DNA technologies

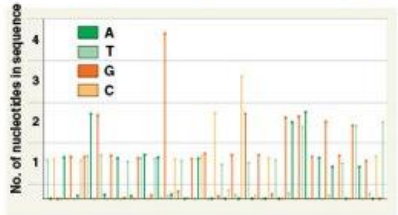
DNA technologies: introduction

- The main technologies for sequencing and manipulating DNA are called **DNA technologies**
- PCR, DNA sequencing and DNA cloning are valuable tools for genetic engineering
- The complementarity of the two DNA strands is the basis for **nucleic acid hybridization**, the base pairing of one strand of nucleic acid to the complementary sequence on another strand
- **Genetic engineering** is the direct manipulation of genes for practical purposes
- **Biotechnology** refers to the manipulation of organisms or their components to make useful products or services

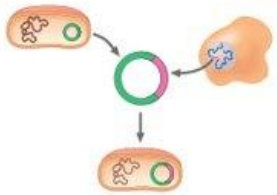
What are the main techniques and applications of biotechnology?

TECHNIQUES

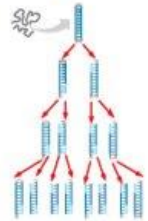
DNA sequencing



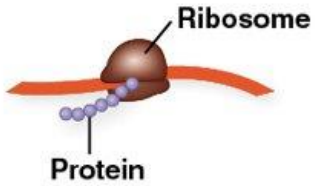
Gene cloning



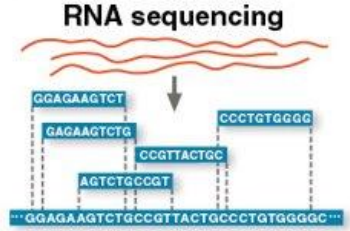
PCR amplification



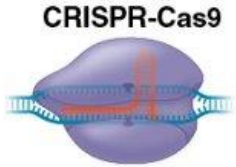
Expressing genes



Analyzing gene expression



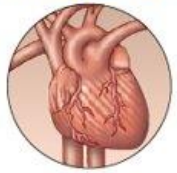
Gene editing



APPLICATIONS

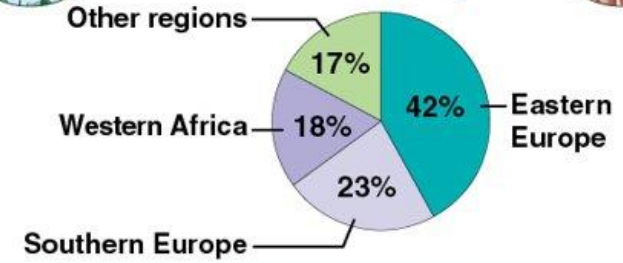
Agriculture

Medicine

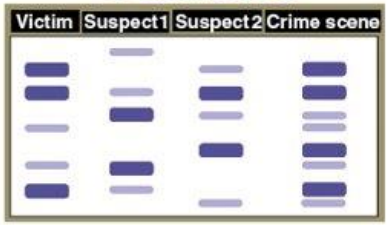


Environmental cleanup

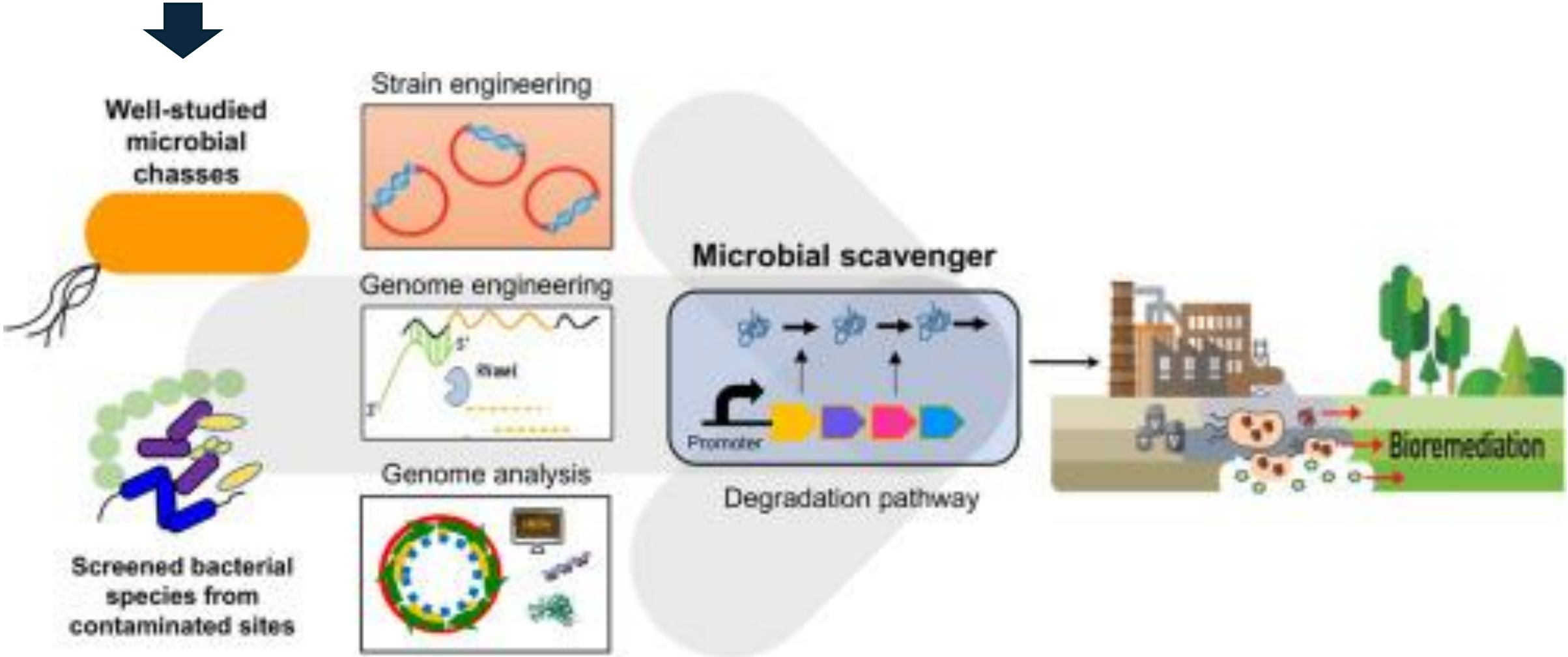
Ancestry



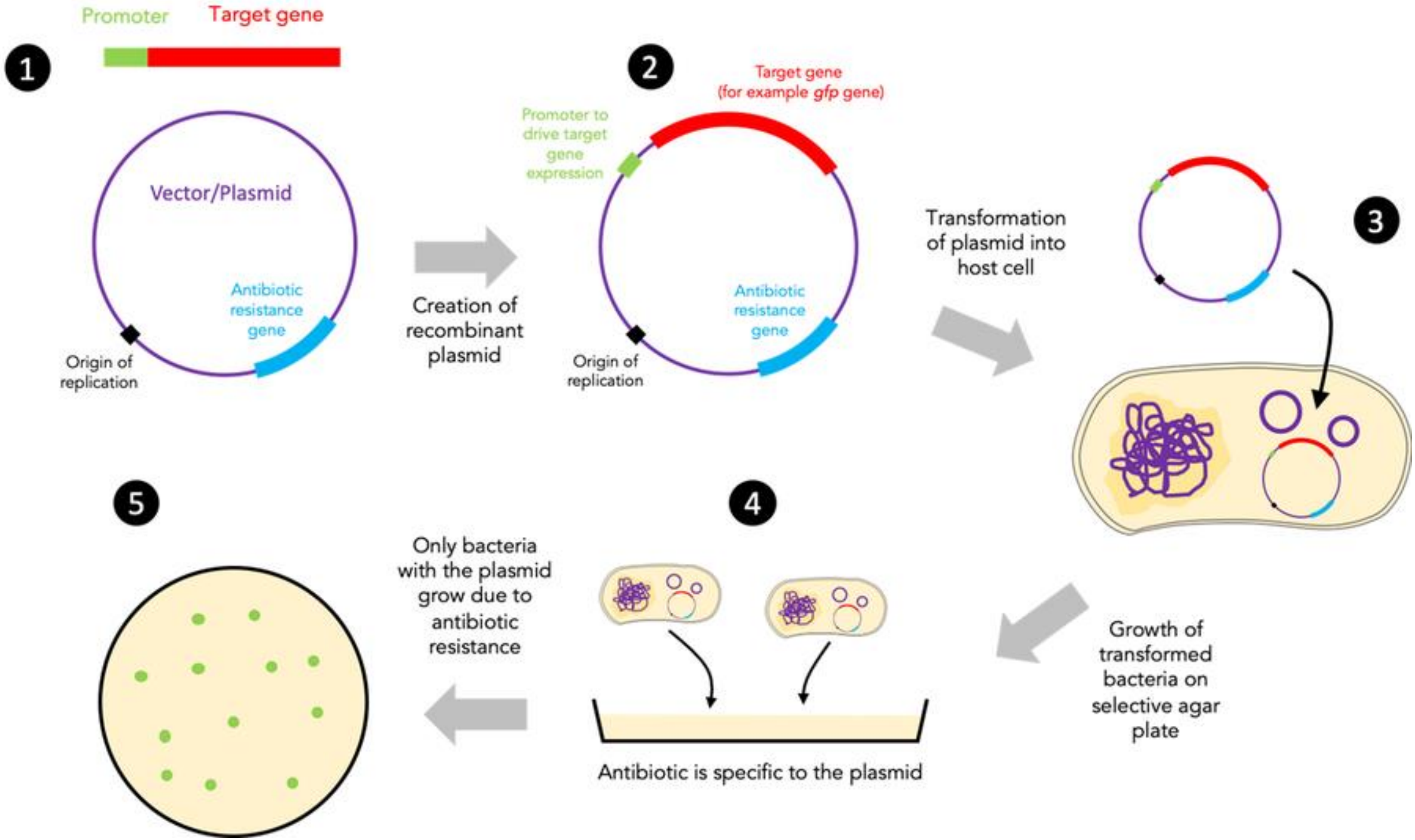
Forensics



PCR, sequencing



Tran et al. 2021


Promoter
Target gene
1
Vector/Plasmid
Antibiotic resistance gene
Origin of replication
Creation of recombinant plasmid
2
Target gene (for example gfp gene)
Promoter to drive target gene expression
Antibiotic resistance gene
Origin of replication
Transformation of plasmid into host cell
3
4
Antibiotic is specific to the plasmid
Growth of transformed bacteria on selective agar plate
Only bacteria with the plasmid grow due to antibiotic resistance
5

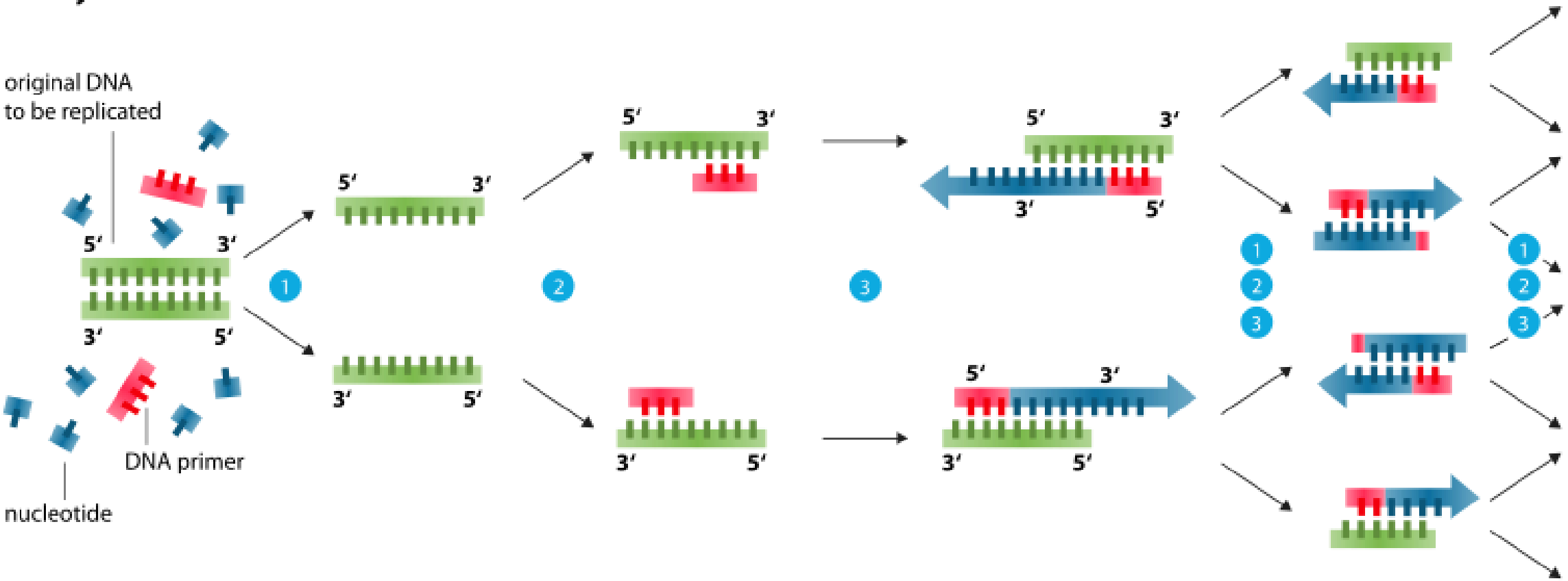
DNA amplification: PCR

Amplifying DNA: The Polymerase Chain Reaction (PCR)

- The **polymerase chain reaction**, **PCR**, can produce many copies of a specific target segment of DNA
- A three-step cycle—heating (denaturing), cooling (annealing), and extension—brings about a chain reaction that produces an exponentially growing population of identical DNA molecules
- The process uses primers, short single-stranded DNA molecules complementary to sequences to either side of the target sequence

- The key to PCR is an unusual, heat-stable enzyme, a DNA polymerase called ***Taq* polymerase**
- Other polymerases may be used as well; some are more accurate and stable than *Taq*, such as *Pfu* polymerase
- The primers used are specific for the sequence to be amplified
- PCR amplification occasionally incorporates errors into the amplified strands and so cannot substitute for gene cloning in cells

