

Lecture 2 - Fundamental Technologies II: Amplification, Sequencing, and Genome Engineering

BENG168

**Instructor: Adam M. Feist, Assistant Professor, Shu Chien -
Gene Lay Department of Bioengineering**

In-Class Announcements & Follow Up

- Update to previous lecture slides
- In-class announcements

Completing Double-Stranded cDNA

- **Second Strand Synthesis:** The hybrid is treated with RNase H, which nicks the mRNA strand.
- DNA Polymerase I uses these nicks as initiation sites to synthesize the second strand.
 - DNA Pol I has 5' exonuclease activity, which removes the remaining RNA fragments as synthesis proceeds.
 - But, it can continue synthesizing DNA even after it reaches a downstream nick or fragment.
- **Last Step (in figure):** Ends of the DNA fragments are polished (blunted) using T4 DNA Polymerase for blunt-end ligation. It can extend (Filling In) as well as shorten (Chewing Back).
- **Final Step:** T4 DNA Ligase can join the blunt ends

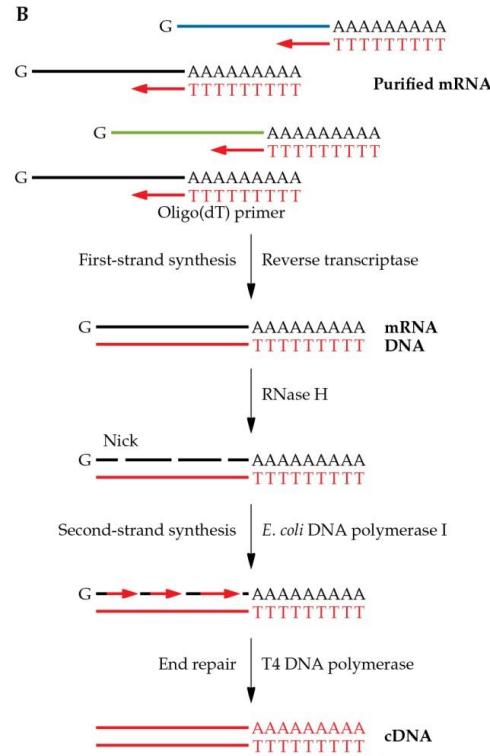


Figure 2.11 Synthesis of double-stranded cDNA using gene-specific primers (**A**) or oligo(dT) primers (**B**). A short oligonucleotide primer is added to a mixture of purified mRNA and anneals to a complementary sequence on the mRNA. Reverse transcriptase catalyzes the synthesis of a DNA strand from the primer using the mRNA as a template. To synthesize the second strand of DNA, the mRNA is nicked by RNase H, which creates initiation sites for *E. coli* DNA polymerase I. The 5' exonuclease activity of DNA polymerase I removes RNA sequences that are encountered as DNA synthesis proceeds. The ends of the cDNA are blunted using T4 DNA polymerase prior to cloning.

Module 1: DNA Amplification and Synthesis

(Source Pages: Chapter 2: 39–47)

- The Polymerase Chain Reaction (PCR)
- Cloning Strategies for Amplicons (T/A Cloning)
- Quantitative Real-Time PCR (qPCR)
- Synthetic Biology: Building Genes from Scratch
- Gibson Assembly and Large-Scale Synthesis

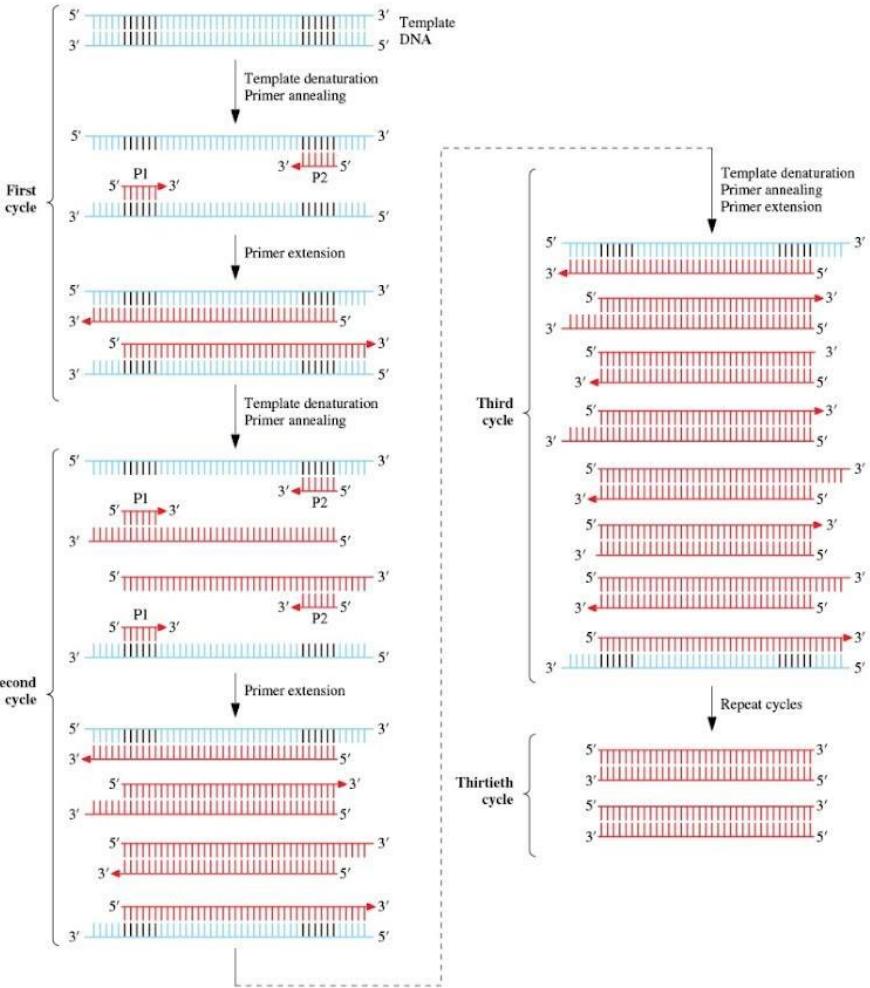
The Power of PCR (Polymerase Chain Reaction) Amplification

- High-speed *in vitro* synthesis of specific DNA targets.
- **Sensitivity:** Capable of producing millions of copies from a single template molecule.
- **Specificity:** Two synthetic oligonucleotide primers (~20 nt) define the boundaries of the reaction.
- **Components:** Requires template DNA, thermostable DNA polymerase (e.g., Taq), dNTPs, and buffer with Mg²⁺



The PCR Cycle

- **Denaturation (95°C):** High heat breaks hydrogen bonds to separate double-stranded DNA into single strands.
- **Annealing (45–68°C):** Primers base-pair to complementary flanking sequences on the target template.
- **Extension (70°C):** Taq polymerase synthesizes a new strand from the 3'-OH of the primers.
- Repeated 25–40 times to achieve exponential amplification of the "short template", products called **amplicons**.



