SECTION 6: TECHNIQUES IN THE CHARACTERIZATION AND MANIPULATION OF NUCLEIC ACIDS

LEARNING OBJECTIVES

By the end of this section, you will be able to do the following:

- Explain the techniques used in recombinant DNA technology
- Discuss current approaches to DNA manipulation and analysis

The study of nucleic acids began with the discovery of DNA structure and the elucidation of the cellular process that it undergoes, and progressed to the study of genes and entire genomes of organisms. Advances in our ability to analyze and manipulate DNA have led to breakthroughs in the diagnosis and treatment of diseases, development of novel proteins and enhanced agricultural products, and a better understanding of heredity, among others.

BASIC TECHNIQUES USED TO MANIPULATE DNA

Basic techniques used in the manipulation and characterization of genetic material include molecular cloning, hybridization, gel electrophoresis, PCR, blotting methods, and DNA sequencing, and genome editing.

Molecular cloning

Molecular cloning refers to the process of combining DNA sequences from multiple sources. The resulting DNA product, called a recombinant DNA molecule, is not usually found in biological organisms. Molecular cloning applications include producing a specific gene or DNA fragment of interest for further study, and the expression of a protein product that is otherwise difficult to extract from its natural source.

Molecular cloning requires two elements:

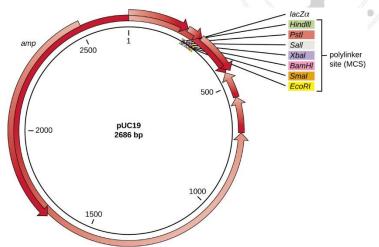
- 1. DNA fragment to be propagated may encode a gene of interest
- Vector DNA molecule used as a "vehicle" in which to insert the desired DNA fragment. A vector needs to be able to self – replicate. DNA cloning vectors include:
 - Plasmids the most commonly used cloning vector
 - self-replicating, extra-chromosomal circular DNA molecules
 - used as cloning vector for small pieces of DNA (typically 50 5000 base pairs)
 - replicates independently of the chromosomal DNA in bacteria
 - carry genes that can contribute favorable traits to the organism, such as antibiotic resistance

Phage vector – phages are viruses that can infect bacteria
 about 1000 times more efficient than the plasmid vector

Cosmids – can be packaged in lambda phage particles for infection into *E. coli* Can accommodate up to 45 kb of foreign DNA

 Yeast artificial chromosomes (YACs) – yeast vectors containing with centromeres and telomeres

- can carry about 200 kb or larger fragments
- Bacteria artificial chromosomes (BACs) one of the most frequently used vectors for large inserts in genome projects



□ can accommodate 100 – 300 kb DNA inserts

Figure 1. The artificially constructed plasmid vector pUC19 is commonly used for cloning foreign DNA. Arrows indicate the directions in which the genes are transcribed. Note the polylinker site, containing multiple unique restriction enzyme recognition sites, found within the *lacZ* reporter gene. Also note the ampicillin (*amp*) resistance gene encoded on the plasmid. From

P. MAN

https://openstax.org/books/microbiology/pages/12-1-microbes-and-the-tools-of-genetic-engineering

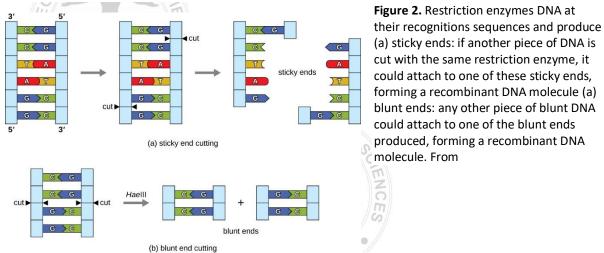
ARTSAN

Recombinant plasmids are commonly used molecular cloning vectors because they are relatively easy to use and carry useful selectable markers, for example, antibiotic resistance genes. Bacteria transformed with recombinant DNA molecule is grown in an antibiotic for which the plasmid carries a resistance gene. Only bacteria that has taken up the recombinant plasmid will grow in this medium.