Polymerase chain reaction - PCR

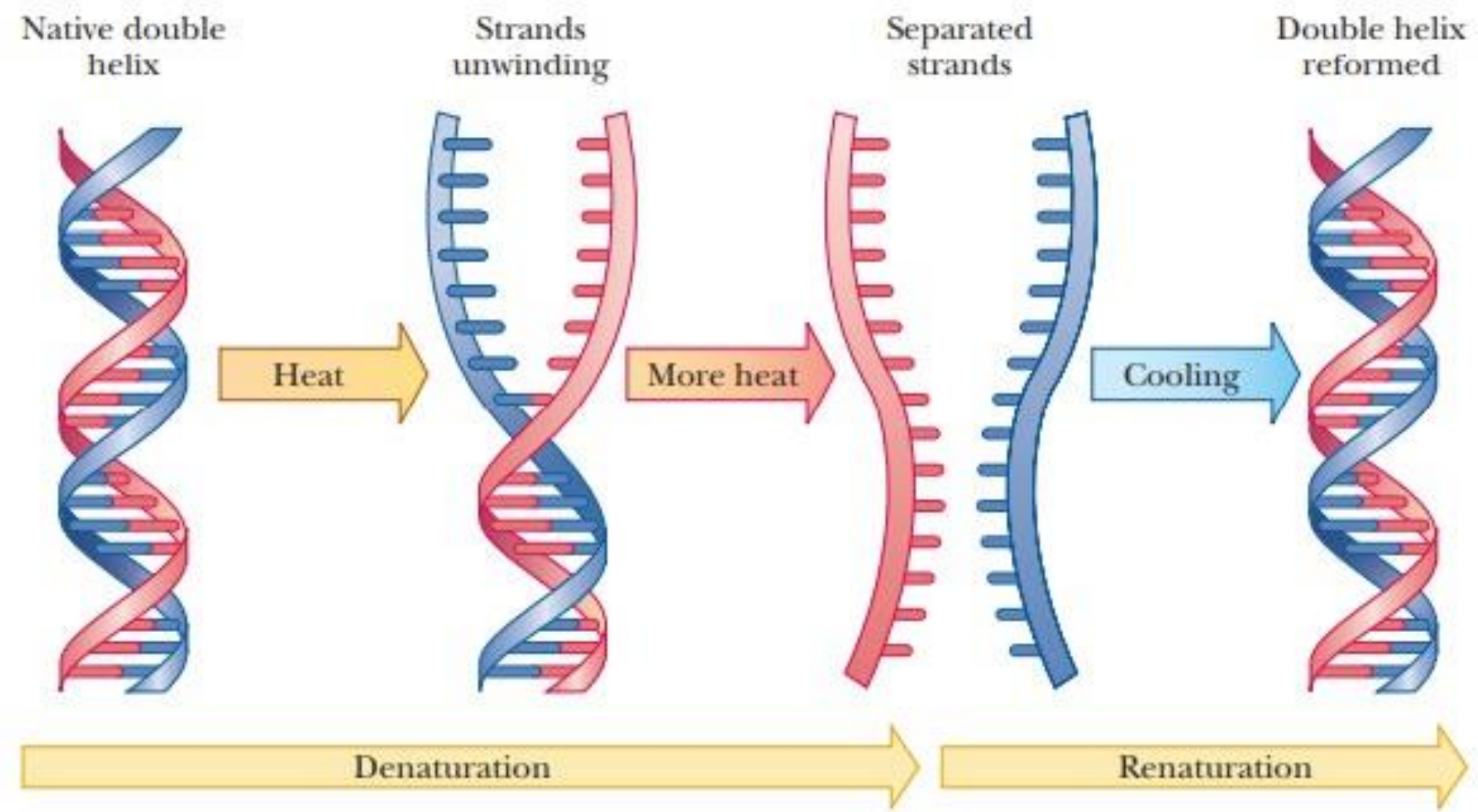


1 **Denaturation** at 94-96°C

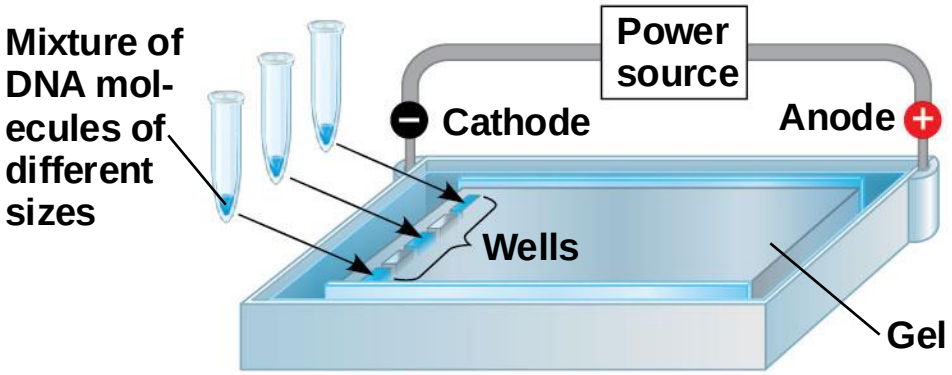
2 **Annealing** at ~68°C

3 **Elongation** at ca. 72 °C

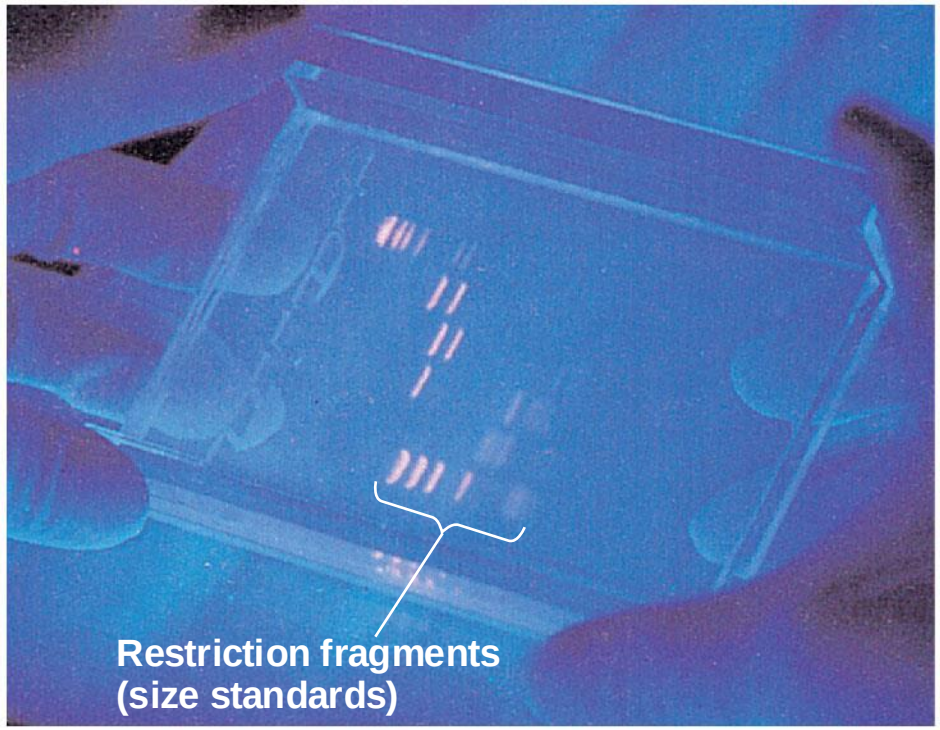
DNA, denaturation and renaturation



PCR:
Agarose electrophoresis



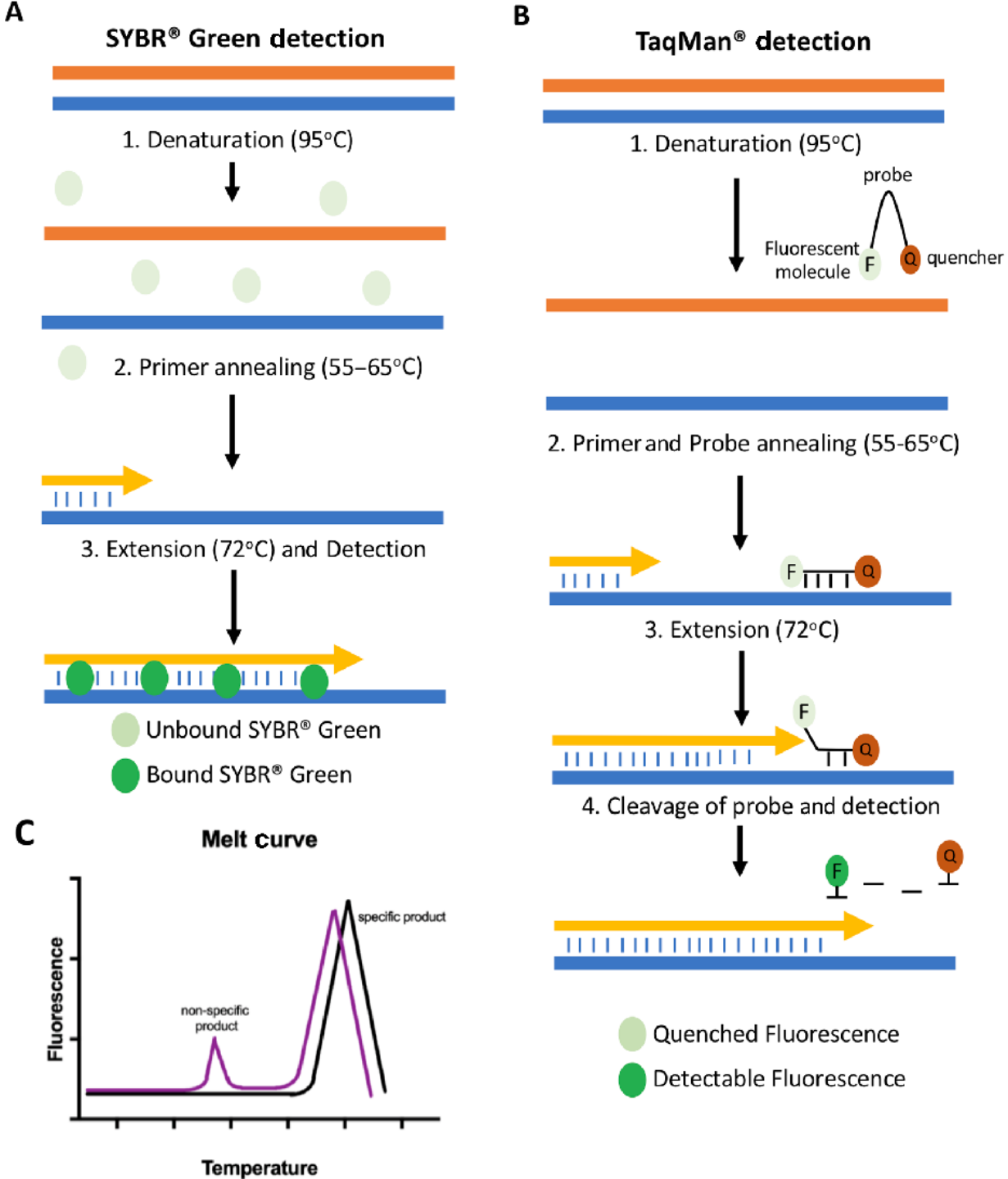
(a) Negatively charged DNA molecules will move toward the positive electrode.



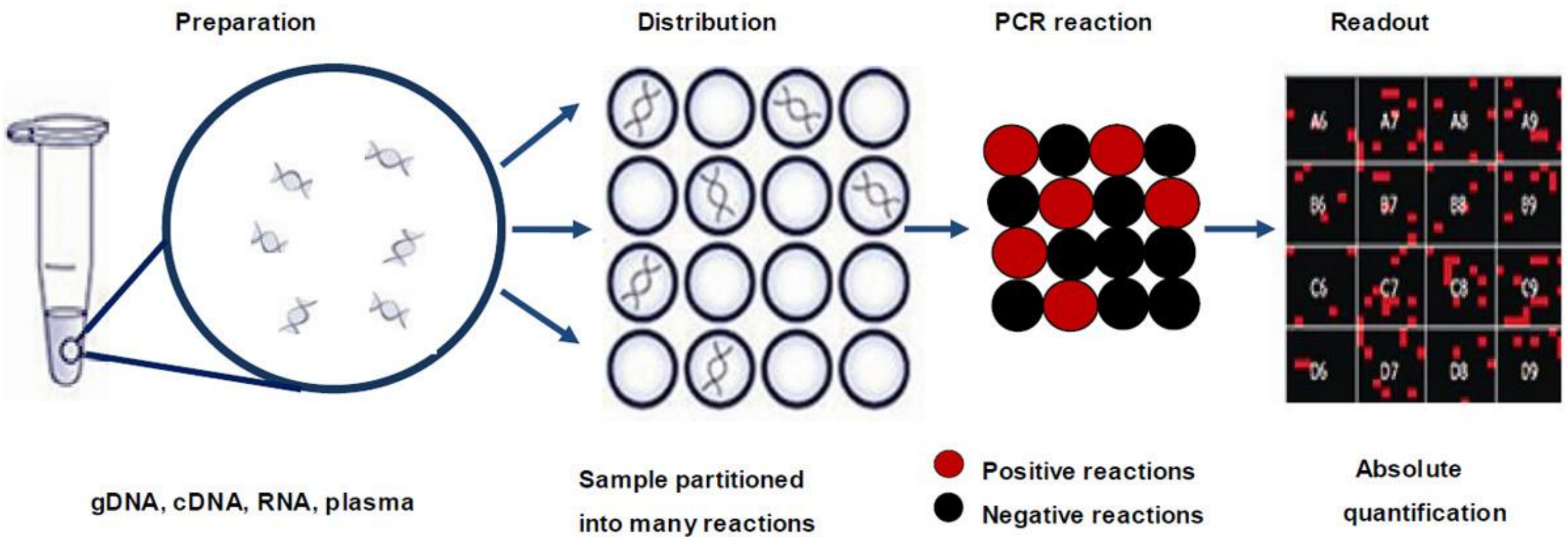
(b) Shorter molecules are slowed down less than longer ones, so they move faster through the gel.

qPCR

- Smaller target fragments (100 bp aprox. instead of few hundred)