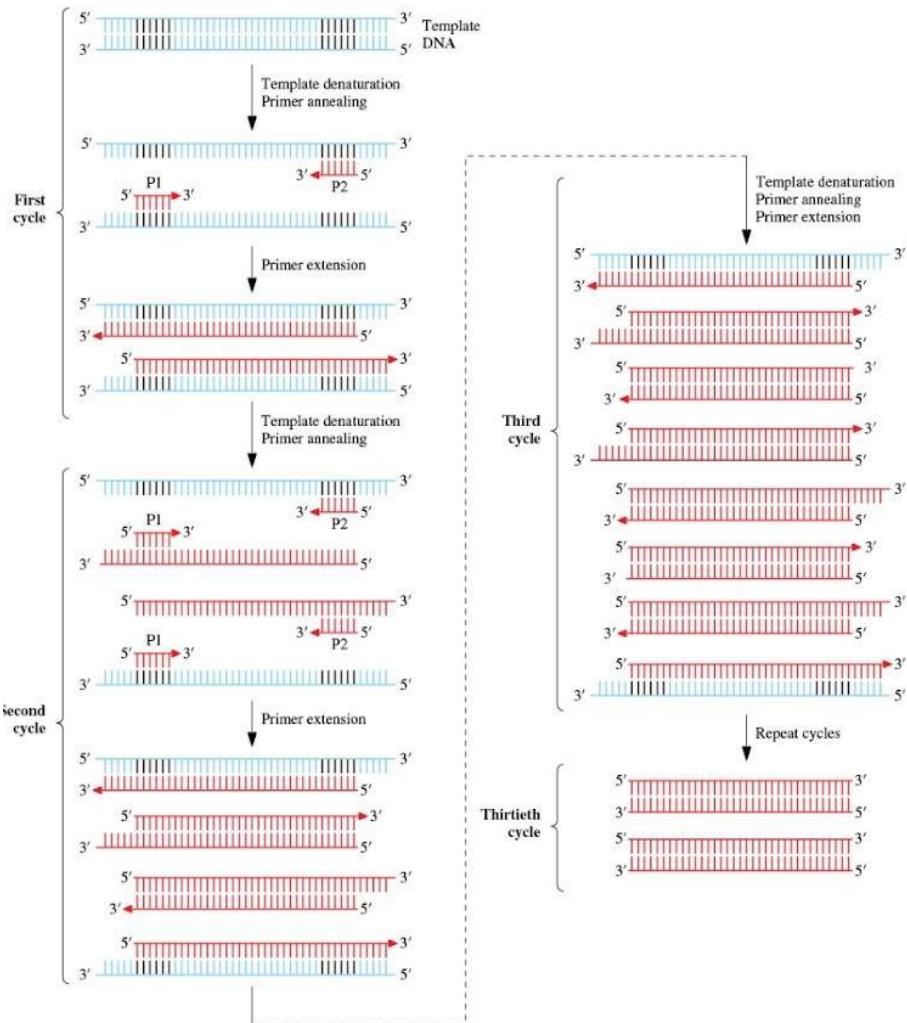


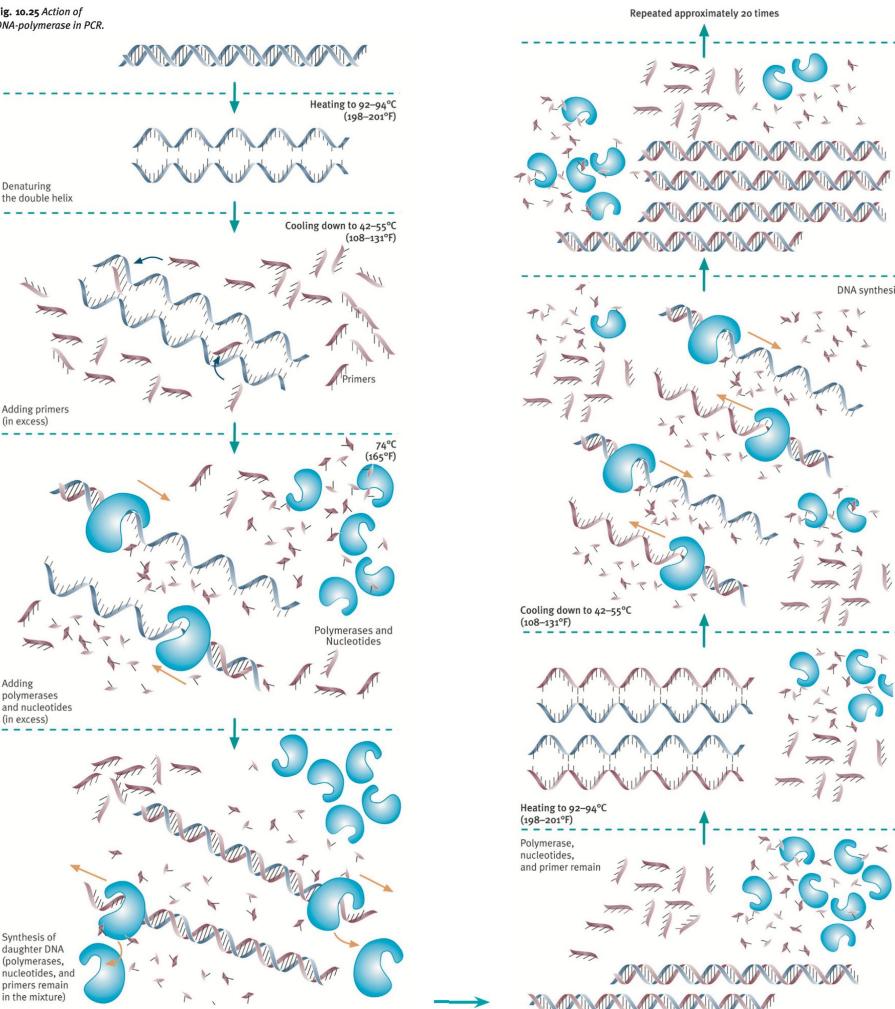
Figure 2.20 PCR. During a PCR cycle, the template DNA is denatured by heating and then slowly cooled to enable two primers (P1 and P2) to anneal to complementary (black) bases flanking the target DNA. The temperature is raised to about 70°C, and in the presence of the four deoxyribonucleotides, *Taq* DNA polymerase catalyzes the synthesis of a DNA strand extending from the 3' hydroxyl end of each primer. In the first PCR cycle, DNA synthesis continues past the region of the template DNA strand that is complementary to the other primer sequence. The products of this reaction are two long strands of DNA that serve as templates for DNA synthesis during the second PCR cycle. In the second cycle, the primers hybridize to complementary regions in both the original strands and the long template strands, and DNA

synthesis produces more long DNA strands from the original strands and short strands from the long template strands. A short template strand has a primer sequence at one end and the sequence complementary to the other primer at its other end. During the third PCR cycle, the primers hybridize to complementary regions of original, long template, and short template strands, and DNA synthesis produces long strands from the original strands and short strands from both long and short templates. By the end of the 30th PCR cycle, the products (amplicons) consist predominantly of short double-stranded DNA molecules that carry the target DNA sequence delineated by the primer sequences. Note that in the figure, newly synthesized strands are differentiated from template strands by a terminal arrow.

Key Components of PCR

- Template DNA (100–3,000 bp).
- Two oligonucleotide primers.
- Thermostable DNA polymerase (Taq).
- dNTPs and Buffer.

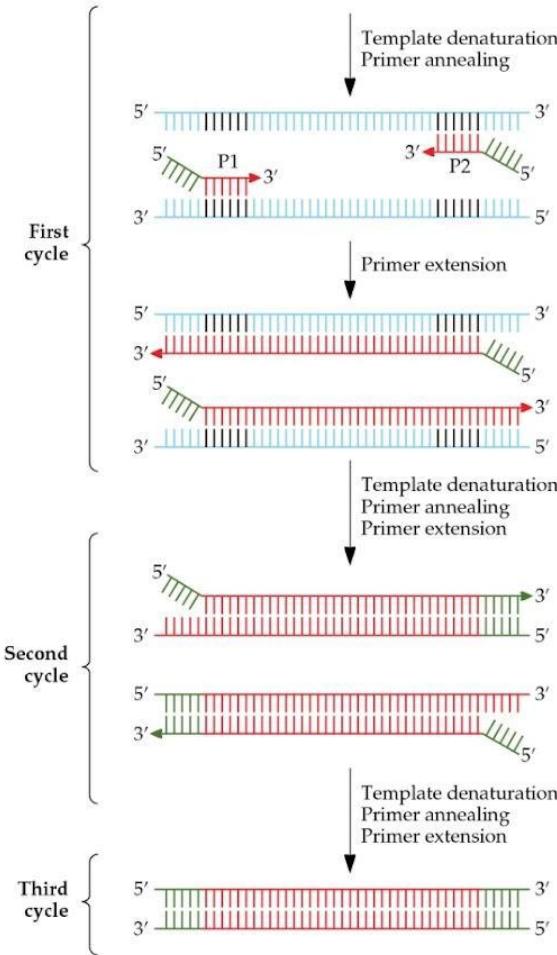
Fig. 10.25 Action of DNA-polymerase in PCR.



Cloning PCR Products

- Adding restriction sites to 5' primer ends for easier ligation (from Lecture 1). Dont bind initial target DNA.

Figure 2.21 Addition of restriction enzyme recognition sites to PCR-amplified target DNA to facilitate cloning. Each of the two oligonucleotide primers (P1 and P2) has a sequence of approximately 20 nucleotides in the 3' end that is complementary to a region flanking the target DNA (shown in black). The sequence at the 5' end of each primer consists of a restriction endonuclease recognition site (shown in green) that does not base-pair with the template DNA during the annealing steps of the first and second PCR cycles. However, during the second cycle, the long DNA strands produced in the first cycle serve as templates for synthesis of short DNA strands (indicated by a terminal arrow) that include the restriction endonuclease recognition sequences at both ends. DNA synthesis during the third and subsequent PCR cycles produces double-stranded DNA molecules that carry the target DNA sequence flanked by restriction endonuclease recognition sequences. These linear PCR products can be cleaved with the restriction endonucleases to produce sticky ends for ligation with a vector. Note that not all of the DNA produced during each PCR cycle is shown.



Cloning PCR Products: T/A Cloning

- PCR amplicons often lack convenient restriction sites for standard ligation (from Lecture 1).
- **Taq Polymerase Oddity:** Taq lacks 3'-5' proofreading and adds a single extra deoxyadenosine (dAMP) to the 3' ends.
 - Only added to the 3' ends of the strands after the polymerase has finished replicating the template
- **T-Vector Strategy:** Linearized vectors with complementary 3' deoxythymidine (dTTP) overhangs are used to capture these "A-tailed" products.
- **Efficiency:** This "universal" method bypasses the need for restriction enzymes during the initial cloning step.

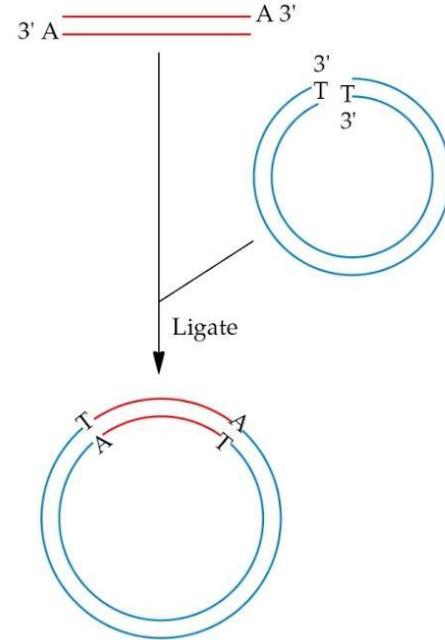


Figure 2.22 Cloning of PCR products without using restriction endonucleases. *Taq* DNA polymerase adds a single dAMP (A) to the ends of PCR-amplified DNA molecules. These extensions can base-pair with complementary single dTMP (T) overhangs on a specially constructed linearized cloning vector. Ligation with T4 DNA ligase results in insertion of the PCR product into the vector.

Quantitative PCR (qPCR)

- Allows us to **quantify** the starting amount of DNA in a sample, which is vital for diagnostics.
- Real-time monitoring of DNA synthesis using fluorescent dyes (e.g., SYBR Green).
- **ΔRn (Normalized Fluorescence):** The magnitude of the signal generated by the given set of PCR conditions.
- **C_t (Threshold Cycle):** The cycle number at which fluorescence exceeds the background threshold.