The most important tools in making recombinant DNA molecules are restriction endonucleases, which are enzymes naturally produced by bacteria as a defense mechanism against foreign DNA. Restriction enzymes recognize and cleave specific palindromic DNA sequences, known as recognition sequences. The reaction usually creates single-stranded overhangs (or "sticky ends"). Two DNA molecules cut with the same restriction enzyme will generate the same sticky ends. When mixed together, they can anneal to form a recombinant DNA molecule. The recombinant DNA molecule is then reintroduced into a host cell, usually bacteria. When the bacteria divides, copies of the recombinant DNA will also be made.

Cloning vectors such as plasmids have been engineered such that they contain multiple restriction sites, which facilitates molecular cloning. The recombinant DNA produced by the

bacteria may be extracted for further characterization. Special plasmid vectors, called expression vectors, may also be used to produce the protein product of the cloned gene.



https://openstax.org/books/microbiology/pages/12-1-microbes-and-the-tools-of-genetic-engineering

Genetic Engineering

Genetic engineering is the alteration of an organism's genotype using molecular cloning techniques to modify an organism's DNA, usually to achieve certain traits. The organism that receives the recombinant DNA is called a genetically modified organism (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is now considered to be transgenic. Bacteria, plants, and animals have been genetically modified since the early 1970s for research, medical and agricultural purposes. In the US, GMOs such as Roundup-ready soybeans and borer-resistant corn are commonly available in the market.

An example of a transgenic microorganism is the bacterial strain that produces human insulin. Insulin gene isolated from humans was inserted into a plasmid, and the recombinant plasmid was then inserted into bacteria. As a result, these transgenic microbes are able to produce and secrete human insulin.

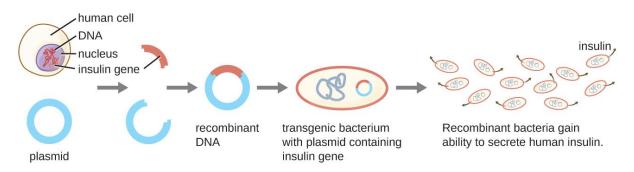


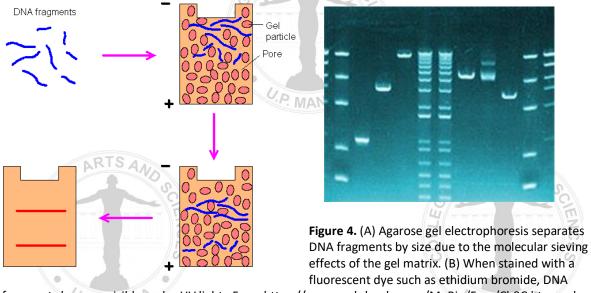
Figure 3. Recombinant DNA technology is the artificial recombination of DNA from two organisms. In this example, the human insulin gene is inserted into a bacterial plasmid. This recombinant plasmid can then be used to transform bacteria, which gain the ability to produce the insulin protein. From https://openstax.org/books/microbiology/pages/12-1-microbes-and-the-tools-of-genetic-engineering

Recombinant DNA Product	Application
Atrial natriuretic peptide	Treatment of heart disease (e.g., congestive heart failure), kidney disease, high blood pressure
DNase	Treatment of viscous lung secretions in cystic fibrosis
Erythropoietin	Treatment of severe anemia with kidney damage
Factor VIII	Treatment of hemophilia
Hepatitis B vaccine	Prevention of hepatitis B infection
Human growth hormone	Treatment of growth hormone deficiency, Turner's syndrome, burns
Human insulin	Treatment of diabetes
Interferons	Treatment of multiple sclerosis, various cancers (e.g., melanoma), viral infections (e.g., Hepatitis B and C)
Tissue plasminogen activator	Treatment of pulmonary embolism in ischemic stroke, myocardial infarction
T T T T T T T T T T T T T T T T T T T	