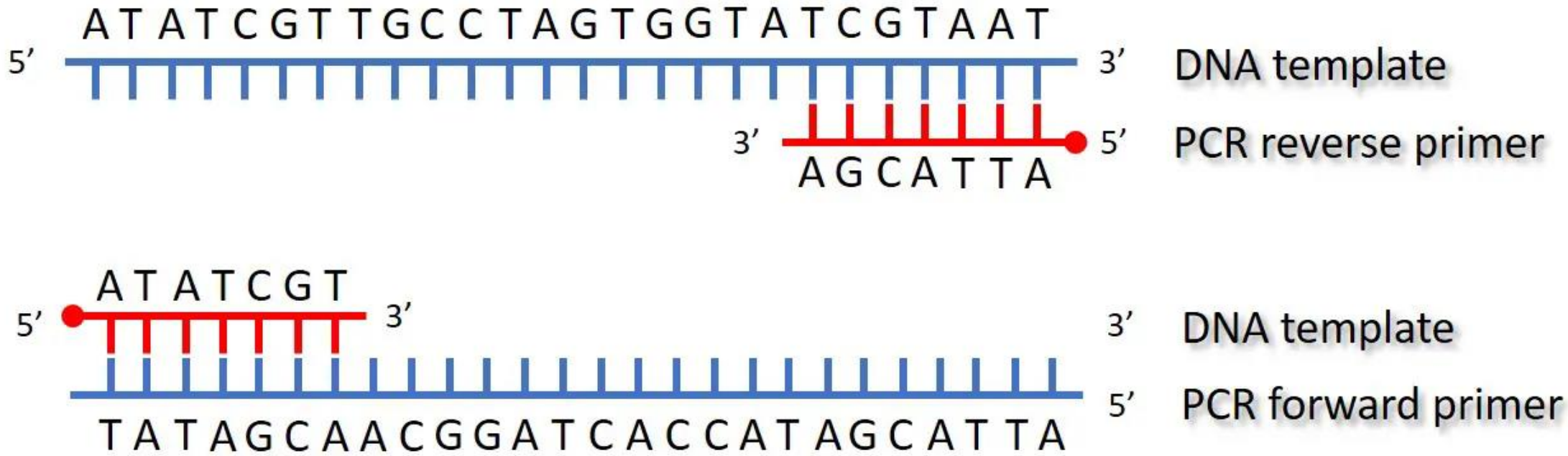


dPCR: digital PCR



Primer design: specificity

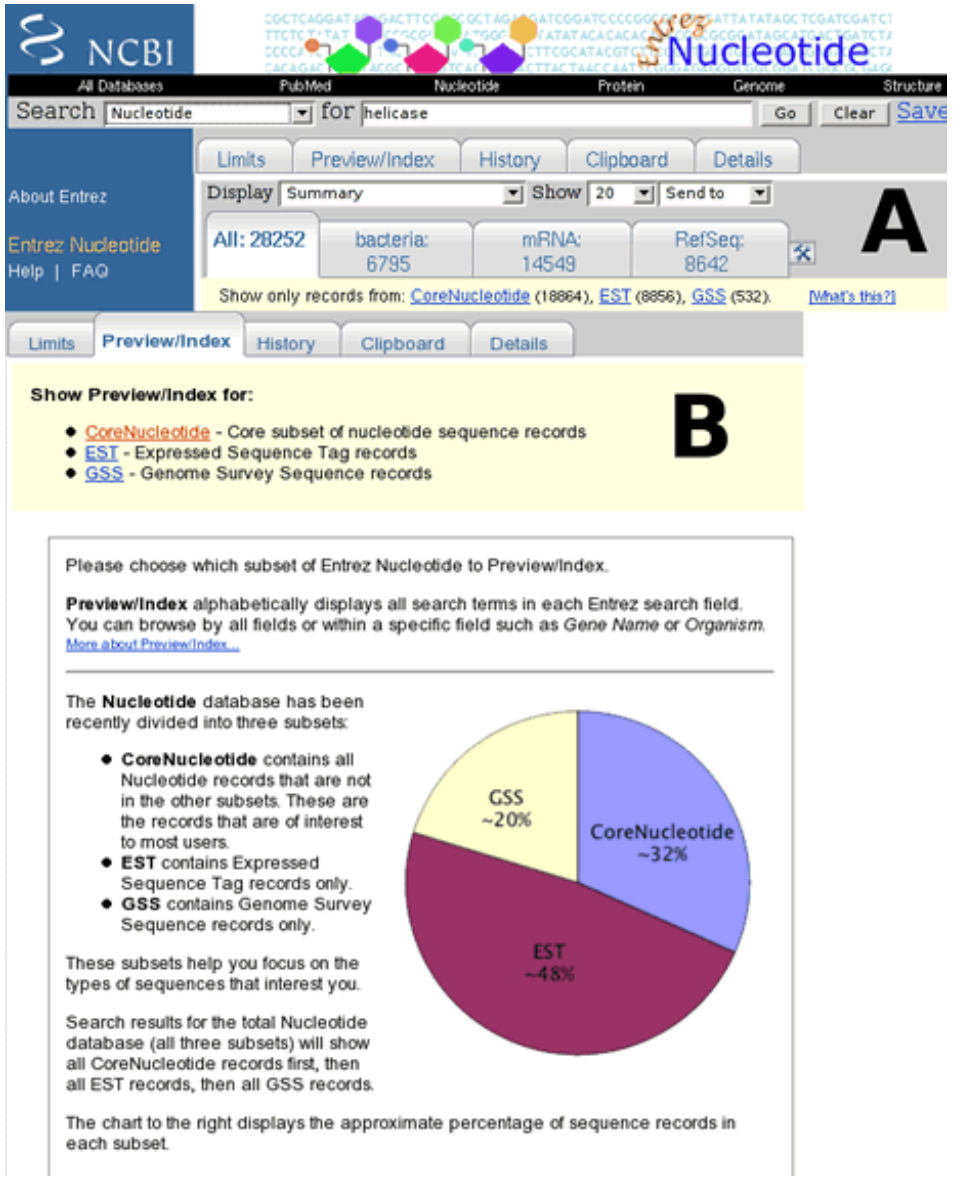
- Sequence
- Temperature optimization (higher temperatures, higher specificity).



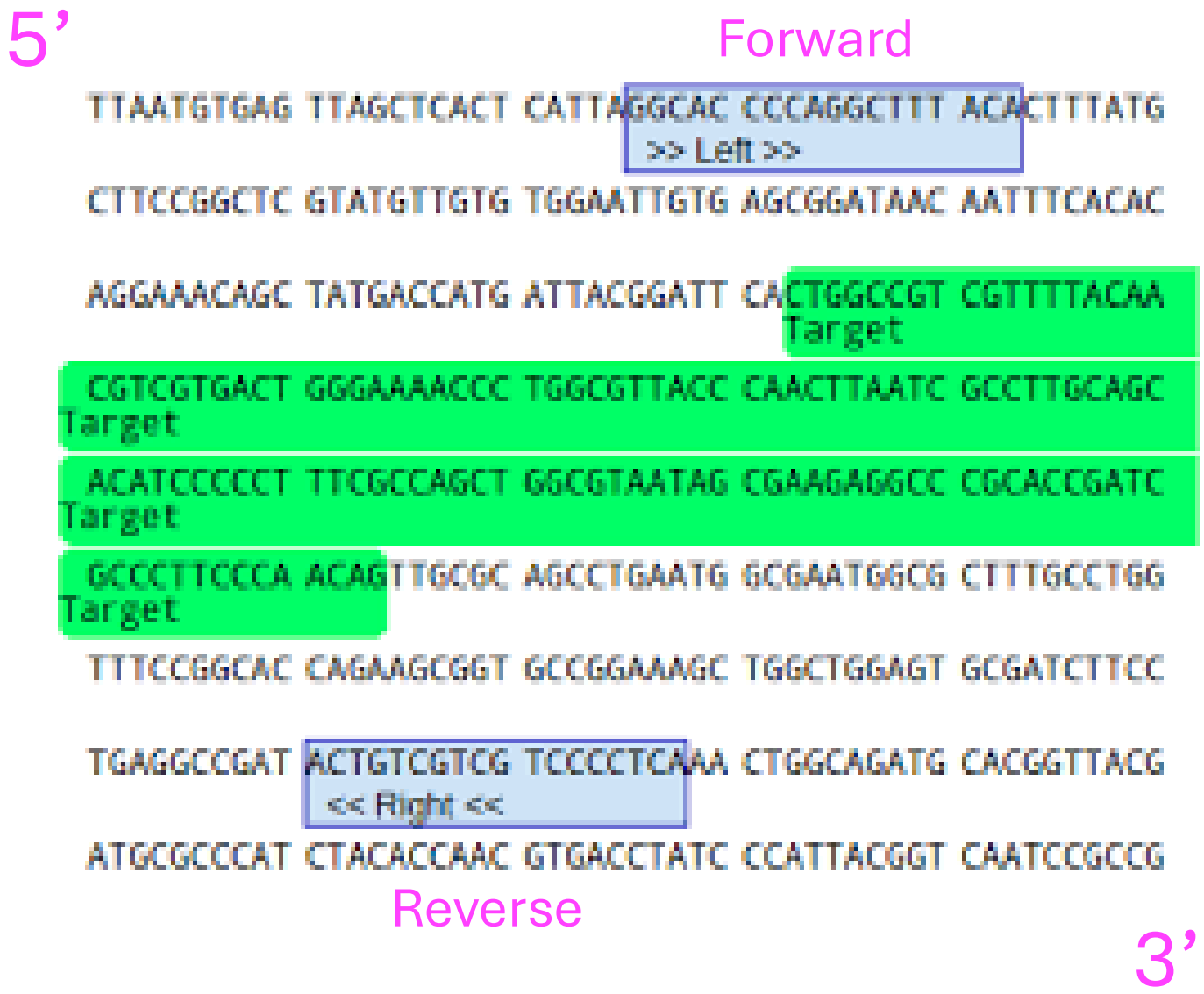
Primer design: specificity

Option 1 “EASY”: amplification of one given sequence. Eg. Virus laboratory strain.
Option 2 “LESS EASY”: amplification of one virus species, or multiple species eg. 16 sRNA

Primer design: retrieving target genome



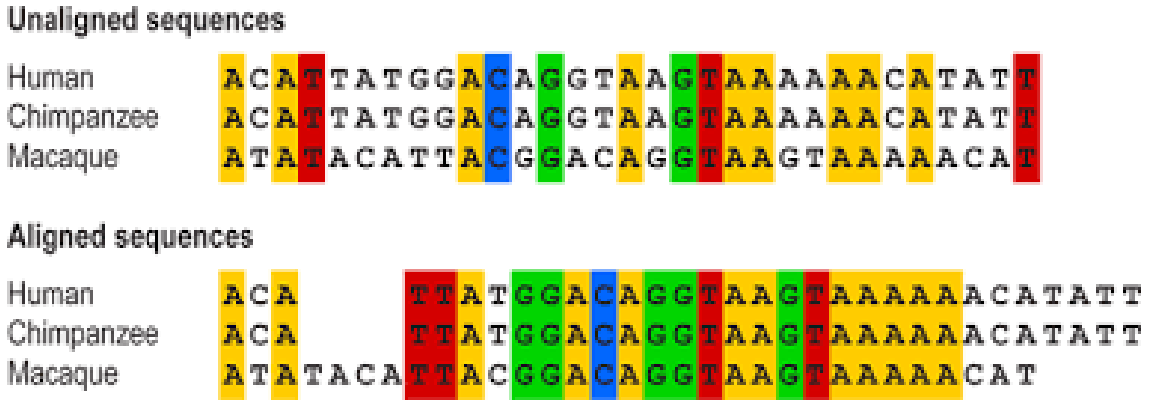
Primer design: primer selection



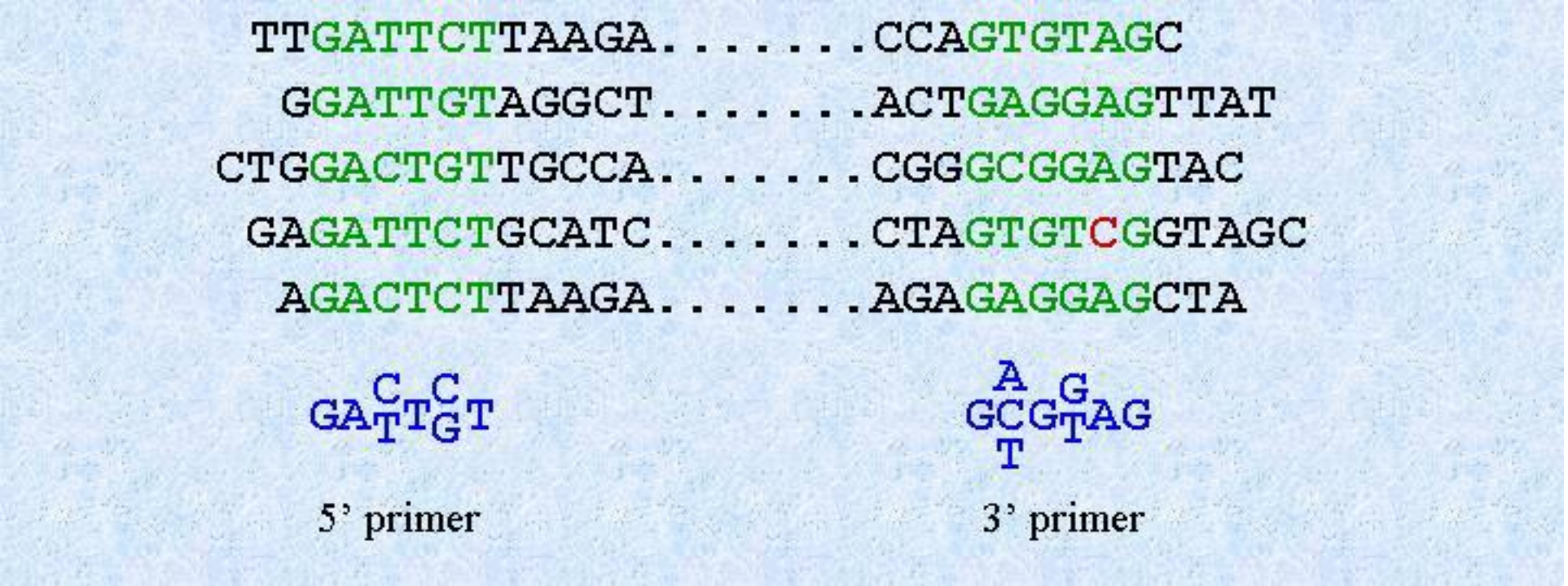
Then, go to **Blast** and search the primer sequences to validate if they are really specific.

Primer design: selection of degenerated primers

1. Download sequences of diverse species 2. Align them (eg. Geneious, Meta…)
3. Identify conserved regions



Primer design: selection of degenerated primers



Primer design and PCR optimization

1. Primer specificity
2. Length
3. Annealing and melting temperatures
4. GC composition
5. GC clamp
6. Secodary structure of the primer

Why primers are usually 20 nt long?