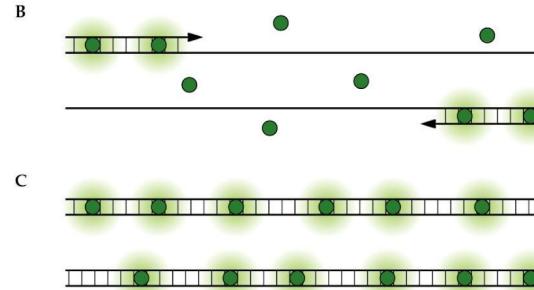


Figure 2.23 The fluorescent dye SYBR green does not bind to single-stranded DNA (A), binds to double-stranded DNA as it is synthesized (B), and is bound to the double-stranded amplified DNA (C). Only the dye-bound DNA fluoresces.

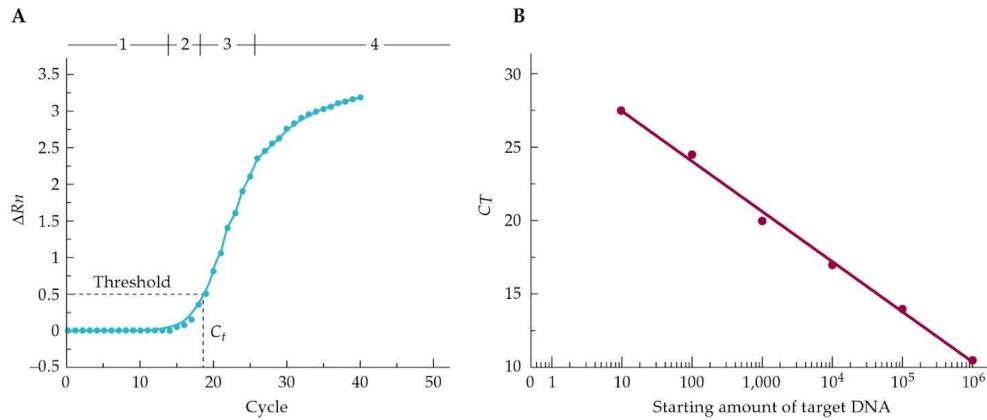


Figure 2.24 (A) Plot of normalized fluorescence (ΔRn) versus cycle number in a real-time PCR experiment. Four phases of PCR are shown: (1) a linear phase, where fluorescence emission is not yet above background level, (2) an early exponential phase, where the fluorescence intensity becomes significantly higher than the background (the cycle at which this occurs is generally known as C_t), (3) an exponential phase, where the amount of product doubles in each cycle, and (4) a plateau phase, where reaction components are limited and amplification slows down. (B) Plot of C_t versus the starting amount of a target nucleotide sequence. Fluorescence detection is linear over several orders of magnitude.

Chemical Synthesis of Genes

- **Oligo Assembly:** Overlapping ~60-mer oligonucleotides are chemically synthesized and annealed.
- **Gap Filling:** DNA Polymerase I uses 3'-OH groups as initiation points to fill single-stranded gaps between annealed oligos.
- **Ligation:** T4 DNA Ligase seals the remaining nicks to create a continuous double-stranded gene.
- **Hierarchical Method:** Larger genes ($\geq 1,000$ bp) are assembled from 500-bp verified building blocks.

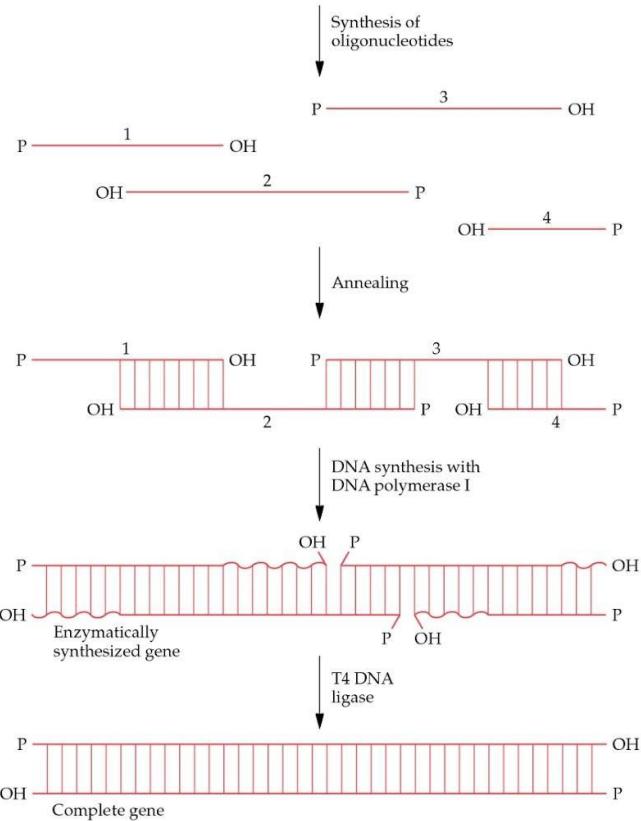


Figure 2.25 Assembly of a gene from oligonucleotides. Individual oligonucleotides are synthesized chemically and then hybridized. The sequences of the oligonucleotides are designed to enable them to form a stable molecule with base-paired regions separated by single-stranded regions (gaps). The gaps are filled in by *in vitro* enzymatic DNA synthesis. The nicks are sealed with T4 DNA ligase.

Gibson Assembly: One-pot Seamless DNA Joining

- **Seamless Joining:** Assembles multiple DNA fragments in a single isothermal reaction without leaving "scars" or using restriction sites.
- **5' Exonuclease:** Chews back 5' ends to expose 3' single-stranded overhangs that anneal to adjacent fragments.
- **Polishing & Sealing:** DNA polymerase fills the gaps, and DNA ligase seals the nicks covalently.
- **Scale:** Can assemble molecules up to several hundred kilobases, used to create the first synthetic bacterial genome.

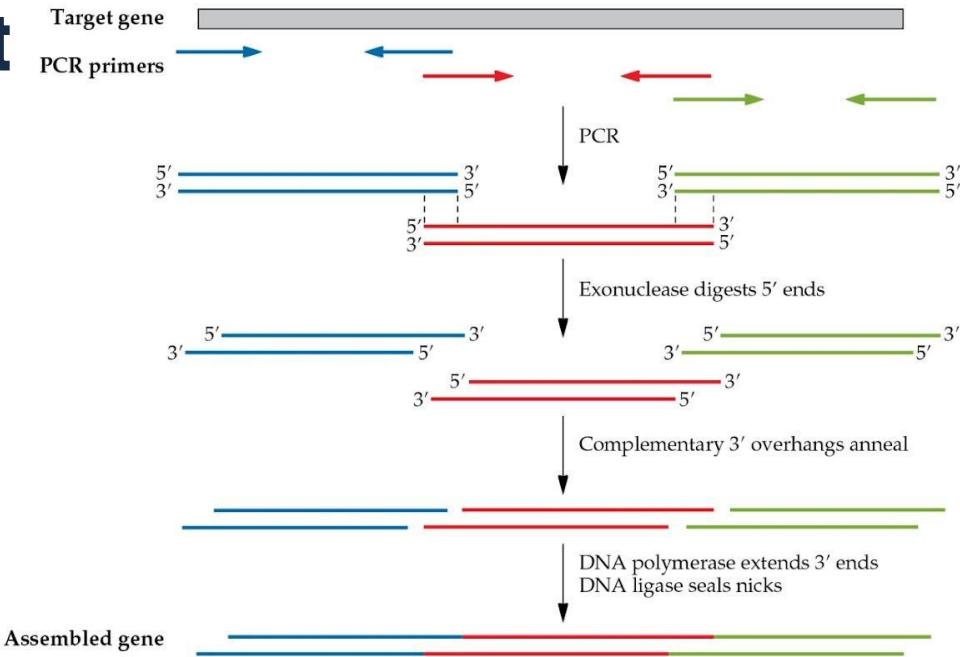


Figure 2.26 Assembly of a gene from PCR products (Gibson assembly). PCR fragments spanning the entire target gene contain sequences (15 to 25 nucleotides) at their ends that are homologous to the ends of adjacent fragments (indicated by broken lines). The fragments are mixed together and incubated with an exonuclease that digests the 5' ends of all fragments to create single-stranded 3' overhangs that are complementary to the 3' overhang on an adjacent fragment. The complementary regions anneal, and DNA polymerase catalyzes the addition of nucleotides to close the gaps. DNA ligase catalyzes the formation of phosphodiester bonds to seal nicks in the backbone of the assembled DNA.

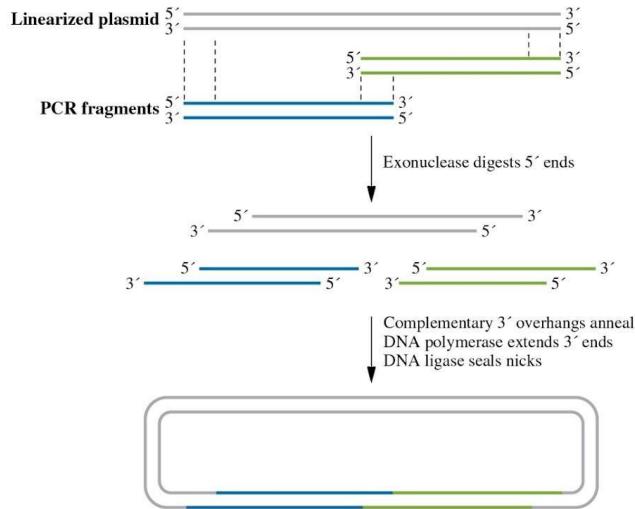


Figure 2.27 Gibson cloning. One or more PCR fragments may be inserted into a plasmid vector using the Gibson assembly procedure described in Fig. 2.26. The ends of the PCR fragments that will be adjacent to the plasmid contain a region of 15 to 25 nucleotides that is homologous to the plasmid sequence at the insertion site (homologous regions are indicated by broken lines). The plasmid sequences are included on the primers used to generate the PCR products. The plasmid is linearized (for example, by restriction enzyme digestion) and mixed with the PCR fragments. Following incubation with a 5' exonuclease, the 3' overhangs on the plasmid anneal to complementary overhangs on the PCR fragments to generate a circular plasmid carrying the cloned PCR fragments.