Table 1. Some genetically engineered pharmaceutical products and applications

DNA electrophoresis

Electrophoresis is a laboratory technique used to identify, quantify, and purify nucleic acid fragments. It involves the application of an electrical field to separate charged molecules through a gel matrix. Because it contains phosphate groups, DNA has an intrinsic negative charge and will move towards the positive electrode. Agarose, the matrix commonly used to separate DNA, has molecular sieving properties, so that shorter DNA fragments will travel faster towards the positive electrode, while the longest fragments will remain closest to the origin of the gel. This results in the separation of fragments based on size.



fragments become visible under UV light. From https://www.web-books.com/MoBio/Free/Ch9C.htm and

https://www.thermofisher.com/ph/en/home/life-science/cloning/cloning-learning-center/invitrogen-schoolof-molecular-biology/na-electrophoresis-education/na-separation-overview.html

DNA libraries

DNA libraries are collections of cloned DNA fragments from a particular organism. Like conventional libraries, DNA libraries are used to collect and store information in the form of DNA. They are critical for the analysis of gene function and for detection of related genes from different sources.

There are two types of DNA libraries:

1. genomic library - contains DNA fragments representing the entire genome of an organism. Genomic DNA from the organism is isolated and digested with a restriction enzyme, and restriction fragments are cloned into a suitable vector. A large number of transformed bacterial colonies must be isolated and kept to ensure that all possible genes from the genome of interest are present in at least one vector. Genomic libraries are commonly used for sequencing applications.

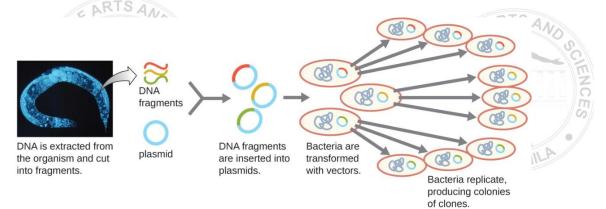


Figure 5. The generation of a genomic library facilitates the discovery of the genomic DNA fragment that contains a gene of interest. (credit "micrograph": modification of work by National Institutes of Health) From https://openstax.org/books/microbiology/pages/12-1-microbes-and-the-tools-of-genetic-engineering

2. cDNA library – contains only complementary DNA molecules synthesized from mRNA molecules in a cell. Total mRNA is isolated mRNA from the cell type of interest using the poly-A tail that is present in all eukaryotic mRNAs, reverse transcribed to synthesize complementary DNA strands. The single stranded DNA molecules are then converted to double stranded form, and then cloned into suitable vectors.





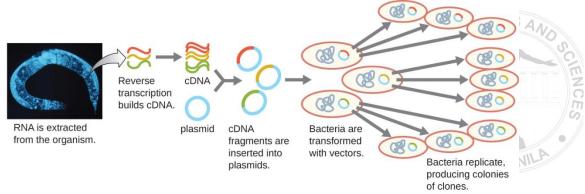


Figure 6. Complementary DNA (cDNA) is made from mRNA by the retroviral enzyme reverse transcriptase, converted into double-stranded copies, and inserted into either plasmid vectors or bacteriophage, producing a cDNA library. (credit "micrograph": modification of work by National Institutes of Health). From https://openstax.org/books/microbiology/pages/12-1-microbes-and-the-tools-of-genetic-engineering

Screening a library by hybridization

To screen a genomic library for a particular gene or sequence of interest, researchers utilize a DNA probe, which is a single-stranded DNA fragment that is complementary to part of the gene of interest. Probes are generally derived from cloned DNA from a related organism,

MANIL

and are labeled with a molecular tag or fluorescent dye.

The target DNA (*i.e.*, the DNA from the library to be probed) is denatured and bound to a nitrocellulose or nylon membrane. The membrane is then incubated with the labeled probe. After washing away excess probe, the membrane is screened by the chosen detection system (e.g., autoradiography).