Primer lengths of 18-25 nucleotides strike a balance between specificity and efficiency. Shorter primers may not provide enough specificity, leading to non-specific binding, while longer primers can reduce the efficiency of the polymerase chain reaction (PCR) due to increased likelihood of secondary structures forming. The 18-25 nucleotide length allows for specific binding to the target sequence while still providing efficient amplification during PCR.

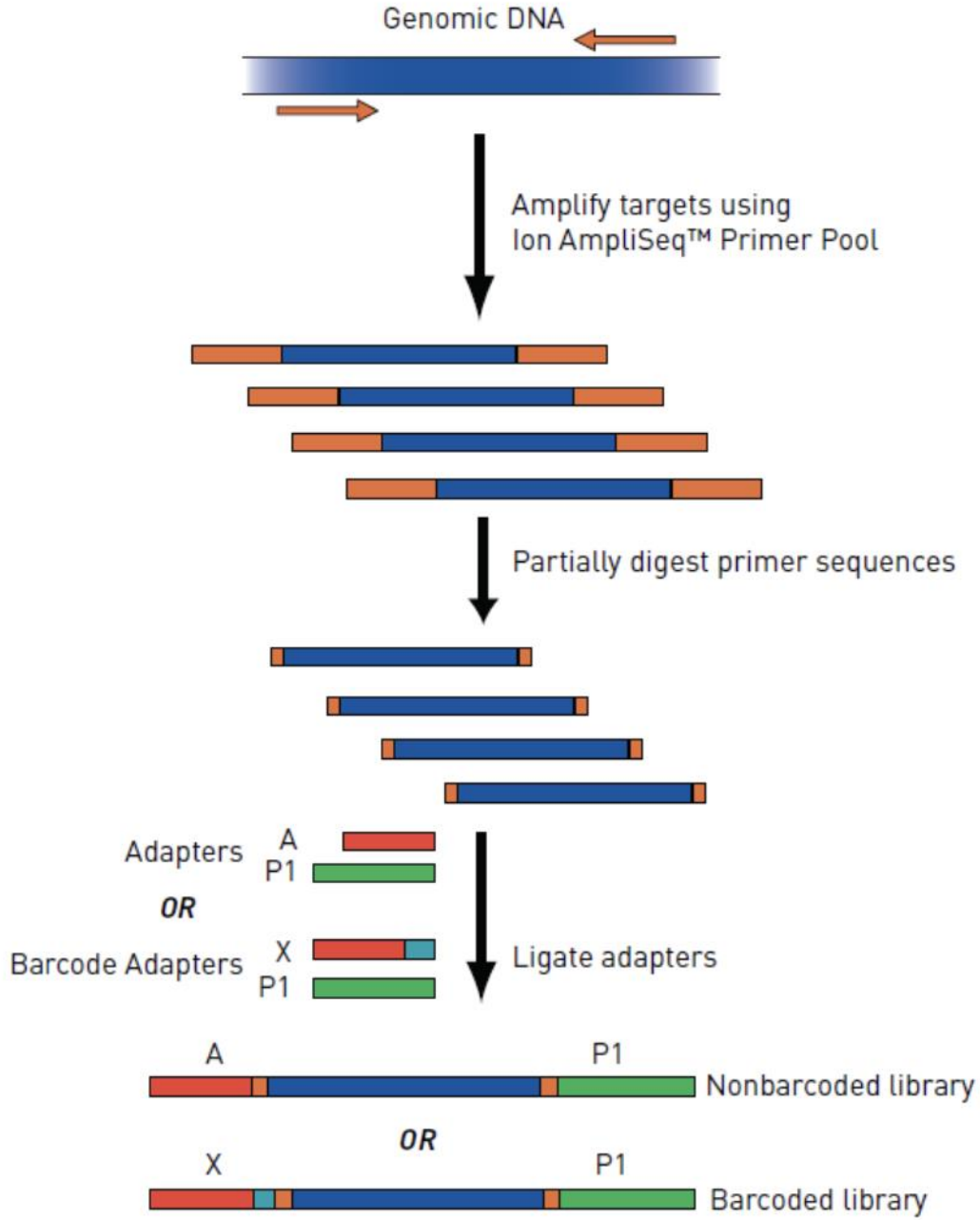
Primer secondary structures

Depending on the sequence of the primer there's a chance it could form **hairpins or dimerize** with itself or the other primer. This means it could base pair with itself or base pair with the other primer instead of the template.

DNA sequencing

Library preparation

- Quantify the concentration of each library using quantitative PCR (qPCR) or another suitable method.
- Normalize the concentration of each library to ensure equal representation in the sequencing pool.

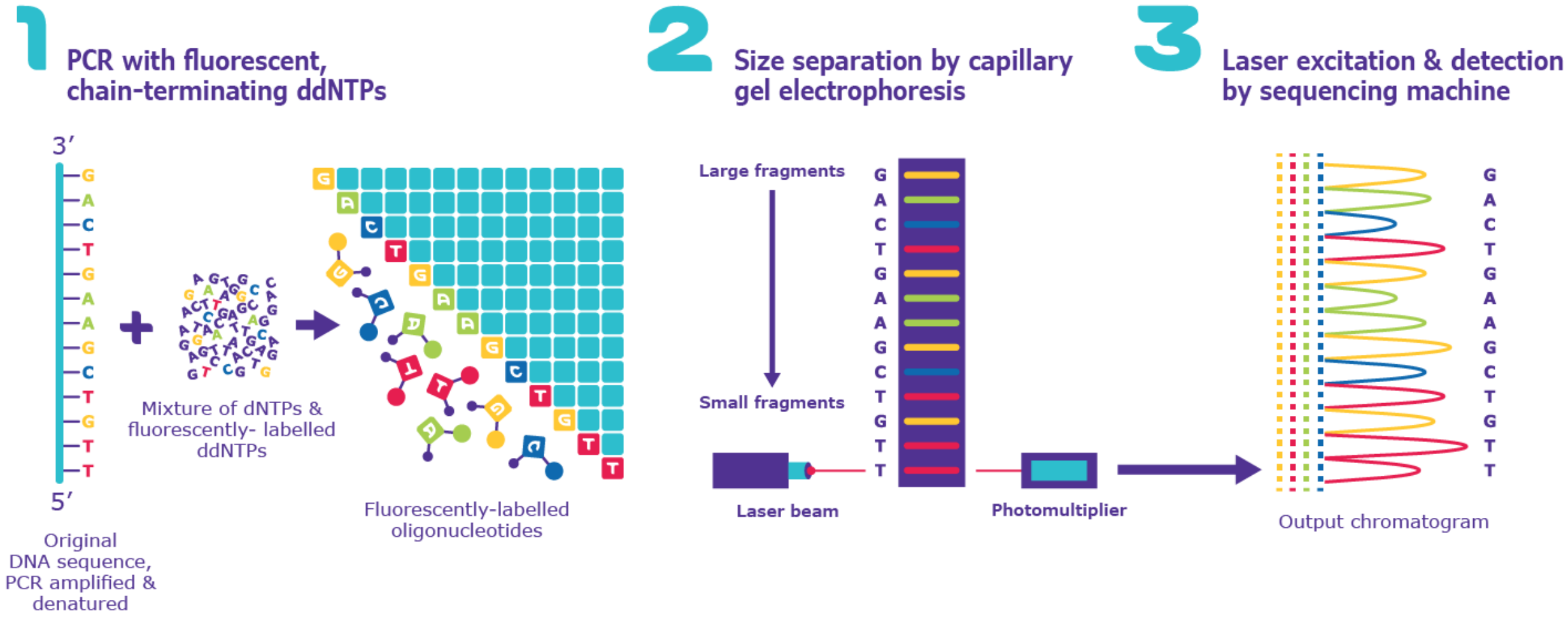


DNA Sequencing

- A gene's complete nucleotide sequence can be determined using a process called **DNA sequencing**
- The first automated procedure was based on a technique called dideoxy or chain termination sequencing, developed by Frederick Sanger
- In the last 15 years, "next-generation sequencing" techniques have been developed that are much faster

- In sequencing by synthesis, many DNA fragments are copied
- A specific strand of each fragment is immobilized, and the complementary strand synthesized one nucleotide at a time
- Thousands or hundreds of thousands of fragments about 300 nucleotides long can be sequenced in parallel
- This is an example of “high-throughput” technology

DNA Sequencing: Sanger sequencing



DNA Sequencing: next generation sequencing

First Generation

Sanger Sequencing
Maxam and Gilbert
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

Second Generation
(Next Generation Sequencing)

454, Solexa,
Ion Torrent
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

Third Generation

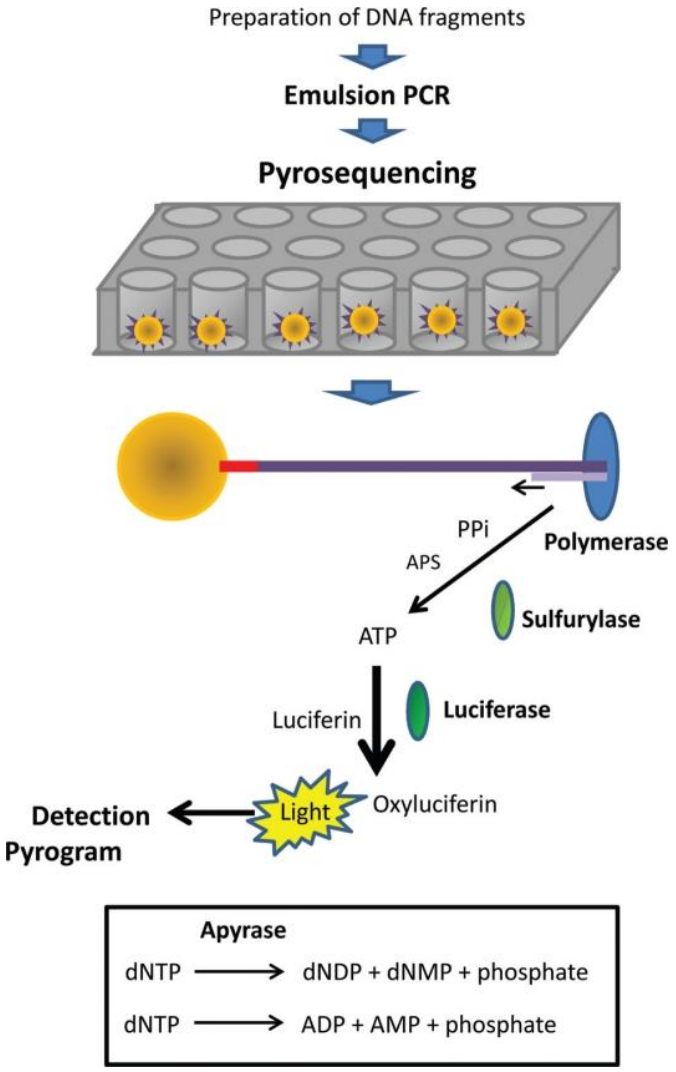
PacBio
Oxford Nanopore

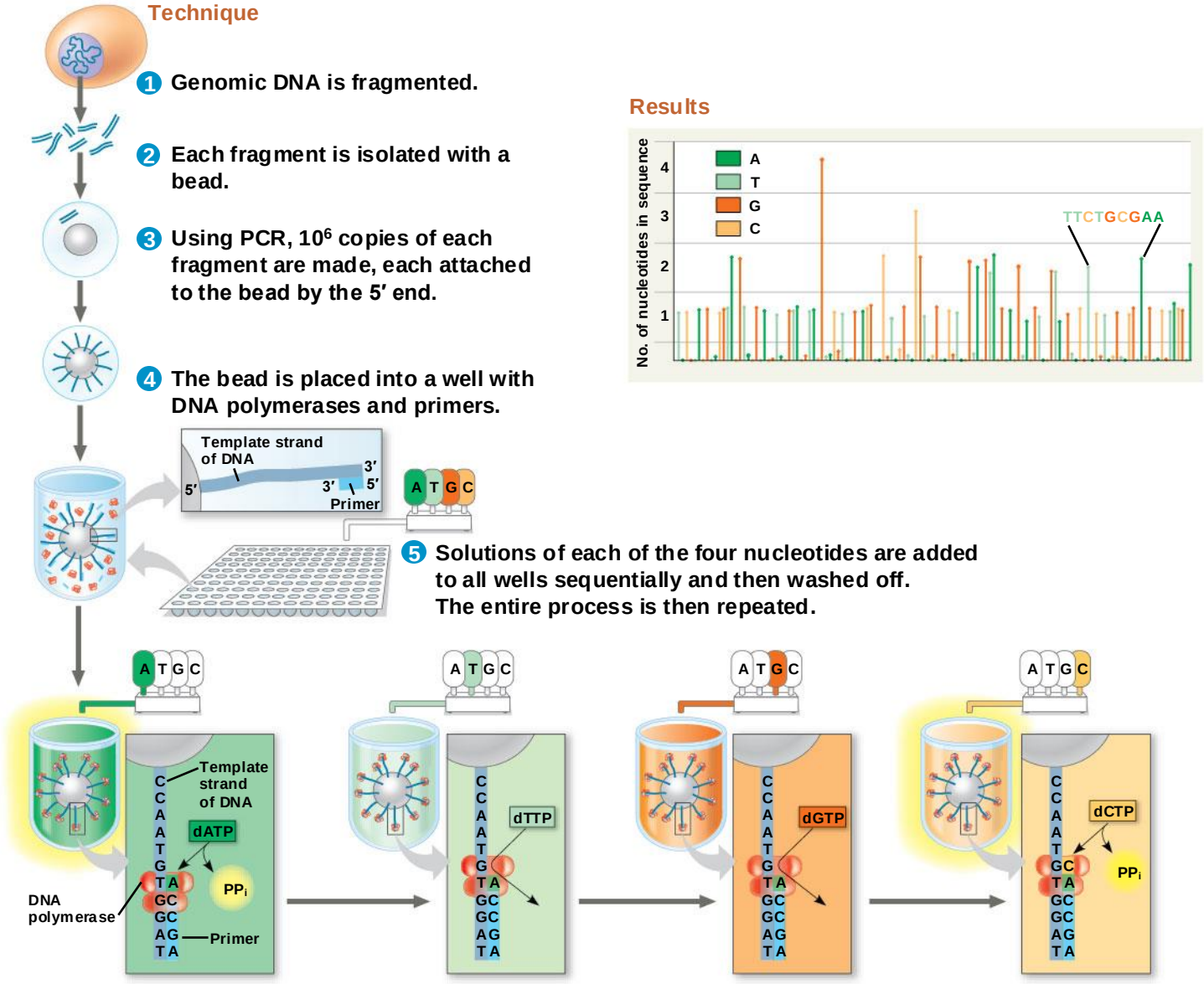
- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