Video - Concepts in the module or a demonstration

- Gibson assembly - developed by Daniel Gibson, a scientist at the J. Craig Venter Institute - 2008 - 2009

Extra Content:

- Published in 2010 - Science
- Used to make a synthetic genome and minimal cell - here in San Diego!
 - Prof. Feist worked on a project evolving it

Module 2: DNA Sequencing (DNAseq)

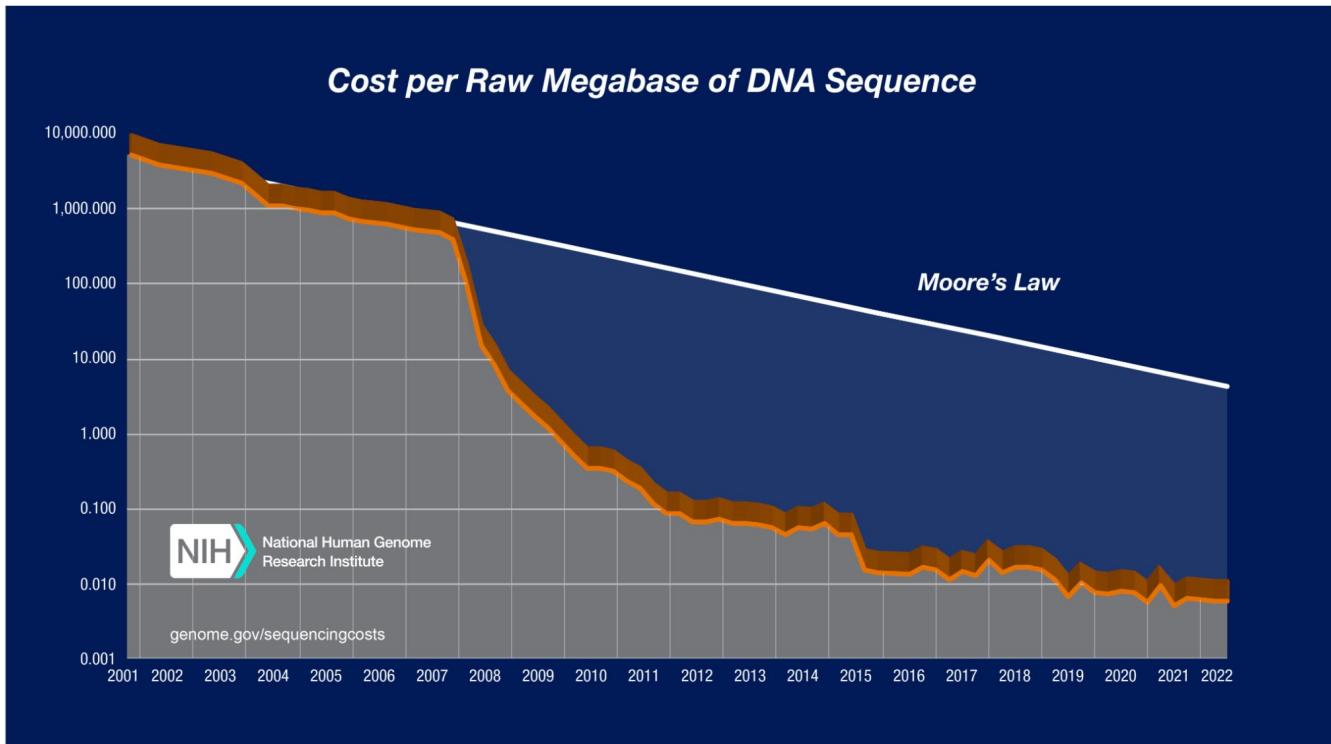
Technologies

(Source Pages: Chapter 2: 48– 61)

- **Sanger Sequencing (The Gold Standard)**
- **Next-Generation Sequencing (NGS/Illumina)**
- **3rd Gen: Single-Molecule Real-Time (PacBio)**
- **3rd Gen: Nanopore (Direct Threading)**
- **Genome Assembly and Metagenomics**

DNA Sequencing Costs: Data

Massive
Technology
Development
efforts have
enabled more
affordable
sequencing!



Sequencing cost per megabase - 2022

Sanger: Dideoxynucleotide Method

- **Chain Termination:** Incorporates ddNTPs (Dideoxynucleoside Triphosphates) that lack a 3'-OH group, preventing further phosphodiester bond formation.
- **Fluorescent Labeling:** Each of the four ddNTPs is tagged with a different color dye for laser identification.
- **Separation:** Capillary electrophoresis separates fragments differing by a single nucleotide, read by a laser as they pass.
- **Read Length:** Typically 500–1,000 bases with 99.9% accuracy.

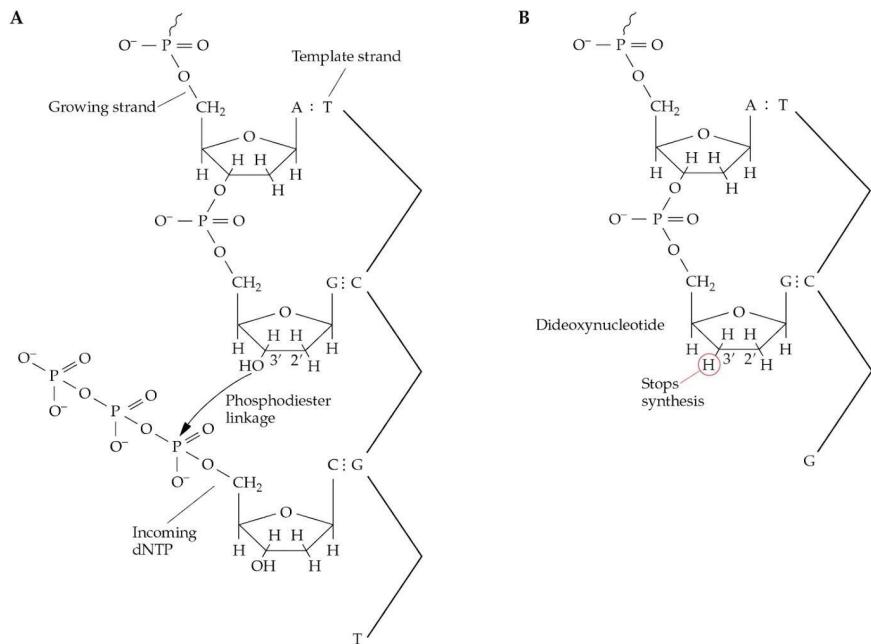


Figure 2.28 Incorporation of a dideoxynucleotide terminates DNA synthesis. **(A)** Addition of an incoming deoxyribonucleoside triphosphate (dNTP) requires a hydroxyl group on the 3' carbon of the last nucleotide of a growing DNA strand. **(B)** DNA synthesis stops if a synthetic dideoxynucleotide that lacks a 3' hydroxyl group is incorporated at the end of the growing chain because a phosphodiester bond cannot be formed with the next incoming nucleotide.

Sanger: Dideoxynucleotide Method

- **Chain Termination:** Incorporates ddNTPs that lack a 3'-OH group, preventing further phosphodiester bond formation.
- **Fluorescent Labeling:** Each of the four ddNTPs is tagged with a different color dye for laser identification.
- **Separation:** Capillary electrophoresis separates fragments differing by a single nucleotide, read by a laser as they pass.
- **Read Length:** Typically 500–1,000 bases with 99.9% accuracy.

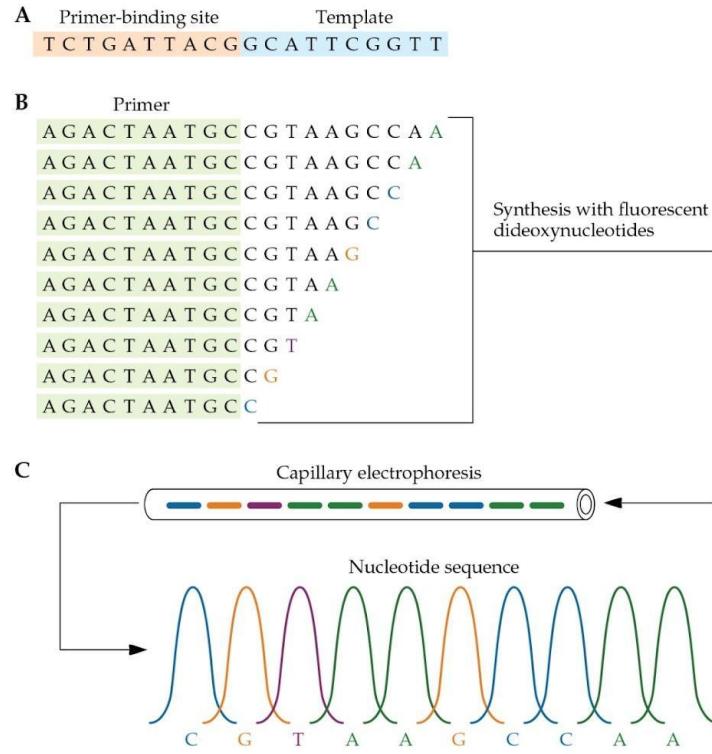


Figure 2.29 Dideoxynucleotide method for DNA sequencing. An oligonucleotide primer binds to a complementary sequence adjacent to the region to be sequenced in a single-stranded DNA template (**A**). As DNA synthesis proceeds from the primer, dideoxynucleotides are randomly added to the growing DNA strands, thereby terminating strand extension. This results in DNA molecules of all possible lengths that have a fluorescently labeled dideoxynucleotide at the 3' end (**B**). DNA molecules of different sizes are separated by capillary electrophoresis, and as each molecule passes by a laser, a fluorescent signal that corresponds with one of the four dideoxynucleotides is recorded. The successive fluorescent signals are represented as a sequencing chromatogram (colored peaks) (**C**).