Many organisms have had their entire genomes sequenced, so the use of library screening has declined in recent years.

Figure 7. Screening a DNA Library by Probing

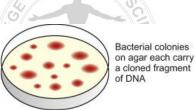
Blotting techniques

Probes can also be used to detect specific molecules in DNA fragments separated by gel electrophoresis. However, probes cannot be applied directly to the gel. The problem

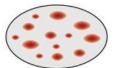
can be solved by using a different approach – blotting. There are three types of blotting methods:

MAN

1. Southern blot – used to find a particular sequence in a DNA sample mixture. DNA fragments are separated on a gel, transferred to a nylon membrane, and incubated with a DNA probe complementary to the sequence of interest

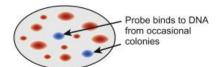








ADD LABELED DNA PROBE



- Northern blot similar to Southern blotting, but detects RNA gel instead of DNA. Northern blots are typically used to detect the amount of mRNA expressed in a tissue or organism sample.
- 3. Western blot protein mixtures are run on an SDS-PAGE gel and detected using protein specific antibodies

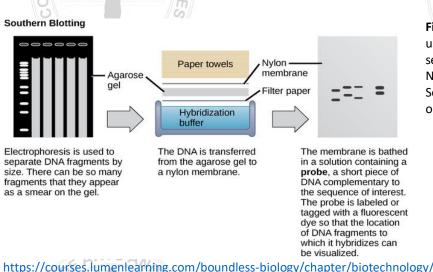


Figure 8. Southern blotting is used to find a particular sequence in a sample of DNA. Northern blotting is similar to Southern blotting, but RNA is run on the gel instead of DNA. From

Microarray Analysis

Microarray analysis is another technique that utilizes hybridization between complementary DNA sequences. It is a major tool in the field of **transcriptomics**, which focuses the complete set of RNA transcripts that are produced by the genome. Microarrays are typically used to compare gene-expression patterns from different samples, for example, cells infected with a virus versus uninfected cells, or cancerous cells versus healthy cells. It allows genome-wide expression profiling, instead of gene-specific analysis used in older techniques such as the Northern blot.

Typically, DNA fragments encompassing an organism's entire genomic library, or cDNA fragments corresponding to an organism's full complement of expressed genes, are incorporated in a DNA chip, or gene chip, with each DNA fragment acting as a probe. Labelled single-stranded DNA (or cDNA from reverse – transcribed RNA) from cell samples are hybridized to the gene chip under high stringency conditions. The amount of







hybridization detected for a specific probe is proportional to the number of nucleic acid fragments in the sample.

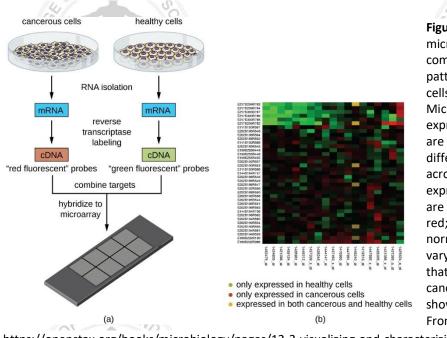


Figure 9. A) The steps in microarray analysis for comparing gene expression patterns between cancerous cells and healthy cells. B) Microarray information can be expressed as a heat map. Genes are shown on the left side; different samples are shown across the bottom. Genes expressed only in cancer cells are shown in varying shades of red; genes expressed only in normal cells are shown in varying shades of green. Genes that are expressed in both cancerous and normal cells are shown in yellow. From

https://openstax.org/books/microbiology/pages/12-2-visualizing-and-characterizing-dna-rna-and-protein

Although microarray technology allows global gene expression comparison between two samples in a short time, it requires sophisticated detection equipment and analysis software; thus, its use is typically limited to research settings. Observed gene expression differences also simply suggest what genes are involved in a process. Their actual roles will require traditional experimental verification.