Short-read sequencing

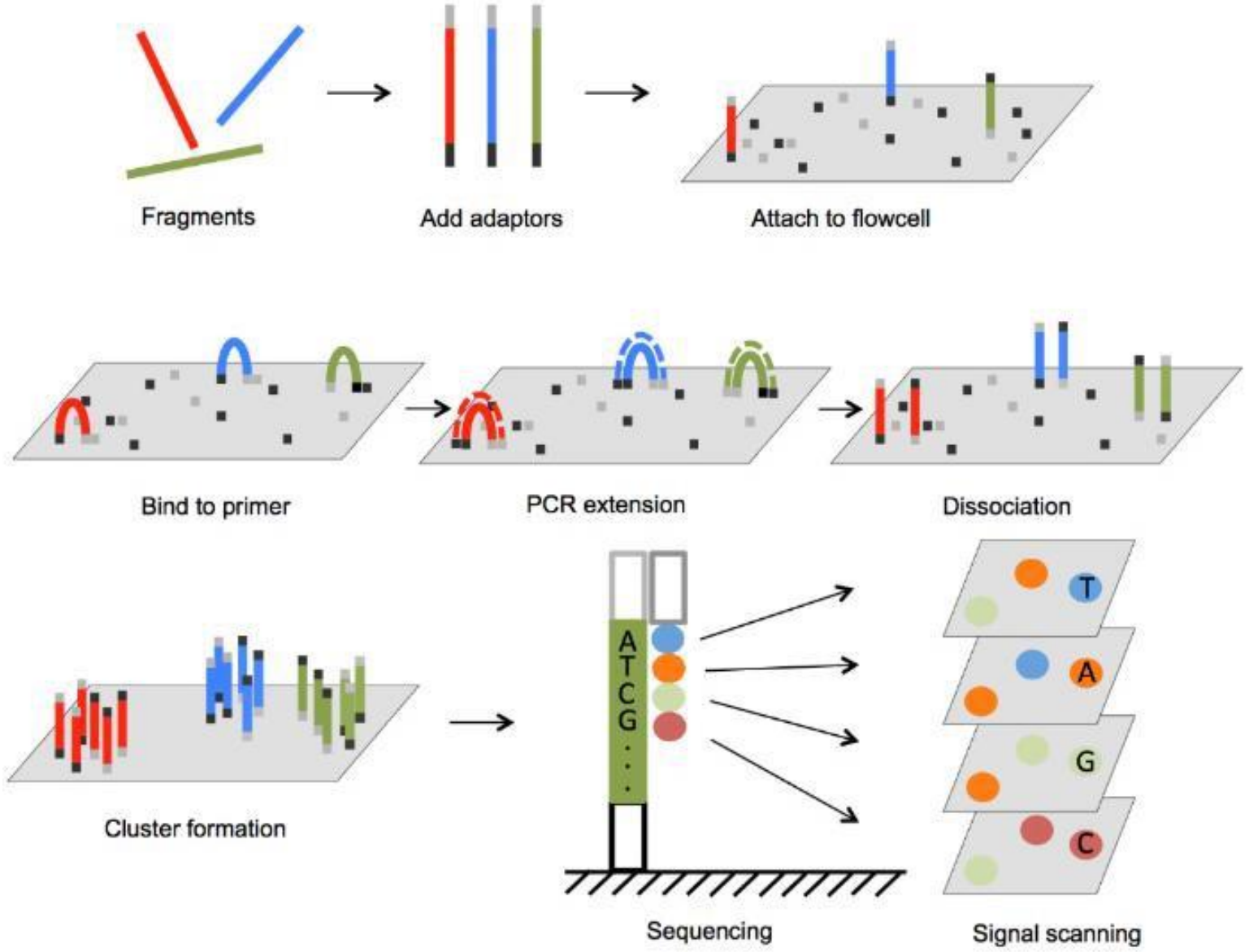
Long-read sequencing

Sequencing by synthesis

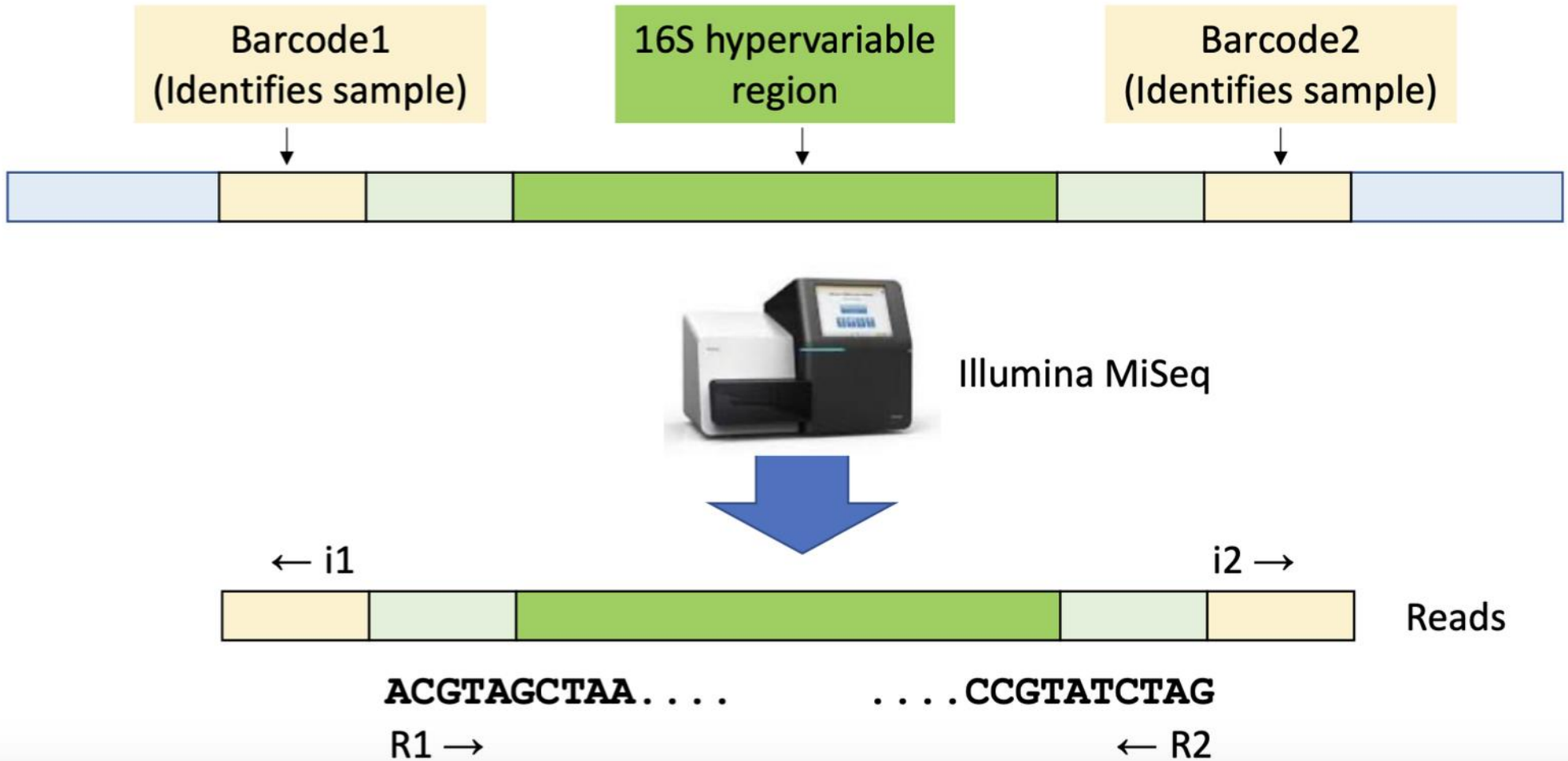




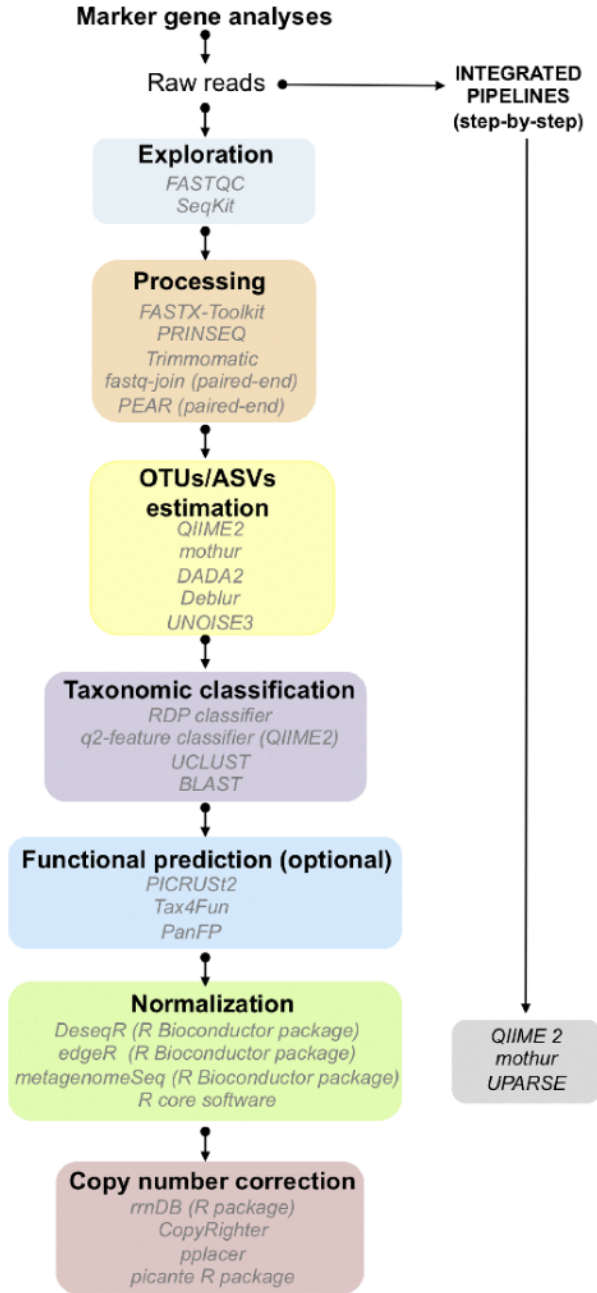
Illumina sequencing



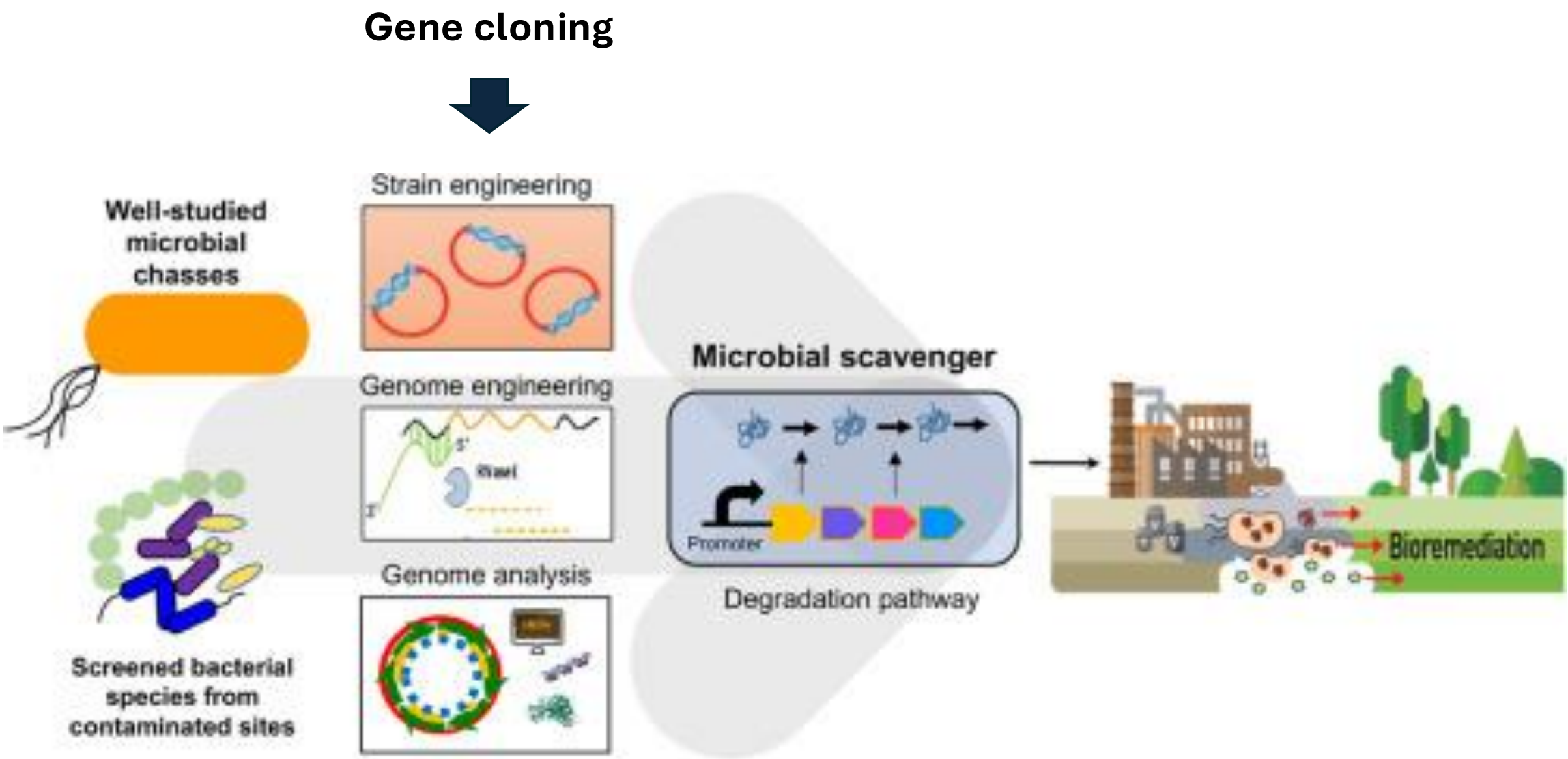
Raw reads


Barcode1
(Identifies sample)
16S hypervariable
region
Barcode2
(Identifies sample)
Illumina MiSeq
← i1
i2 →
Reads
ACGTAGCTAA. . . .
. . . .CCGTATCTAG
R1 →
← R2

