Brassica rapa mutants. Plant Biotechnology, 31(2), 185-190.

Morrow, J. F. (1979). [1] Recombinant DNA techniques. Recombinant DNA, 3– 24. doi:10.1016/0076-6879(79)68003-5

Munung, N.S., Nnodu, O.E., Moru, P.O. et al. Looking ahead: ethical and social challenges of somatic gene therapy for sickle cell disease in Africa. Gene Ther 31, 202–208 (2024). https://doi.org/10.1038/s41434-023-00429-7

Nishikawa, M., & Hashida, M. (2002). Nonviral Approaches Satisfying Various Requirements for Effective in Vivo Gene Therapy. Biological & Pharmaceutical Bulletin, 25(3), 275–283. doi:10.1248/bpb.25.275

Pangasa, Aradhana, et al. "High resolution melting-curve (HRM) analysis for the diagnosis of cryptosporidiosis in humans." Molecular and cellular probes 23.1 (2009): 10-15.

Paun O, Schönswetter P. Amplified fragment length polymorphism: an invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. Methods Mol Biol. 2012;862:75-87.

Philippe Gallusci, philippe.gallusci· Zhanwu Dai, zhanwu., Michel Génard·[2017] Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. 22, 7, p610-623.

Quan, P. L., Sauzade, M., & Brouzes, E. (2018). dPCR: a technology review. Sensors, 18(4), 1271.

Rajeevan MS, Vernon SD, Taysavang N, Unger ER. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. The Journal of Molecular Diagnostics. 2001;3(1):26-31

Redman, M., King, A., Watson, C., & King, D. (2016). What is CRISPR/Cas9? Archives of Disease in Childhood - Education & Practice Edition, 101(4), 213–215. doi:10.1136/archdischild-2016-310459

Ricroch A., Clairand P., Harwood W. (2017). Use of CRISPR systems in plant genome editing: Toward new opportunities in agriculture. Emerg. Top. Life Sci. 1 (2), 169–182.

Rohit A, Maiti B, Shenoy S, Karunasagar I. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) for rapid diagnosis of neonatal sepsis. Indian J Med Res. 2016 Jan;143(1):72-8. doi: 10.4103/0971-5916.178613.

Samiec M. Molecular Mechanism and Application of Somatic Cell Cloning in Mammals-Past, Present and Future. Int J Mol Sci. 2022 Nov 9;23(22):13786. doi: 10.3390/ijms232213786. PMID: 36430264; PMCID: PMC9697074.

Sauer N. J., Narváez-Vásquez J., Mozoruk J., Miller R. B., Warburg Z. J., Woodward M. J., et al. (2016). Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. Plant Physiol. 170 (4), 1917–1928.

Savageau, M.A., Voit, E.O. and Irvine, D.H. (1987). Biochemical systems theory and metabolic control theory: I. Fundamental similarities and differences. Mathematical Biosciences 86:127-145.

Shahzad, S., Afzal, M., Sikandar, S., & Afzal, I. (2020). Polymerase Chain Reaction. Genetic Engineering - A Glimpse of Techniques and Applications. doi:10.5772/intechopen.81924

Sharma, K. K., Bhatnagar-Mathur, P., & Thorpe, T. A. (2005). Genetic transformation technology: status and problems. In Vitro Cellular & Developmental Biology-Plant

Shehata HR, Hassane B and Newmaster SG (2024) Real-time PCR methods for identification and stability monitoring of Bifidobacterium longum subsp. longum UABl-14 during shelf life. Front. Microbiol. 15:1360241. doi: 10.3389/fmicb.2024.1360241

Shi J., Gao H., Wang H., Lafitte H. R., Archibald R. L., Yang M., et al. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol. J. 15 (2), 207–216.

Shimada, H., Kondo, K., Fraser, P. D., Miura, Y., Saito, T and, Misawa, N. (1998). Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. Applied Microbiology and Biotechnology 64:2676-2680.

Smirnoff, N. (1998). Plant resistance to environmental stress. Current Opinion in Biotechnology 9:214-219.

Stephanopoulos, G. and Vallino, J.J. (1991). Network rigidity and metabolic engineering in metabolite overproduction. Science 252:1675-1681.

Stephen B, Mueller R. Realtime reverse transcription PCR (qRTPCR) and its potential use in clinical diagnosis. Clinical Science. 2005;109:365-379

Stratagene. (2007). Introduction to quantitative PCR: methods and applications guide.

Syndercombe Court, D. (2021). Mitochondrial DNA in forensic use. Emerging Topics in Life Sciences, 5(3), 415-426.

Terpe, K. (2005). Protein Tags. In: Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Springer, Berlin, Heidelberg . https://doi.org/10.1007/3-540- 29623-9_3650

Tian, X.C., Kubota, C., Enright, B. et al. Cloning animals by somatic cell nuclear transfer – biological factors. Reprod Biol Endocrinol 1, 98 (2003). https://doi.org/10.1186/1477-7827- 1-98

Upta A., Christensen R. G., Rayla A. L., Lakshmanan A., Stormo G. D., Wolfe S. A. (2012). An optimized two-finger archive for ZFN-mediated gene targeting. Nat. Methods. 9 (6), 588– 590.

Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research. 1995;23:4407–4414. doi: 10.1093/nar/23.21.4407.

WANG, D.-K., SUN, Z.-X., & TAO, Y.-Z. (2006). Application of TILLING in Plant Improvement. Acta Genetica Sinica, 33(11), 957–964. doi:10.1016/s0379-4172(06)60130-3

Wang D, Gao G. State-of-the-art human gene therapy: part II. Gene therapy strategies and clinical applications. Discov Med. 2014 Sep;18(98):151-61. PMID: 25227756; PMCID: PMC4440458.

Wells DN, Laible G, Tucker FC, Miller AL, Oliver JE, Xiang T, Forsyth JT, Berg MC, Cockrem K, L'Huillier PJ, Tervit HR, Oback B: Coordination between donor cell type and cell cycle stage improves nuclear cloning efficiency in cattle. Theriogenology. 2003, 59: 45-59. 10.1016/S0093-691X(02)01273-6.

Wiesner, R.J., Ruegg, J.C. and Morano, I. (1992) Counting target molecules by exponential polymerase chain reaction: copy number mitochondrial DNA in rat tissues. Biochem. Biophys.

Res. Commun. 183, 553–559

Wilson, A. J., Morgan, E. R., Booth, M., Norman, R., Perkins, S. E., Hauffe, H. C., … Fenton, A. (2017). What is a vector? Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1719), 20160085.

Wivel, N. A., & Wilson, J. M. (1998). METHODS OF GENE DELIVERY.

Hematology/Oncology Clinics of North America, 12(3), 483–501. doi:10.1016/s0889-8588(05)70004-6

Wolf, D.P., Mitalipov, P.A. & Mitalipov, S.M. Principles of and strategies for germline gene therapy. Nat Med 25, 890–897 (2019). https://doi.org/10.1038/s41591-019-0473-8

Zhao, X., Li, G., & Liang, S. (2013). Several Affinity Tags Commonly Used in Chromatographic Purification. Journal of Analytical Methods in Chemistry, 2013, 1–8.

Zong Y., Song Q., Li C., Jin S., Zhang D., Wang Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. 36, 950–953.