Amplification of nucleic acid fragments: the polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a technique used to amplify specific regions of DNA for further analysis. Developed in 1983 by Kary Mullis, it can generate millions of copies of a target DNA sequence. Amplification is achieved by a series of three steps:

- 1. Denaturation: double-stranded DNA templates are heated to separate the strands
- 2. Annealing: short DNA molecules called primers bind to flanking regions of the target DNA
- 3. Extension: DNA polymerase extends the 3' end of each primer along the template strands

These steps are repeated ("cycled") 25–35 times to exponentially produce exact copies of the target DNA. PCR is made possible by thermostable DNA polymerases, first found in hot springs, which can withstand temperatures up to 96°C.





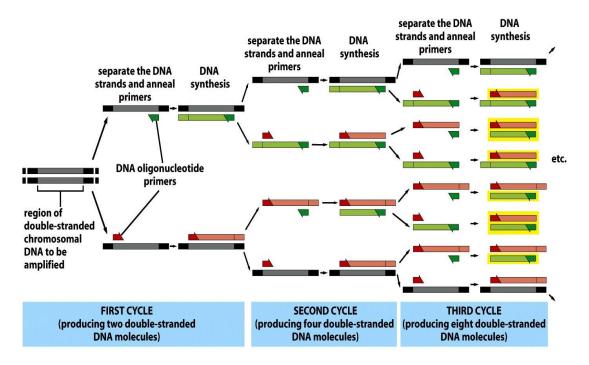


Figure 10. Three steps of PCR: denaturation, annealing, and extension, as shown in the first cycle, and the exponential amplification of target DNA with repeated cycling. From Figure 8-45b *Molecular Biology of the Cell* (© Garland Science 2008)

PCR is one of the most commonly used assays for obtaining a particular segment of DNA or RNA, and has multiple applications in research, forensic, and clinical laboratories, including:

- determining the sequence of nucleotides in a specific region of DNA
- amplifying a target region of DNA for cloning into a plasmid vector
- identifying the source of a DNA sample left at a crime scene
- analyzing samples to determine paternity
- comparing samples of ancient DNA with modern organisms
- determining the presence of difficult to culture, or unculturable, microorganisms in humans or environmental samples

PCR variations

Several modifications to PCR have been developed to further increase its utility.

MAN

1. Reverse transcriptase PCR (RT-PCR)

This PCR variation is used to amplify RNA targets. The RNA template must first be reverse transcribed to convert mRNA molecules into cDNA. That cDNA is then used as a template for traditional PCR amplification. RT-PCR can detect whether a specific gene has been expressed in a sample.

2. Real-time PCR, or quantitative PCR (qPCR)

Standard PCR and RT-PCR protocols are not quantitative because amplification products are only detected at the end of reactions, when they are run in agarose gel

10

electrophoresis. It is not possible to accurately determine how many original template molecules were present in the sample at the start of the reaction.

In qPCR, fluorescent tags are used to monitor the increase in target sequence concentration in real time during PCR amplification and to determine C_T , the cycle at which the fluorescence signal, which corresponds to target DNA concentration, exceeds a defined threshold. C_T inversely proportional to no. of copies of original cDNA template. The data obtained can then be used to quantify the amount of the original target sequence.

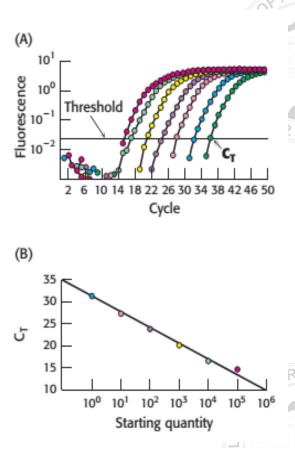


Figure 11. Real time PCR quantitatively measures nucleic acid concentrations. A) C_T , or the number of cycles required for the amplification to shift from the initial linear to the exponential phase, is a measure of the abundance of the original template. B) The starting concentration of an unknown sample may be determined from a standard plot using samples with known concentrations.