Amplicon sequencing analysis



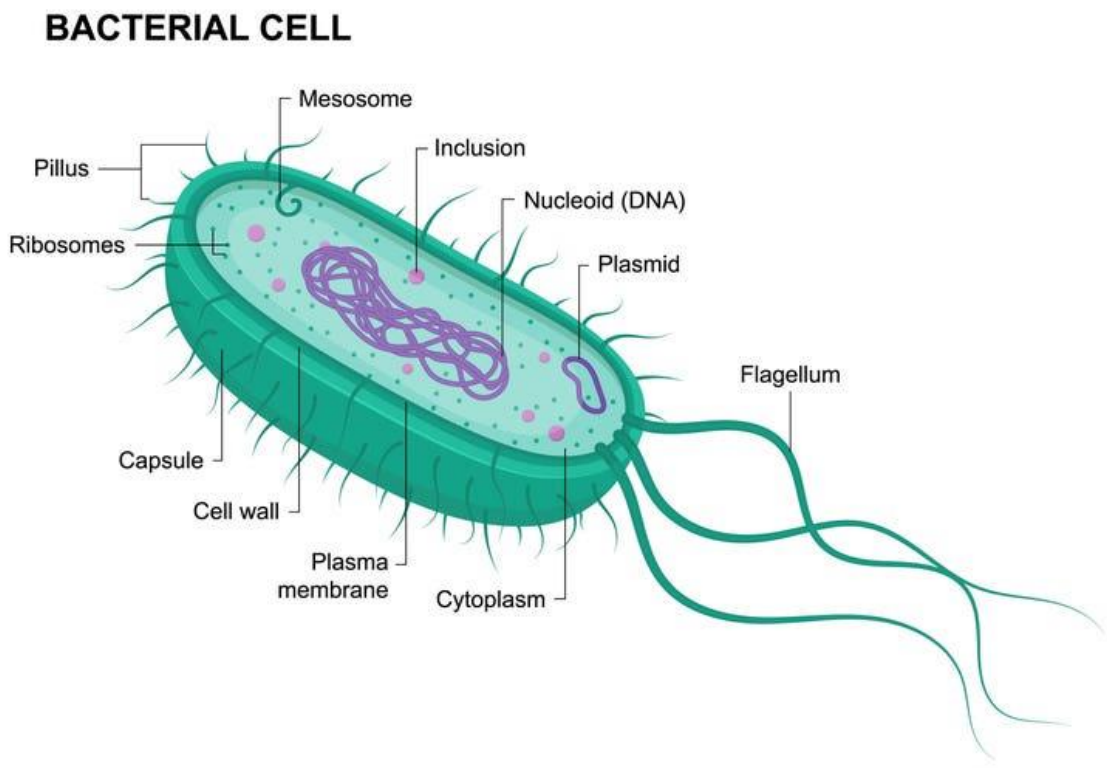
DNA cloning



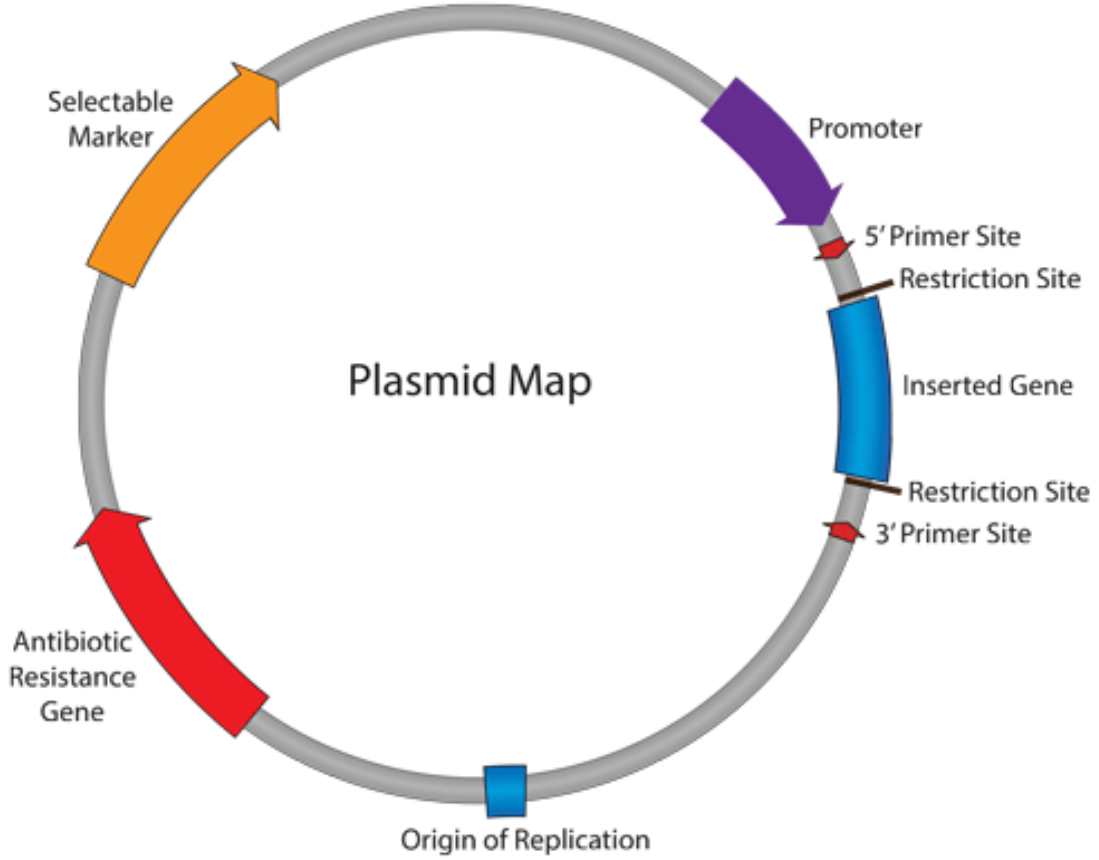
Tran et al. 2021

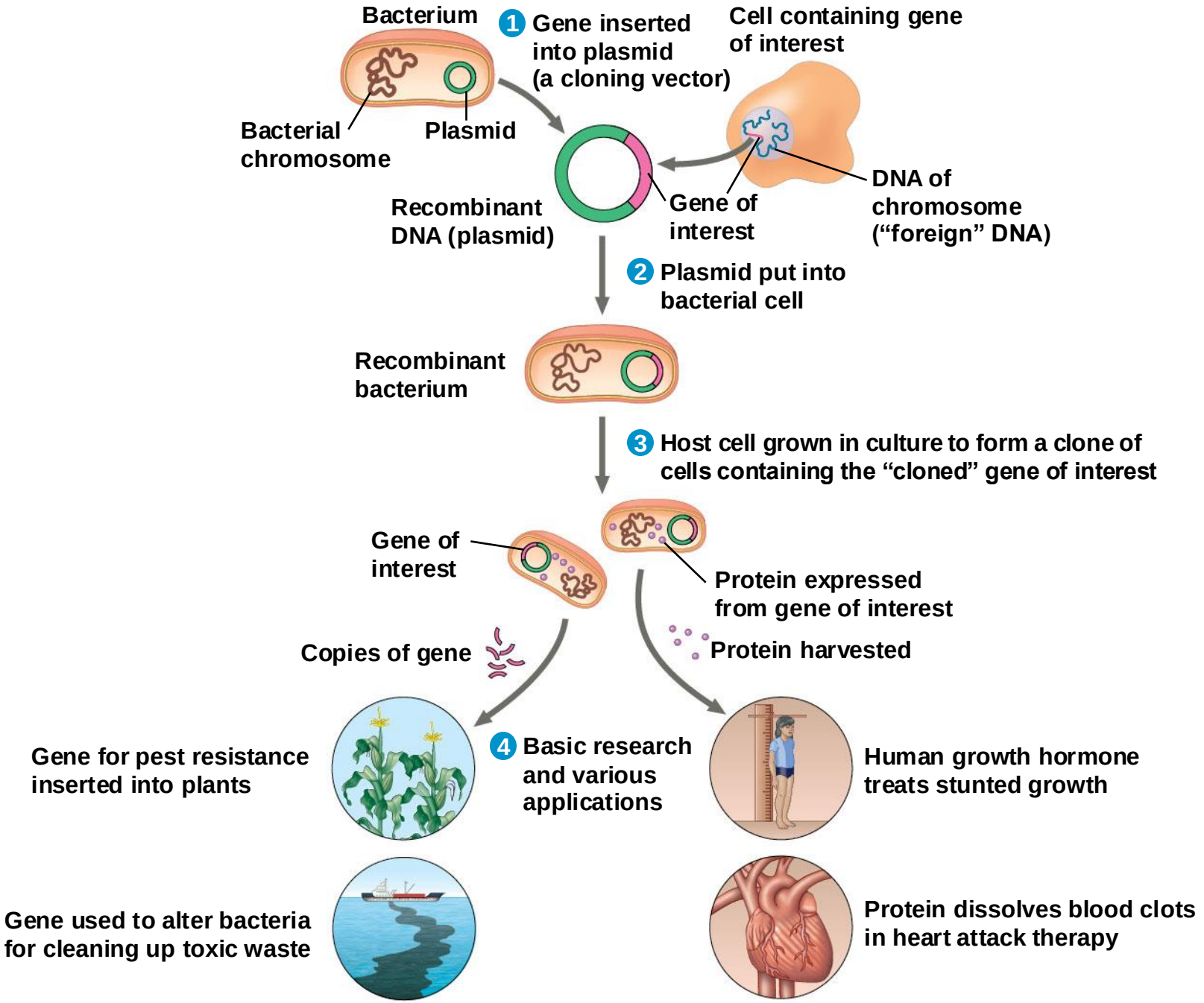
Making Multiple Copies of a Gene or Other DNA Segment

- To work directly with specific genes, scientists prepare well-defined DNA segments in multiple identical copies by a process called **DNA cloning**
- **Plasmids** are small, circular DNA molecules that replicate separately from the bacterial chromosome
- Researchers can insert DNA into a plasmid to produce a **recombinant DNA molecule**, which contains DNA from two different sources



- Reproduction of a recombinant plasmid in a bacterial cell results in cloning of the plasmid including the foreign DNA
- This production of multiple copies of a single gene is a type of DNA cloning called **gene cloning**

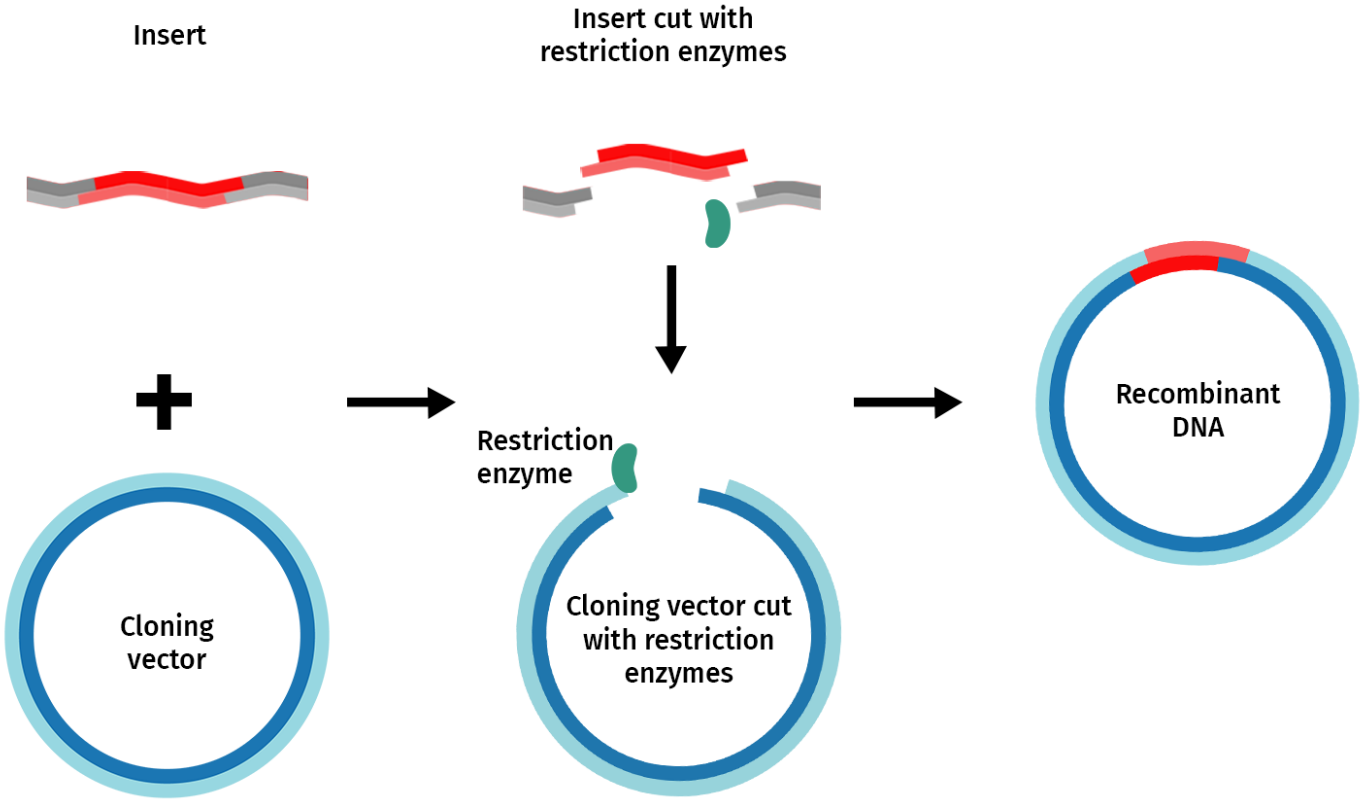



Bacterium
1 Gene inserted into plasmid (a cloning vector)
Cell containing gene of interest
Bacterial chromosome
Plasmid
Recombinant DNA (plasmid)
Gene of interest
DNA of chromosome ("foreign" DNA)
2 Plasmid put into bacterial cell
Recombinant bacterium
3 Host cell grown in culture to form a clone of cells containing the "cloned" gene of interest
Gene of interest
Protein expressed from gene of interest
Copies of gene
Protein harvested
Gene for pest resistance inserted into plants
4 Basic research and various applications
Human growth hormone treats stunted growth
Gene used to alter bacteria for cleaning up toxic waste
Protein dissolves blood clots in heart attack therapy