Next-Gen: Illumina Sequencing

- Reversible Chain Terminators:**
Modified dNTPs have a removable blocking group on the 3'-OH and a cleavable fluorescent dye.
- Process:** One base is added, the color is recorded, the block and dye are removed, and the cycle repeats.
- Massive Parallelization:**
Simultaneously sequences millions of fragments on a glass slide flow cell.
- Outcome:** High throughput at a significantly lower cost per base

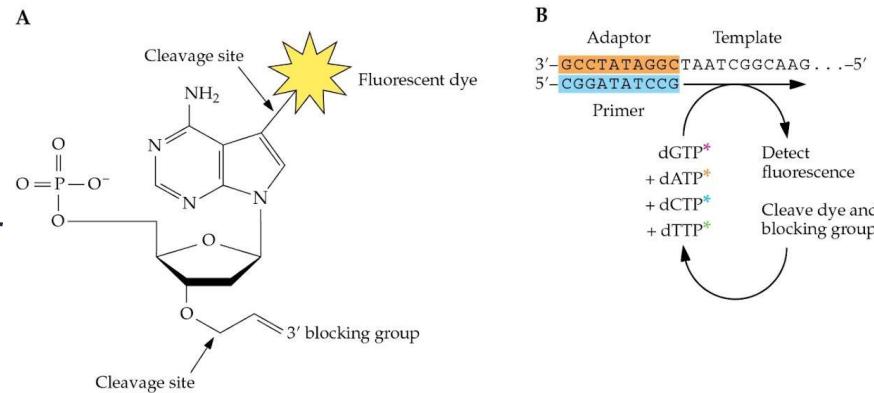


Figure 2.30 Sequencing using reversible chain terminators. **(A)** Reversible chain terminators are modified nucleotides that have a removable blocking group on the oxygen of the 3' position of the deoxyribose sugar to prevent addition of more than one nucleotide per sequencing cycle. To enable identification, a different fluorescent dye is attached to each of the four nucleotides via a cleavable linker. Shown is the fluorescent dye attached to adenine. **(B)** An adaptor sequence is added to the 3' end of the DNA sequencing template that provides a binding site for a sequencing primer. All four modified nucleotides are added in a single cycle, and a modified DNA polymerase extends the growing DNA chain by one nucleotide per cycle. Fluorescence is detected, and then the dye and the 3' blocking group are cleaved before the next cycle. Removal of the blocking group restores the 3' hydroxyl group for addition of the next nucleotide.

Bridge PCR and Cluster Generation

- **Capture:** Single-stranded DNA templates hybridize to oligonucleotides covalently bound to a glass slide surface.
- **Amplification:** Bridge PCR generates clusters of ~1,000 identical copies of each original DNA fragment.
- **Signal Boost:** This amplification ensures the fluorescent signal from each added base is strong enough for the camera to detect.
- **Cycle:** Hundreds of millions of clusters are sequenced simultaneously.

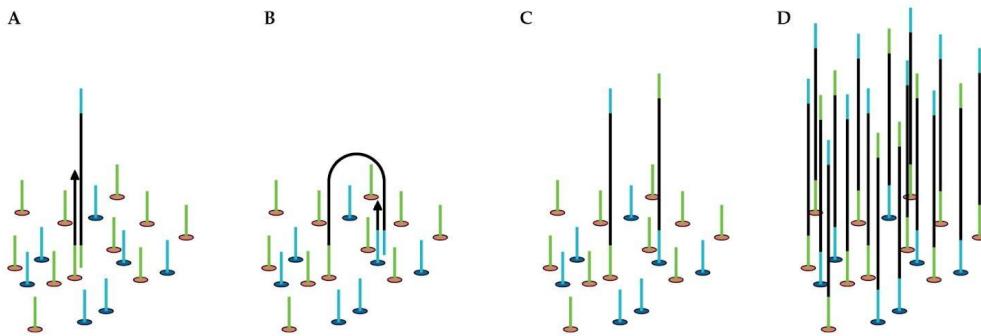


Figure 2.35 Generation of clusters of sequencing templates. Denatured genomic DNA library fragments are captured on a glass slide by annealing to a bound oligonucleotide via a complementary adaptor sequence (**A**). The oligonucleotide primes the synthesis of the complementary strand (denoted by the black arrow). The resulting double-stranded DNA is denatured, and the original library fragment is washed away. The newly synthesized strand remains anchored to the glass slide at one end and binds to an adjacent oligonucleotide primer by the adaptor sequence at the other end (**B**). The complementary strand is synthesized by extension of the second primer in a process known as bridge amplification. Denaturation of the double-stranded product results in two single-stranded DNA molecules that are bound to the glass slide (**C**). The process is repeated many times to generate clusters of about a thousand copies of each sequencing template (**D**).

3rd Gen: SMRT Sequencing (PacBio)

- **SMRT:** Single-Molecule Real-Time sequencing uses a single immobilized DNA polymerase in a nanoscale well.
- **Real-Time Detection:** Fluorescent tags are attached to the terminal phosphate; they are released and detected as the base is added.
- **Continuous Reads:** Generates long sequence reads (30–50 kb) because synthesis is not interrupted by blocking groups.
- **Advantages:** Can detect epigenetic modifications like methylated bases.

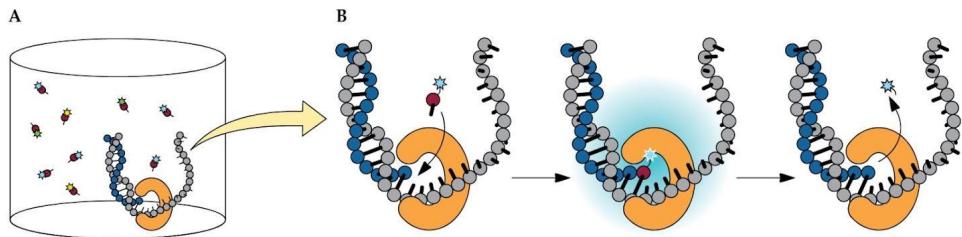


Figure 2.31 Single-molecule real-time sequencing. One molecule of DNA polymerase (orange shape) is attached to the bottom of a nanoscale well. A single-stranded DNA molecule (gray strand) bound to a primer (blue strand) is captured in the active site of the polymerase (**A**). Each of the four different nucleoside triphosphates is attached to a different fluorophore (colored stars) at the terminal phosphate, which is released during template-dependent nucleotide incorporation into the growing DNA strand. Fluorescence emission from a zeptoliter (10^{-21} liter) volume at the bottom of the well is detected by a laser before the cleaved pyrophosphate with attached fluorophore diffuses away (**B**).

3rd Gen: Nanopore Sequencing

- **Direct Threading:** A motor protein (e.g., modified DNA polymerase) threads a single DNA strand through a tiny protein pore in a synthetic membrane.
- **Electrical Signal:** As bases pass through the pore, they disrupt an electric current in a unique, characteristic way.
- **Portability:** Available in handheld formats (MINION) for environmental monitoring and even use on the International Space Station.
- **Ultra-Long Reads:** Can read single strands longer than 100 kb.

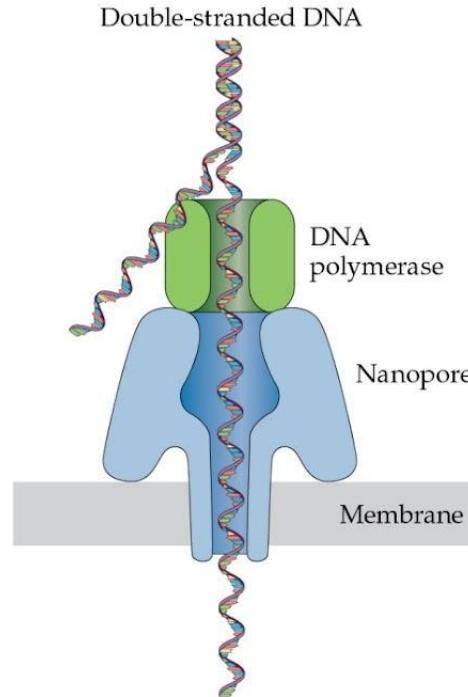


Figure 2.32 Nanopore sequencing. A double-stranded DNA molecule is bound to an adapter and a tethering protein that direct it to a protein nanopore embedded in a synthetic membrane. A motor protein such as DNA polymerase or helicase unwinds the DNA and threads one strand through the nanopore. As the DNA passes through the pore, an electrical current fluctuates in a manner that is characteristic of the nucleotides in the pore which enable their identification. The processivity of DNA polymerase facilitates the sequencing of long DNA molecules (several hundred kilobases) in a single run.