The use of qPCR has allowed researchers to determine the number of DNA copies, and even organisms, present in a sample. In clinical settings, qRT-PCR is used to determine viral load in HIV-positive patients to evaluate the effectiveness of their therapy. More recently, qPCR has become widely accepted as the gold standard diagnostic test for the detection of nucleic acid from SARS-CoV-2, the virus that causes COVID-19. Interestingly, since SARS-CoV-2 is an RNA virus, the COVID-19 RT-PCR test is a real-time reverse transcription polymerase chain reaction.

Genomics and sequencing

Genomics focuses on the comparison of entire genomes, including their complete set of genes, nucleotide sequence and organization. It utilizes recombinant DNA techniques, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of genomes. Advances in the field of genomics coincide with advances in DNA sequencing technologies.

The basic sequencing technique used for most modern day sequencing projects is the dideoxy chain termination method developed by Fred Sanger in the 1970s. In Sanger sequencing, the DNA strand to be sequenced serves as a template for DNA synthesis. A DNA primer is designed to be the starting point for DNA synthesis. DNA synthesis reactions are performed. Each reaction contains normal deoxynucleotides (dATP, dGTP, dCTP and DTTP), as well as small amounts of one of four dideoxynucleotides (ddNTPs): ddATP, ddGTP, ddCTP,

or ddTTP. Since a ddNTP lacks a 3' hydroxyl group, when it is incorporated into a chain of nucleotides instead of the regular nucleotide, synthesis terminates. Since the ddNTPs are randomly incorporated, synthesis terminates at different positions for each reaction.

Following synthesis, the sequencing products are separated using electrophoresis. Since the different ddNTPs are each labeled by a different fluorescent color, determining the order by which they reach a fluorescence detector will determine the order of added nucleotides in the sequence.

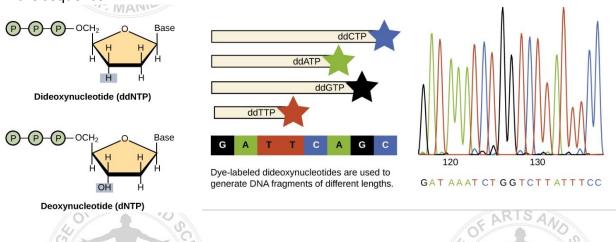


Figure 12. Sanger's method. A) A dideoxynucleotide (ddNTP) is similar in structure to a deoxynucleotide (dNTP), except for a missing 3'OH group. B) When a fluorescent-tagged ddNTP is incorporated into a DNA strand, synthesis stips. In a reaction containing the different ddNTPs, DNA fragement synthesis can be terminated at different points. The synthesis products are then separated on the basis of size, and fluorescent tags are used to determine the order of added nucleotides in the sequence.

Genomes are large – typically in the millions of baseparis – and therefore cannot be sequenced end-to-end in a single step. To sequence a genome, its DNA must first be broken down into smaller pieces, with each resulting fragment sequenced individually. The established base order for the specific DNA fragment is called a 'sequence read'. Computational tools are then utilized to assemble the various fragments and deduce the sequence of the starting genome. This process was historically called 'shotgun sequencing'.

The first major foray into whole genome DNA sequencing was the Human Genome Project (HGP), which was spearheaded by the US National Institutes of Health. Completed in 2003, the HGP utilized Sanger sequencing of genomic clones from DNA obtained from multiple individuals to produce a representative sequence of the human genome.

Since 2005, major genome sequencing projects have utilized **next generation sequencing** (NGS), a catch-all term that describes a number of different modern sequencing technologies capable of generating sequences millions of short fragments (25 to 600 base pairs) in a single day. These newer technologies use sophisticated software to get through the tedious process of determining the order of all fragments generated, and can produce entire genome sequences faster and at a much cheaper cost. Next generation sequencing technologies include:

• Illumina (Solexa) sequencing: works by simultaneously identifying DNA bases, as each base emits a unique fluorescent signal, and adding them to a nucleic acid chain.