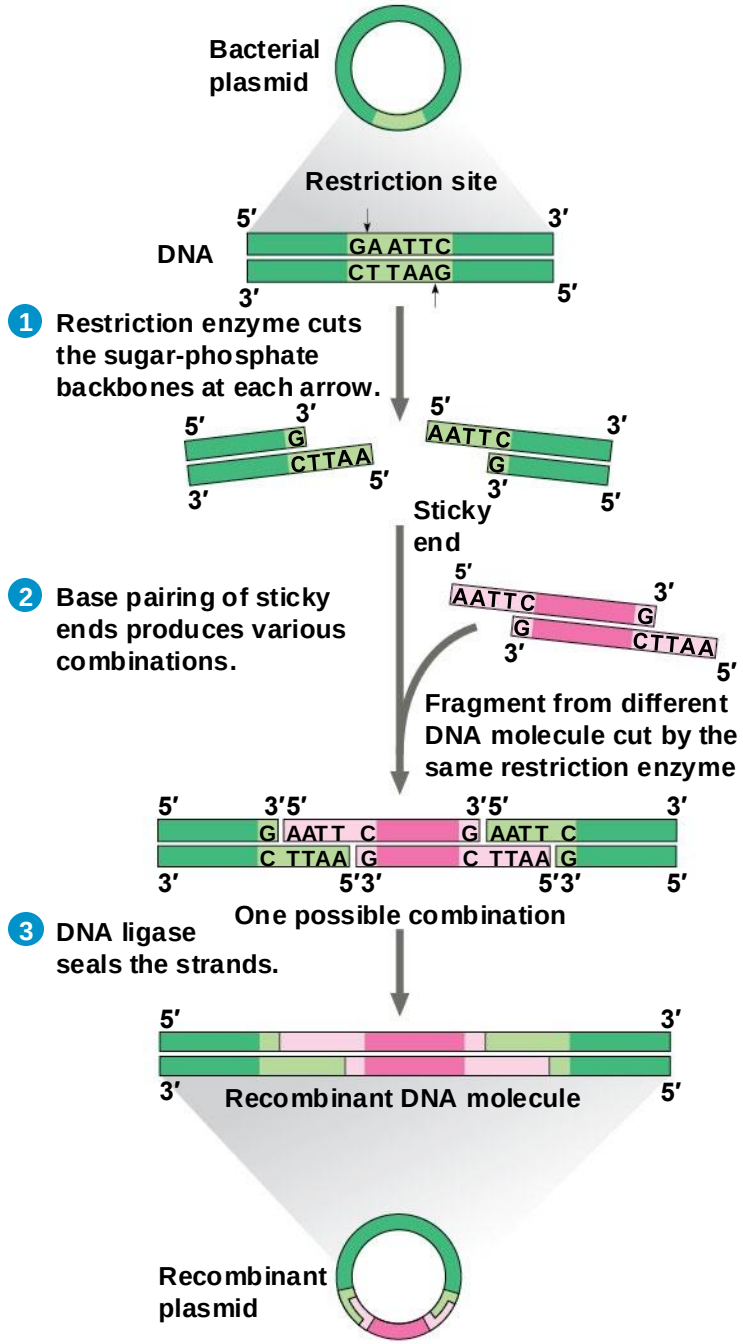
- A plasmid used to clone a foreign gene is called a **cloning vector**
- Bacterial plasmids are widely used as cloning vectors because they are readily obtained, easily manipulated, and easily introduced into bacterial cells, and once in the bacteria they multiply rapidly
- Gene cloning is useful for amplifying genes to produce a protein product for research, medical, or other purposes

Scissors and glue for gene cloning: **restriction enzymes**

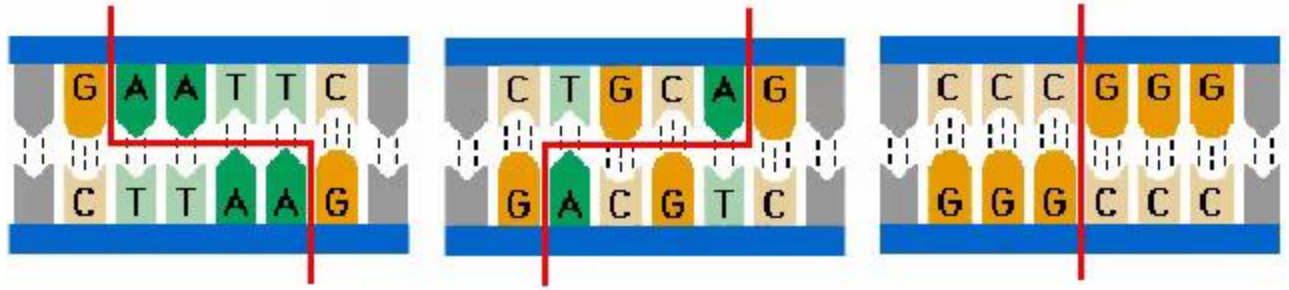
- Bacterial **restriction enzymes** cut DNA molecules at specific DNA sequences called **restriction sites**
- A restriction enzyme usually makes many cuts, yielding **restriction fragments**
- The most useful restriction enzymes cut DNA in a staggered way, producing fragments with at least one single-stranded end called a **sticky end**

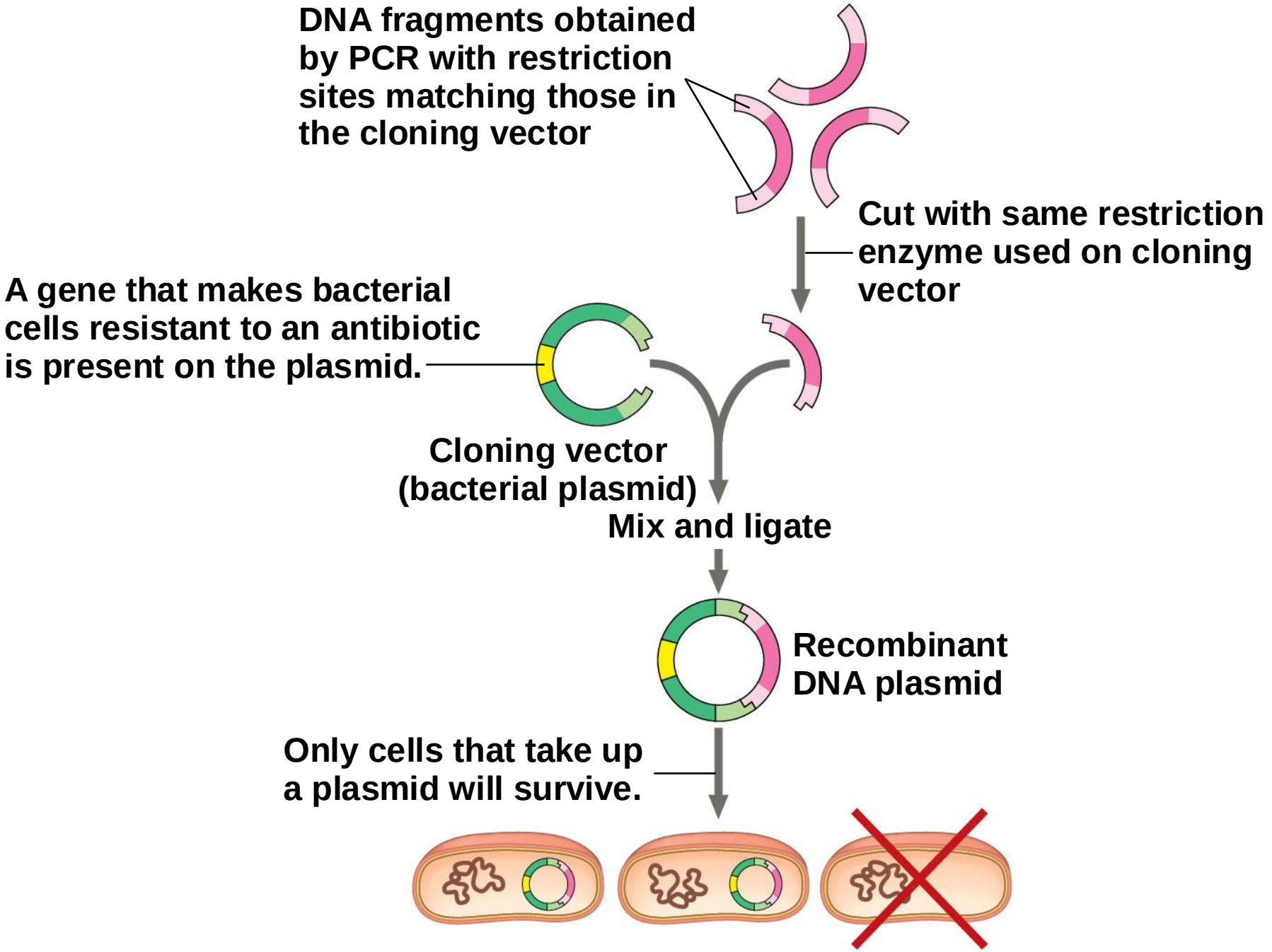


- Sticky ends can bond with complementary sticky ends of other fragments
- **DNA ligase** is an enzyme that seals the bonds between restriction fragments
- This allows researchers to join two DNA fragments from different sources

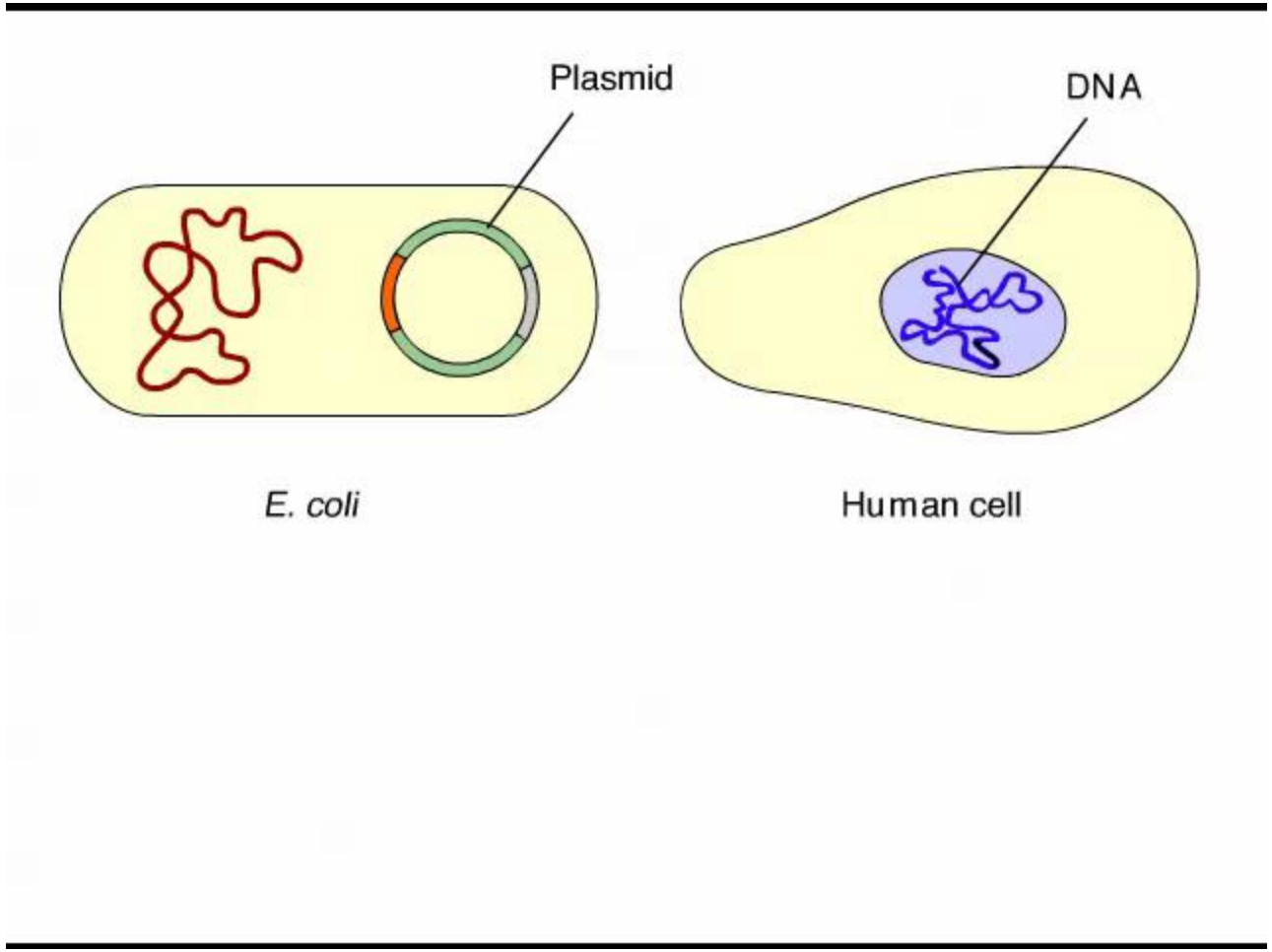

Bacterial plasmid
Restriction site
DNA
5′
3′
GAATTC
CTTAAG
3′
5′
1
Restriction enzyme cuts the sugar-phosphate backbones at each arrow.
5′
3′
G
CTTAA
5′
3′
5′
AATTC
3′
G
3′
5′
Sticky end
2
Base pairing of sticky ends produces various combinations.
5′
AATTC
3′
G
G
CTTAA
3′
5′
Fragment from different DNA molecule cut by the same restriction enzyme
5′
3′5′
3′5′
3′
G
AATT C
G
AATT C
C TTAA
G
C TTAA
G
3′
5′3′
5′3′
5′
One possible combination
3
DNA ligase seals the strands.
5′
3′
3′
Recombinant DNA molecule
5′
Recombinant plasmid