Comparison of Sequencing Technologies

- Choosing which sequencing method to use is influenced by a few key differentiating factors.
 - Cost is not shown here
 - Error rates improving rapidly

Table 2.3 Comparison of sequencing technologies

Technology	Method	Read length (kb)	Error rate (%)	Advantages
Dideoxynucleotide sequencing (Sanger)	Synthesis	0.5–1	0.001	High accuracy
Sequencing using reversible chain terminators (Illumina)	Synthesis	0.1–0.3	0.1	Low cost, accurate
Single-molecule real-time sequencing (PacBio)	Synthesis	30–50	5–15	Rapid, continuous long reads, template amplification not required
Nanopore sequencing	Nanopore	10–200	5–20	Very long reads, portable, template amplification not required

Genome Assembly: Contigs and Scaffolds

- **Reads:** Short sequences (50–300 bp for Illumina) generated by the sequencers.
- **Contigs:** Computer programs align overlapping reads to form long, contiguous stretches of nucleotides.
- **Scaffolds:** Paired-end reads (sequences from both ends of a known fragment) help order and orient contigs despite repetitive regions.
- **Gap Closure:** Small gaps (<10 kb) can be closed by targeted PCR and Sanger sequencing.

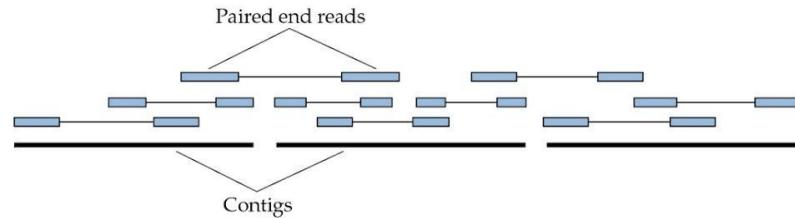


Figure 2.36 Genome sequence assembly. Sequence data generated from both ends of a DNA fragment are known as paired-end reads (paired ends are shown in blue for each fragment, and the distance between them is represented by a thin, black line). A large number of reads are generated and assembled into longer contiguous sequences (contigs) using a computer program that matches overlapping sequences. Paired ends help to determine the order and orientation of contigs as they are assembled into scaffolds. Shown is a scaffold consisting of three contigs.

Genome Assembly: Contigs and Scaffolds

- Genome Assembly is common and foundational for current microbial research



AMERICAN
SOCIETY FOR
MICROBIOLOGY



Microbiology
Resource Announcements

GENOME SEQUENCES



Gapless, Unambiguous Genome Sequence for *Escherichia coli* C, a Workhorse of Industrial Biology

Jonathan E. Pekar,^a Patrick Phaneuf,^a Richard Szubin,^a Bernhard Palsson,^a Adam Feist,^a Jonathan M. Monk^a

^aBioengineering Department, University of California San Diego, La Jolla, California, USA

ABSTRACT *Escherichia coli* C is a commonly used strain in the bioprocessing industry, but despite its utility, the publicly available sequence of the *E. coli* C genome has gaps and 4,180 ambiguous base calls. Here, we present an updated, high-quality, unambiguous genome sequence with no assembly gaps.

*E*scherichia coli C, an often-used industrial strain, was originally isolated by Ferdinand Hueppe from soured cow's milk and described in his 1884 publication in German (1, 2). The C strain was used extensively by Bertani and collaborators in studies of phage P2 and by many others as a strain lacking a type I restriction-modification system (3). The strain was originally termed NCTC 122 at the National Collection of Type Cultures in London, United Kingdom, where its entry states that it was deposited by the Lister Institute (London) in 1920 as "*Escherichia coli*" (4). This strain was recently featured in a publication comparing seven commonly used *E. coli* platform strains and was shown to have high anaerobic growth rates and predicted to have high relative production

Metagenomics: The Uncultured World

- **Definition:** The study of collective genomes extracted directly from environmental samples like soil, seawater, or the human gut.
- **The 1% Rule:** Fewer than 1% of all bacterial species can be cultured in a lab; metagenomics bypasses the need for cultivation.
- **Discovery:** Used to find novel biosynthetic gene clusters for new antiviral or antitumor polyketides in sponges.

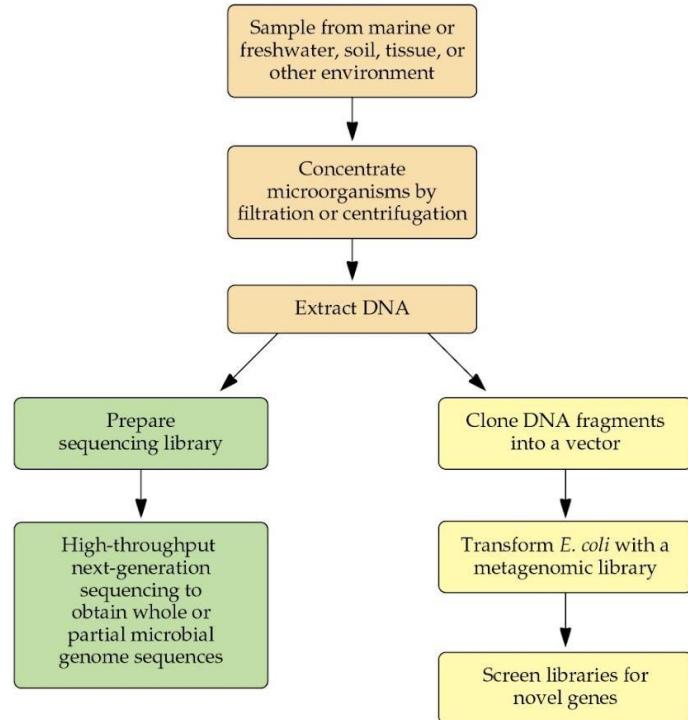
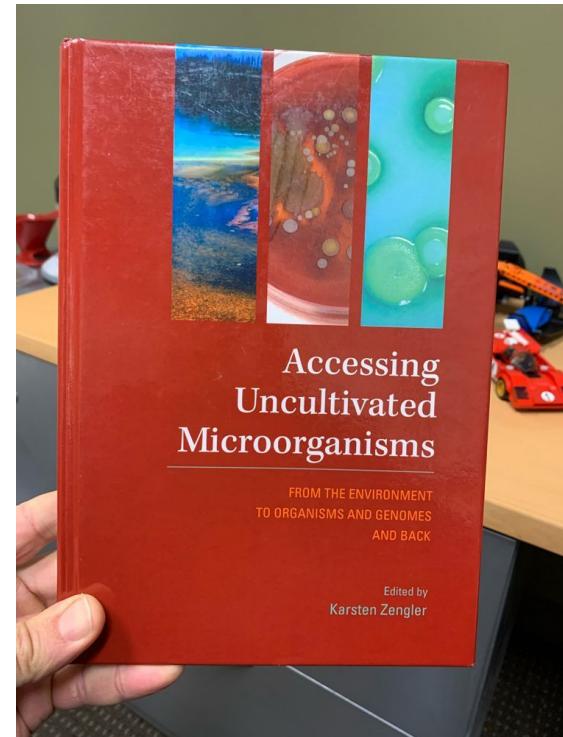


Figure 2.37 Construction of metagenomic libraries. Bacteria and/or viruses in samples from various environments or tissues are concentrated before extracting and then fragmenting the DNA. Libraries containing the DNA fragments are sequenced or screened for novel genes.

Metagenomics: The Uncultured World

- **Definition:** The study of collective genomes extracted directly from environmental samples like soil, seawater, or the human gut.
- **The 1% Rule:** Fewer than 1% of all bacterial species can be cultured in a lab; metagenomics bypasses the need for cultivation.
- **Discovery:** Used to find novel biosynthetic gene clusters for new antiviral or antitumor polyketides in sponges.

- Zengler, K. (ed) **Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back**, 2008.



Video - Concepts in the module or a demonstration

- Video visualizing different sequencing techniques.
 - [Sanger sequencing Technique](#) (1min30s)
 - [Illumina Sequencing](#) (5min12s)
 - [PacBio SMRT Sequencing](#) (1min30s)
 - [Nanopore Sequencing](#) (1min42s)

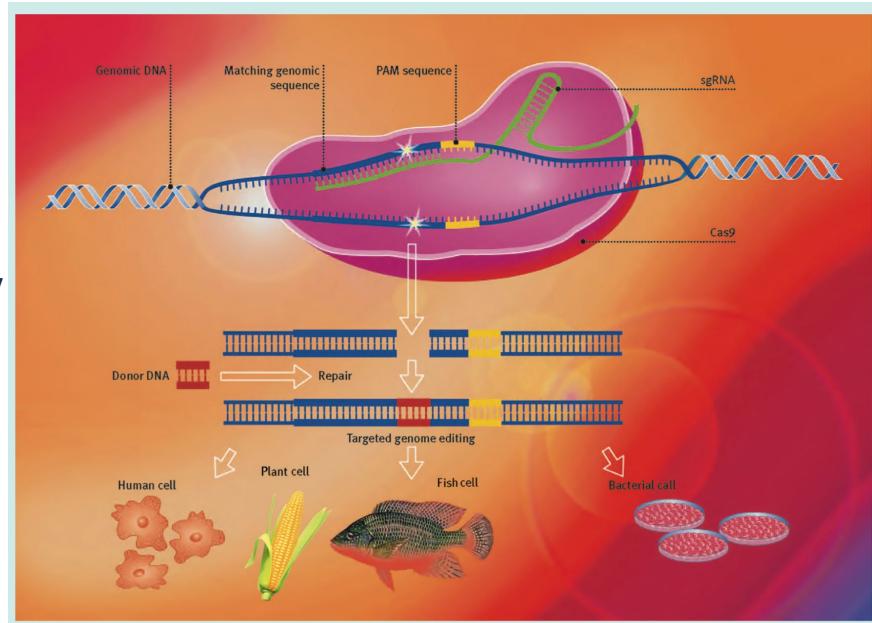
Module 3: Genome Engineering Using CRISPR-Cas9

(Source Pages: Chapter 2: 36–39)

- The Bacterial Origin of CRISPR
- Cas9: The Molecular Scissors
- sgRNA: The Programmable Guide
- Repair Pathways: NHEJ (Nonhomologous End Joining) vs. HDR (Homology-Directed Repair)

Why Care About CRISPR?