- Roche 454 sequencing: based on pyrosequencing, which detects pyrophosphate release using fluorescence after nucleotides are incorporated by polymerase to a new strand of DNA.
- Ion Torrent: Proton/PGM sequencing measures the direct release of H⁺ (protons) from the incorporation of individual bases by DNA polymerase and therefore differs from the previous two methods as it does not measure light.
- Nanopore-based DNA sequencing: involves threading single DNA strands through extremely tiny pores in a membrane. DNA bases are read one at a time as they squeeze through the nanopore. The bases are identified by measuring differences in their effect on ions and electrical current flowing through the pore.

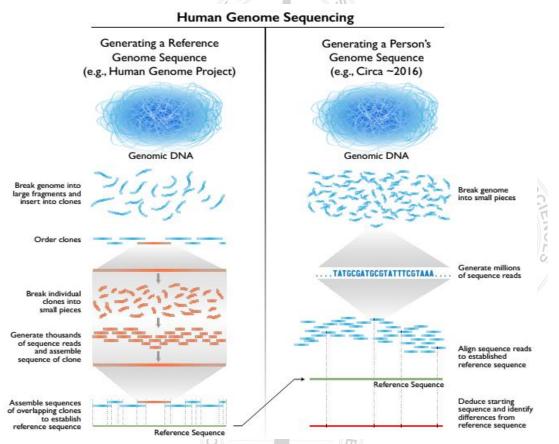
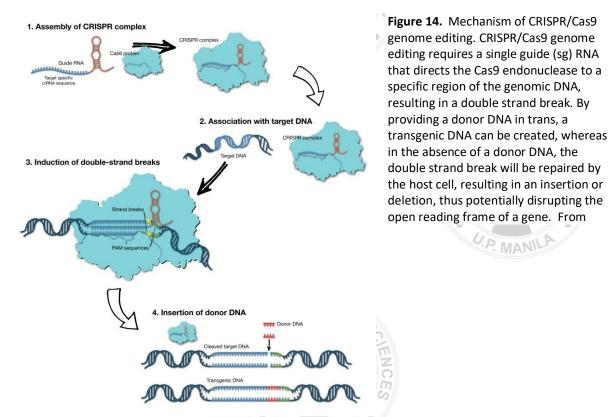


Figure 13. A comparison of approaches to sequencing whole genomes. The Human Genome Project utilized Sanger dideoxy sequencing, while newer projects utilize next generation sequencing methods that can generate entire genomes much more cheaply and rapidly. From https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

The National Center for Biotechnology Information houses a widely used genetic sequence database called GenBank where researchers deposit genetic information for public use. Upon publication of sequence data, researchers upload it to GenBank, giving other researchers access to the information. The collaboration allows researchers to compare newly discovered or unknown sample sequence information with the vast array of sequence data that already exists.

Genome editing

Genome editing (also called gene editing) is a group of technologies that can be used to modify an organism's DNA, for example, the addition, removal or alteration of DNA segments at particular locations. Several approaches to genome editing have been developed, the most well – known of which is CRISPR-Cas9 (which stands for clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9). CRISPR-Cas9 was adapted from a naturally occurring genome editing system in bacteria. The bacteria capture snippets of DNA from invading viruses and use them to create DNA segments known as CRISPR arrays. The CRISPR arrays allow the bacteria to "remember" the viruses (or closely related ones). If the viruses attack again, the bacteria produce RNA segments from the CRISPR arrays to target viral DNA. The bacteria then use Cas9 or a similar enzyme to cut the DNA apart, disabling the virus.



https://www.ncbi.nlm.nih.gov/books/NBK464635/figure/gen_edit.F2/

The CRISPR-Cas9 system works similarly in the laboratory. Researchers create a small piece of RNA with a short "guide" sequence that binds to a specific target sequence of DNA in the genome. The RNA also binds to the Cas9 enzyme. As in bacteria, the modified RNA is used to recognize the DNA sequence, and the Cas9 enzyme cuts the DNA at the targeted location. Although Cas9 is the enzyme that is used most often, other enzymes (for example, Cpf1) can also be used. Once the DNA is cut, researchers use the cell's own DNA repair machinery to add or delete pieces of genetic material, or to make changes to the DNA by replacing an existing segment with a customized DNA sequence.