- It acts as a universal, programmable tool for precise genome editing in diverse organisms.
- CRISPR offers the potential to cure over ten thousand monogenic diseases at their source.
- Gene drives could eradicate malaria by rapidly spreading modifications through specific wild mosquito populations.
- For the first time, humans can purposefully manipulate the entire biological world.
- Future bioengineers must navigate profound ethical dilemmas like the boundary between healing and perfection.



CRISPR: The Bacterial Immune System

- **Natural Defense:** Protects bacteria from invading bacteriophages and plasmids.
- **Memory (Spacers):** Fragments of viral DNA are incorporated as "spacers" between repeated sequences in the CRISPR locus.
- **Endonuclease:** enzyme that cuts or cleaves internal phosphodiester bonds within a DNA or RNA strand
- **Mechanism:** On re-infection, spacers are transcribed into crRNA that guides an endonuclease (Cas) to cleave the viral DNA.

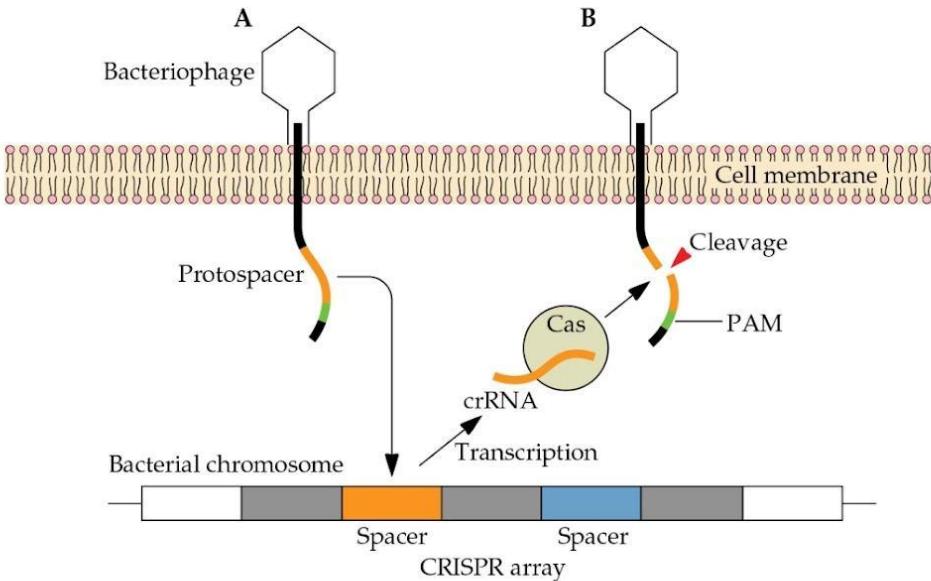


Figure 2.17 Bacterial CRISPR-Cas system for protection against invading bacteriophage. **(A)** Fragments of bacteriophage DNA (protospacer) are incorporated into the host bacterial genome as spacers between repeat sequences (gray) in the CRISPR array. **(B)** On subsequent invasion, the spacer DNA is transcribed to produce CRISPR RNA (crRNA) that guides an endonuclease (Cas) to a sequence in the invading DNA that is homologous to the spacer sequence and is adjacent to a protospacer-adjacent motif (PAM). The viral genome is cleaved.

The CRISPR-Cas9 Toolkit

- **Cas9:** A dual-RNA-guided DNA endonuclease that generates double-stranded breaks.
- **sgRNA (Single Guide RNA):** An engineered fusion of crRNA (CRISPR RNA) and tracrRNA (~80–100 nt) that targets a 20-nt sequence in the genome.
 - For genome engineering, we combine them instead of them forming their natural hybrid.
- **PAM (Protospacer-Adjacent Motif):** Cas9 only cleaves if the target is adjacent to the "NGG" motif; this prevents the bacteria from cutting its own genome.

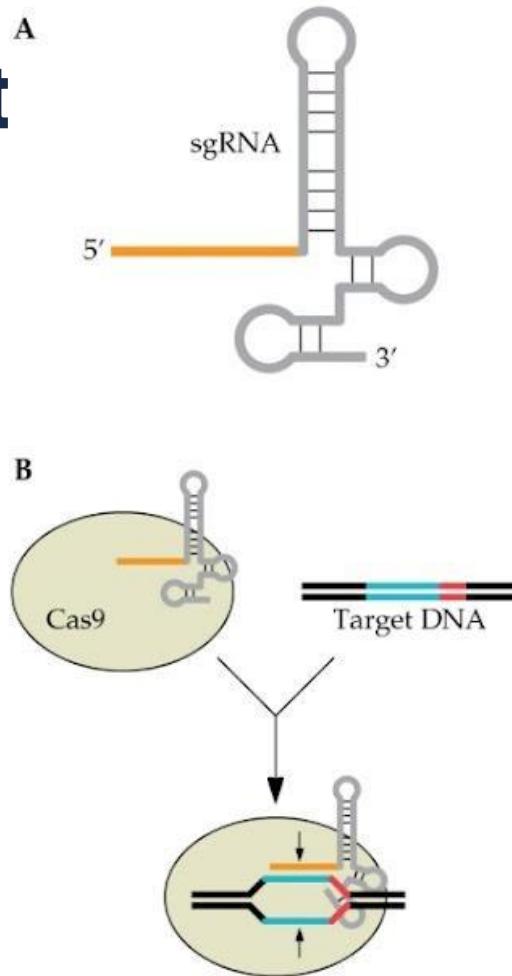


Figure 2.18 CRISPR-Cas system for genome editing. **(A)** An 80- to 100-nucleotide-long single guide RNA (sgRNA) is constructed that contains a 20-nucleotide guide sequence (orange) that is complementary to the target site. The secondary structure, stabilized by intramolecular base pairing between regions of the fused crRNA and tracrRNA sequences, is required for binding to the Cas9 endonuclease. **(B)** The sgRNA guides Cas9 to the target sequence (blue) in the genome. Target recognition requires an adjacent PAM sequence (red) NGG and complementarity between the guide sequence and the target sequence. Cas9 makes a double-stranded break in the target DNA (arrows) which is repaired by homologous recombination or nonhomologous end joining. The repair systems generate deletions and insertions at the target site.

The Cleavage and Targeting Mechanism

- Cas9 scans the DNA for a PAM sequence (NGG). millions of these in a genome!
- The sgRNA unzips the DNA and base-pairs with the target sequence.
- If there is perfect complementarity, Cas9's nuclease domains make a precise double-strand break (DSB) **exactly 3 bp upstream of the PAM**.
- This DSB **activates the cell's internal DNA repair machinery**.

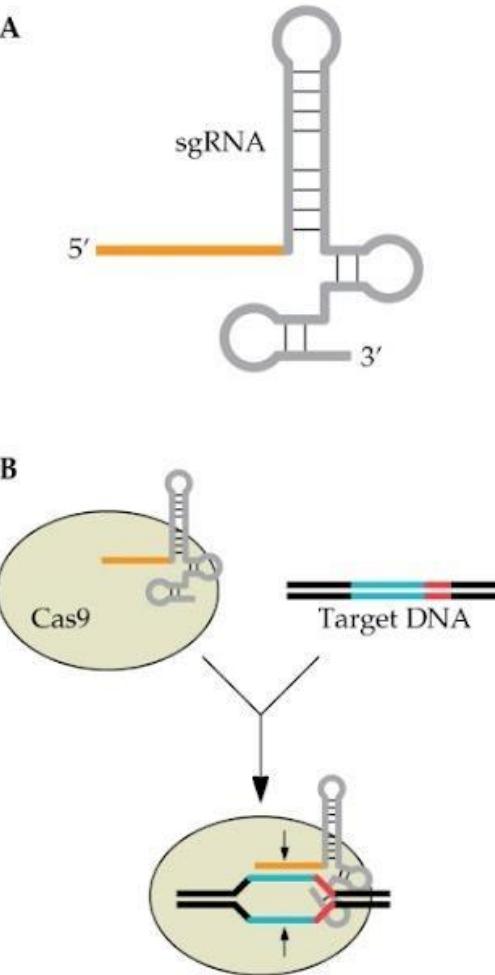


Figure 2.18 CRISPR-Cas system for genome editing. **(A)** An 80- to 100-nucleotide-long single guide RNA (sgRNA) is constructed that contains a 20-nucleotide guide sequence (orange) that is complementary to the target site. The secondary structure, stabilized by intramolecular base pairing between regions of the fused crRNA and tracrRNA sequences, is required for binding to the Cas9 endonuclease. **(B)** The sgRNA guides Cas9 to the target sequence (blue) in the genome. Target recognition requires an adjacent PAM sequence (red) NGG and complementarity between the guide sequence and the target sequence. Cas9 makes a double-stranded break in the target DNA (arrows) which is repaired by homologous recombination or nonhomologous end joining. The repair systems generate deletions and insertions at the target site.

NHEJ: The "Knockout" Repair Pathway

- **Nonhomologous End Joining (NHEJ):** The cell's primary repair mechanism, which is rapid but highly error-prone.
- **Indels:** The process often results in small insertions or deletions of nucleotides.
- **Gene Disruption:** These "indels" cause frameshift mutations that disrupt the reading frame, effectively "knocking out" the gene function.

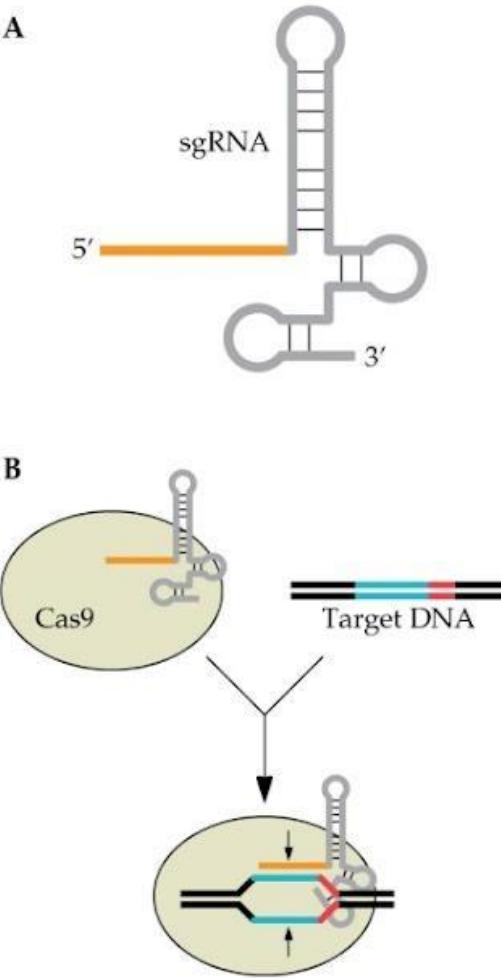
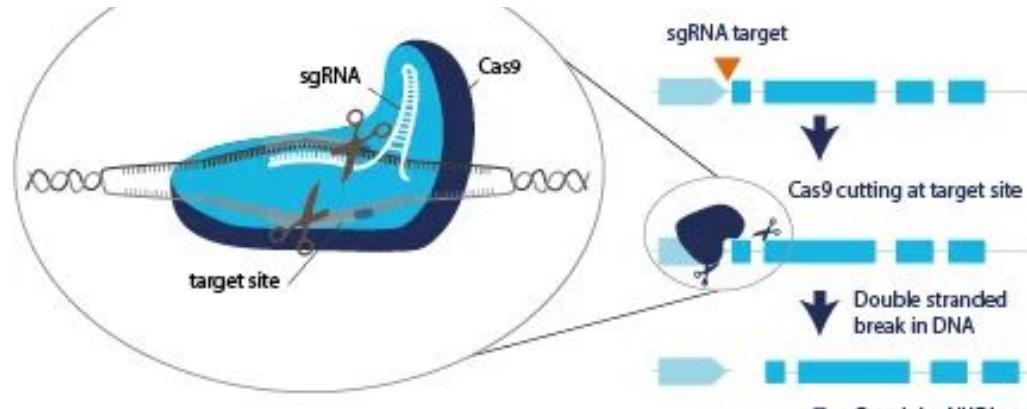


Figure 2.18 CRISPR-Cas system for genome editing. **(A)** An 80- to 100-nucleotide-long single guide RNA (sgRNA) is constructed that contains a 20-nucleotide guide sequence (orange) that is complementary to the target site. The secondary structure, stabilized by intramolecular base pairing between regions of the fused crRNA and tracrRNA sequences, is required for binding to the Cas9 endonuclease. **(B)** The sgRNA guides Cas9 to the target sequence (blue) in the genome. Target recognition requires an adjacent PAM sequence (red) NGG and complementarity between the guide sequence and the target sequence. Cas9 makes a double-stranded break in the target DNA (arrows) which is repaired by homologous recombination or nonhomologous end joining. The repair systems generate deletions and insertions at the target site.

NHEJ: The "Knockout" Repair Pathway

- **Nonhomologous End Joining (NHEJ):** The cell's primary repair mechanism, which is rapid but highly error-prone.



www.zeclinics.com

INDELS Contribution Sequence

-9	43%	CITGACCAAAATCTAGAGCCAGAT	CITGICATAGGTGCTGACATTTGGCATE
-3	17%	CITGAGGAAAAATCTAGAGCCAGATC	CGGACATCTGCTCATAGGCCTGACATTTGGCATE
-26	6%	CITGAGGAAAAATCTAGAGCCAGAT	TAGGTGGTGTGACATTTGGCATE
-16	4%	CITGAGGAAAAATGTAGAGCCAGAT	TAGGTGGTGTGACATTTGGCATE
+8	3%	CITGAGGAAAAATGTACACCCACGATC	NNNNNNNN	ATGCCGATCTTCATCATAGGTGGTCACA
+6	3%	CITGAGGAAAAATGTAGAGCCAGATC	NNNNNN	ATGCCGATCTTGTCATAGGTGGTCACA

Different resulting sequences can be obtained. Sequencing reveals molecular state.

HDR: The "Paste/Correct" Repair Pathway

- **Homology-Directed Repair (HDR):** A **high-fidelity** repair mechanism that uses a **DNA template** to fix breaks
- **Donor DNA:** Researchers provide an exogenous donor DNA fragment carrying a "correct" sequence flanked by homologous regions.
- **Gene Insertion:** Recombinases (enzyme that cuts and rejoins DNA) integrate the donor DNA into the genome, allowing for precise gene correction or insertion.
 - Double-strand break stimulates the cell's recombinase proteins

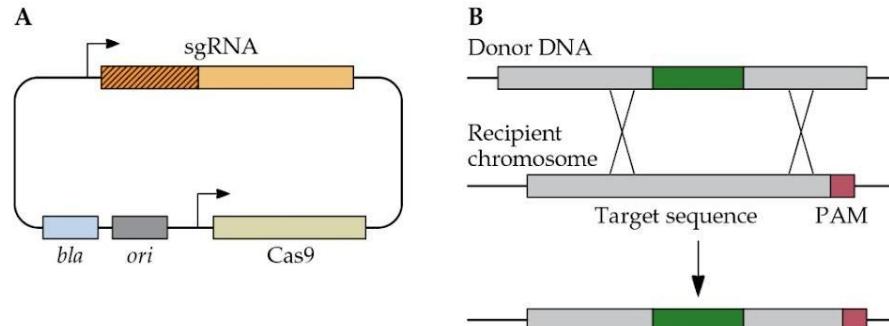


Figure 2.19 Vector for production of sgRNA and Cas9 in host cells **(A)**. The gene encoding sgRNA contains a 20-nucleotide sequence (hatched region) that is complementary to the target site in the host genome. Promoters (arrows) for the sgRNA and Cas9 genes, and codon usage for Cas9, must be suitable for expression in host cells. An origin of replication (*ori*) and a selectable marker (e.g., *bla* encoding β -lactamase, which confers resistance to the antibiotic ampicillin) are included for initial vector construction in *E. coli*. The vector and donor DNA are introduced into a recipient cell. Following expression, the sgRNA guides the Cas9 endonuclease to the target sequence in the recipient cell chromosome, and the endonuclease makes a double-stranded break in the target DNA. **(B)** The donor DNA sequence (green) is flanked by regions that are homologous to the target site (gray) for insertion by homologous recombination. Activation of recombinases that mediate DNA repair results in recombination between homologous sequences on the vector and in the recipient chromosome and, thereby, insertion of the donor DNA into the genome at the target site.

Video - Concepts in the module or a demonstration

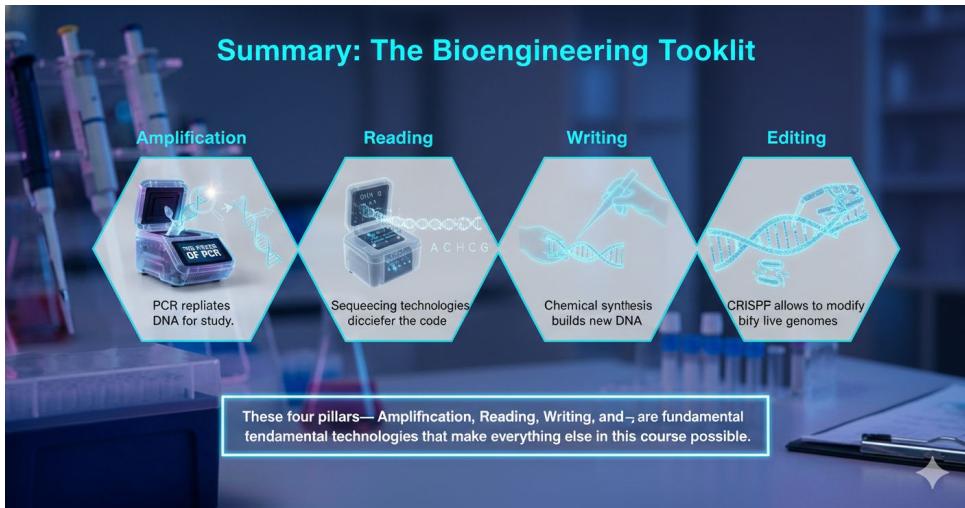
- [CRISPR Explained](#) (1min39s)
- [How CRISPR Lets You Edit DNA](#) (5min28s)
- [CRISPR: Gene Editing & Beyond](#) (4min31s)

Extra

- [Understanding CRISPR-CAs9](#) - long video

Summary: The Bioengineering Toolkit

- **Amplification:** PCR replicates DNA for study.
- **Reading:** Sequencing technologies decipher the code.
- **Writing:** Chemical synthesis builds new DNA.
- **Editing:** CRISPR allows us to modify live genomes.



The End