The CRISPR-Cas9 system has generated a lot of excitement in the scientific community because it is faster, cheaper, more accurate, and more efficient than other existing genome

editing methods. Currently, most research on genome editing is done to understand diseases using cells and animal models. Scientists are also exploring the use of CRISPR-Cas9 to treat a wide variety of diseases, including single-gene disorders such as cystic fibrosis, hemophilia, and sickle cell disease. However, its safety and efficacy in humans has yet to be established.

There are also ethical concerns in the potential use of genome editing to alter human genomes. Most of the changes introduced with genome editing are limited to somatic cells, which are cells other than egg and sperm cells. These changes affect only certain tissues and are not passed from one generation to the next. However, changes made to genes in egg or sperm cells, or in the genes of an embryo, could be passed to future generations. Germline cell and embryo genome editing bring up a number of ethical challenges, including whether it would be permissible to use this technology to enhance normal human traits (such as height or intelligence). Based on concerns about ethics and safety, germline cell and embryo genome editing are currently illegal in many countries.

REFERENCES

Stryer 8th ed.

https://bio.libretexts.org/Bookshelves/Genetics/Book%3A Working with Molecular Genet ics (Hardison)

DNA sequence assembly: https://media.hhmi.org/biointeractive/click/Sequence Assembly/05.html

https://www.ebi.ac.uk/training/online/course/functional-genomics-ii-commontechnologies-and-data-analysis-methods/design-considerations

https://geneticeducation.co.in/real-time-pcr-principle-procedure-advantages-limitationsand-applications/

https://courses.lumenlearning.com/boundless-biology/chapter/biotechnology/

http://www.bio.miami.edu/dana/dox/restrictionenzymes.html

David P. Clark, Nanette J. Pazdernik, in Molecular Biology (Second Edition), 2013

https://openstax.org/books/microbiology/pages/12-2-visualizing-and-characterizing-dnarna-and-protein

P. MANIL

https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

https://courses.lumenlearning.com/boundless-biology/chapter/whole-genome-sequencing/

https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-

https://www.nature.com/scitable/definition/dna-sequencing-205/

https://ghr.nlm.nih.gov/primer/genomicresearch/genomeediting

OTHER VIDEO RESOURCES

- Restriction Endonucleases HD Animation <u>https://www.youtube.com/watch?v=crb</u> -<u>e2eDqg&list=PLYCGVJq0DVwKrmoSIvhzAOh0SpNUgfqFf&index=30</u>
- Cloning a Gene HD Animation
 <u>https://www.youtube.com/watch?v=ryIAvGYPPzU&list=PLYCGVJq0DVwKrmoSlvhzAOh0SpNUgfqFf&i
 ndex=79</u>
- cDNA Complementary DNA HD Animation
 <u>https://www.youtube.com/watch?v=QR14GZK8QkQ&list=PLYCGVJq0DVwKrmoSlvhzAOh0SpNUgfqFf</u> <u>&index=82</u>
- Polymerase Chain Reaction HD Animation
 <u>https://www.youtube.com/watch?v=PB46LKwNU90&list=PLYCGVJq0DVwKrmoSlvhzAOh0SpNUgfqFf</u>
 <u>&index=40</u>
- RT PCR overview
 https://www.youtube.com/watch?v=1kvy17ugl4w
- RT-qPCR for diagnosing COVID-19 (former 2019-nCoV) • https://www.youtube.com/watch?v=5f wieig4iQ
- Using a DNA Microarray HD Animation <u>https://www.youtube.com/watch?v=aV8K8fy-wj4</u>
- CRISPR Explained <u>https://www.youtube.com/watch?v=UKbrwPL3wXE</u>