Animation: Restriction Enzymes



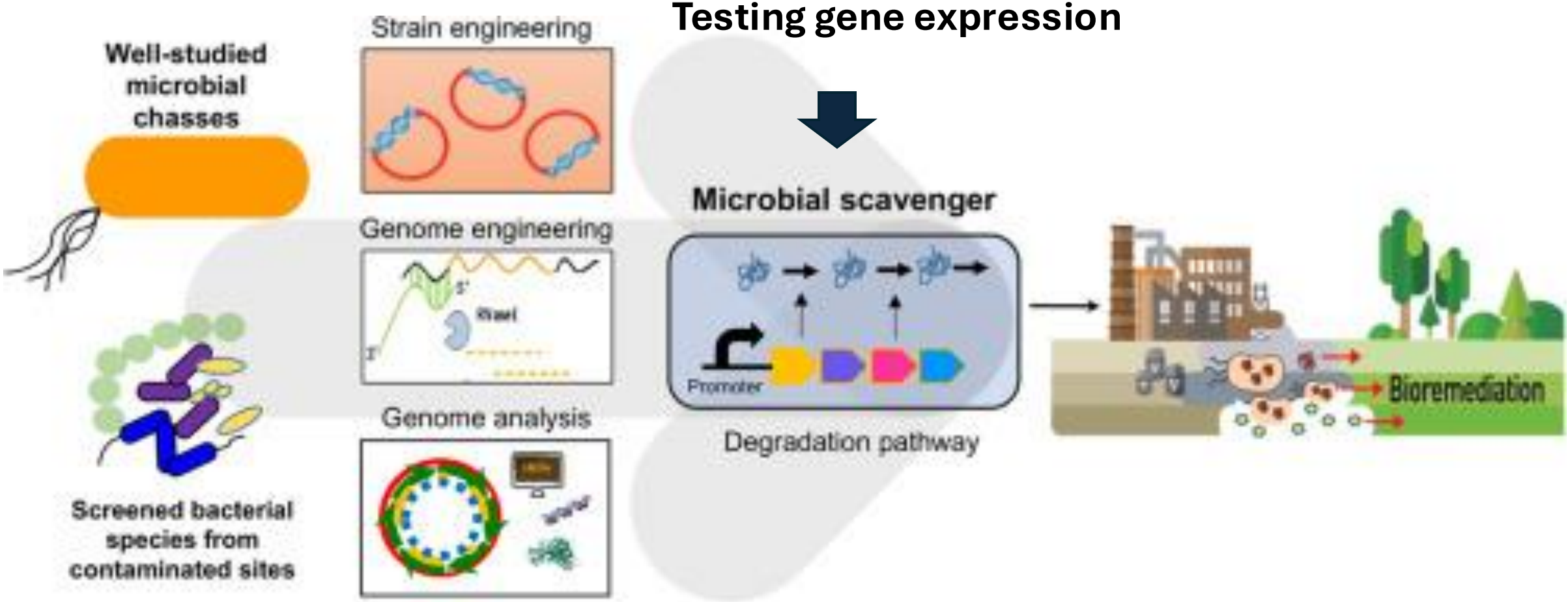

DNA fragments obtained by PCR with restriction sites matching those in the cloning vector
Cut with same restriction enzyme used on cloning vector
A gene that makes bacterial cells resistant to an antibiotic is present on the plasmid.
Cloning vector (bacterial plasmid)
Mix and ligate
Recombinant DNA plasmid
Only cells that take up a plasmid will survive.

Animation: Cloning a Gene



Gene expression and editing

Testing gene expression

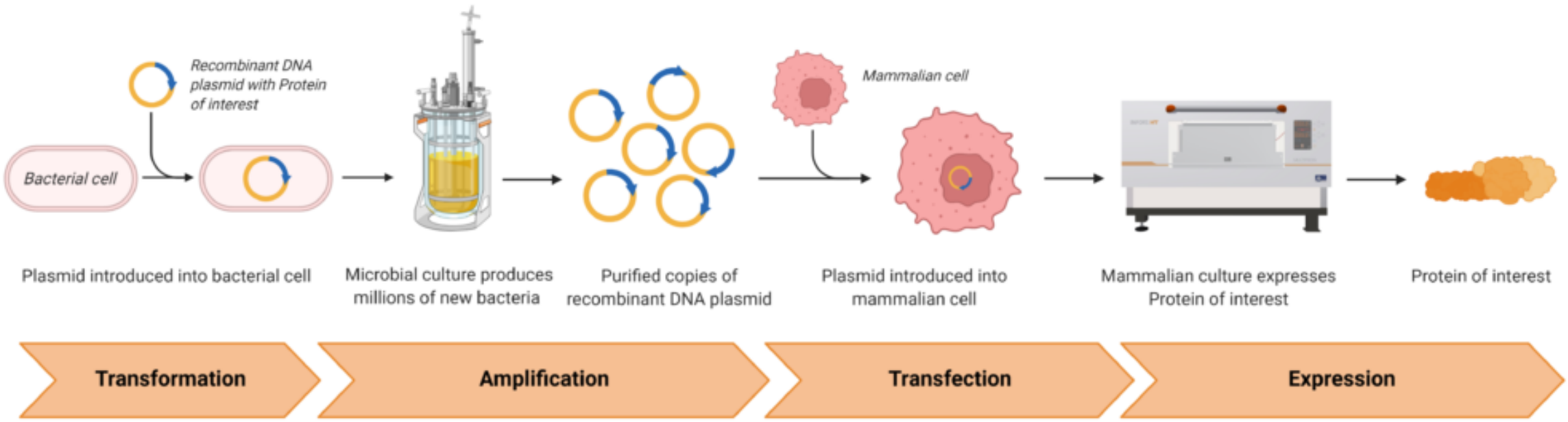


Tran et al. 2021

Expressing Cloned Genes

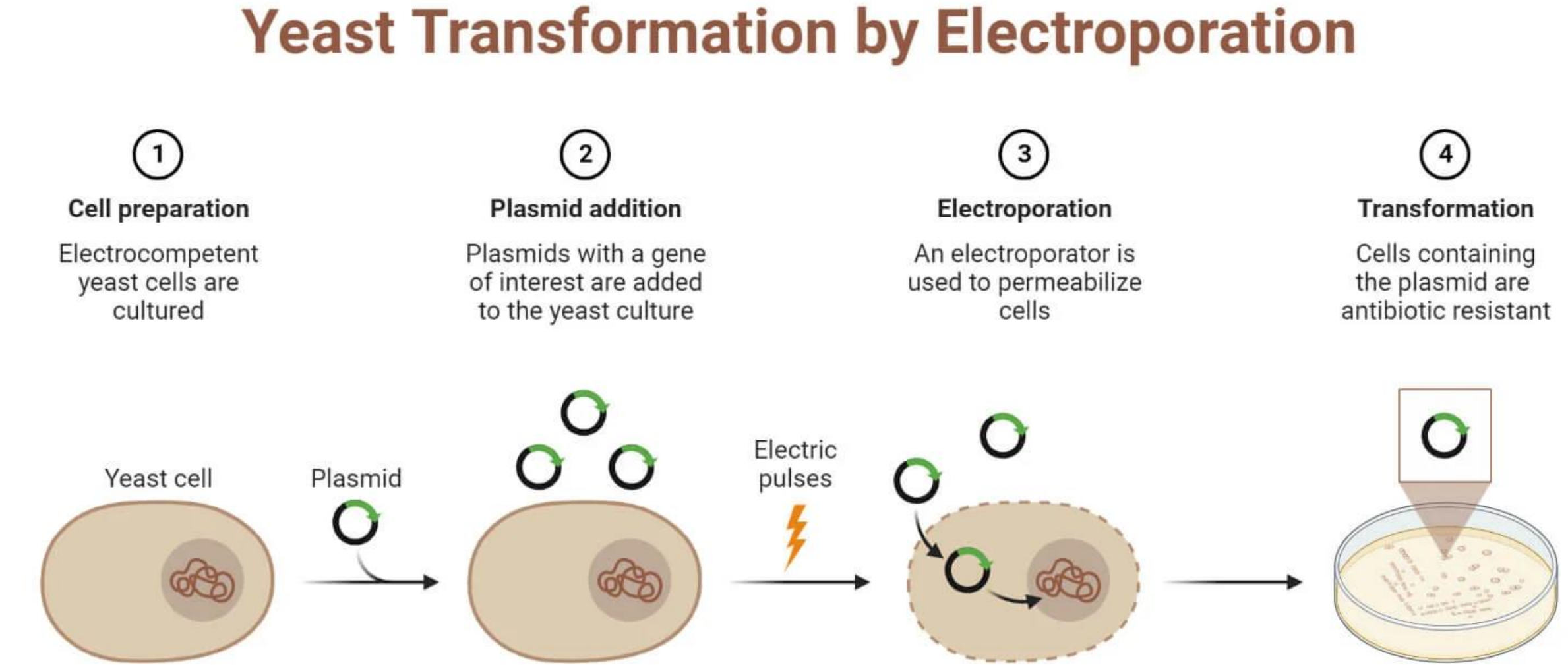
- After a gene has been cloned, its protein product can be produced in larger amounts for research or practical applications
- Cloned genes can be expressed in either bacterial or eukaryotic cells
- Several technical difficulties hinder expression of cloned eukaryotic genes in bacterial host cells
- To overcome differences in promoters and other DNA control sequences, scientists usually employ an **expression vector**, a cloning vector that contains a highly active bacterial promoter
- Molecular biologists can avoid eukaryote-bacterial incompatibility issues by using eukaryotic cells, such as yeasts, as hosts for cloning and expressing genes

- Another problem with eukaryotic gene expression in bacteria is the presence of introns in most eukaryotic genes
- Researchers can avoid this problem by using **cDNA** (short for copy DNA or complementary DNA) is synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase), which contains only exons.



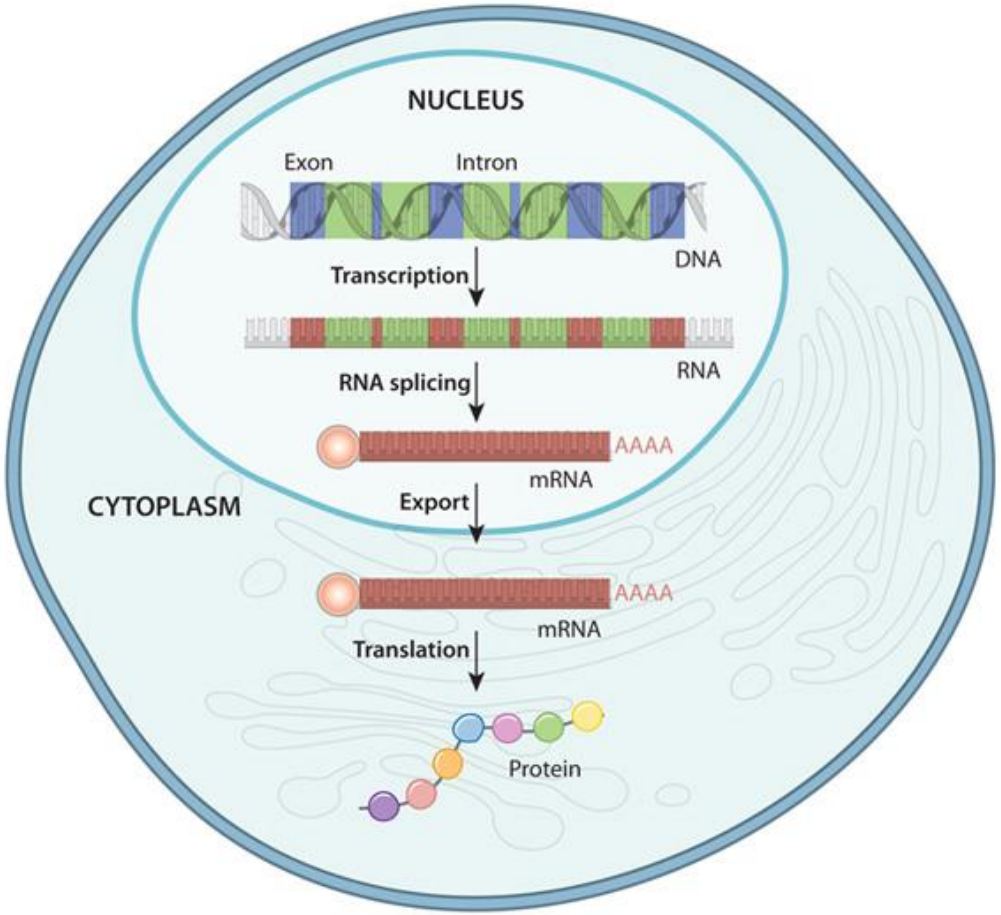
- One method of introducing recombinant DNA into eukaryotic cells is **electroporation**, applying a brief electrical pulse to create temporary holes in plasma membranes
- Alternatively, scientists can inject DNA into cells using microscopically thin needles or viral vectors
- Once inside the cell, the DNA is incorporated into the cell's DNA by natural genetic recombination

Example of transformation



Expressing Cloned Genes

- Analysis of when and where a gene or group of genes is expressed can provide important clues about gene function
- The most straightforward way to discover which genes are expressed in certain cells is to identify the **mRNAs** being made

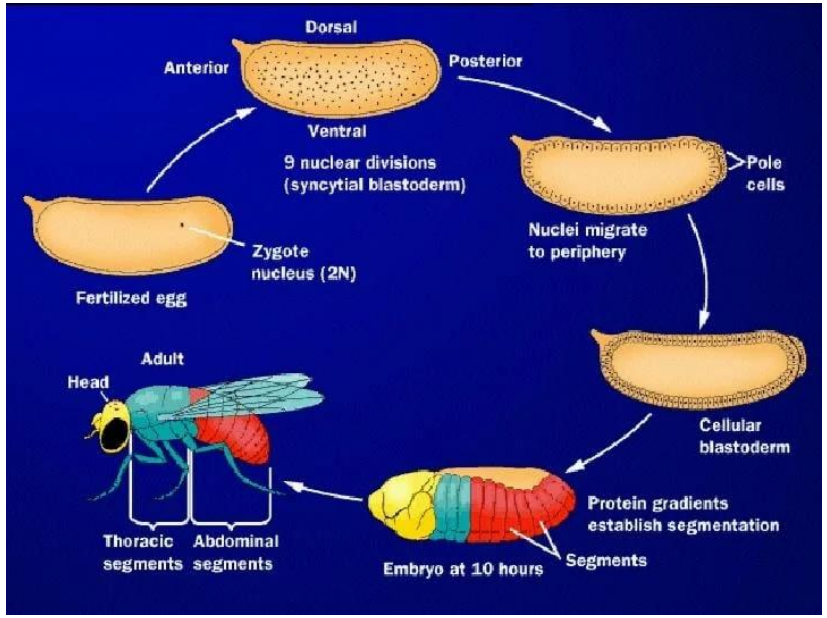


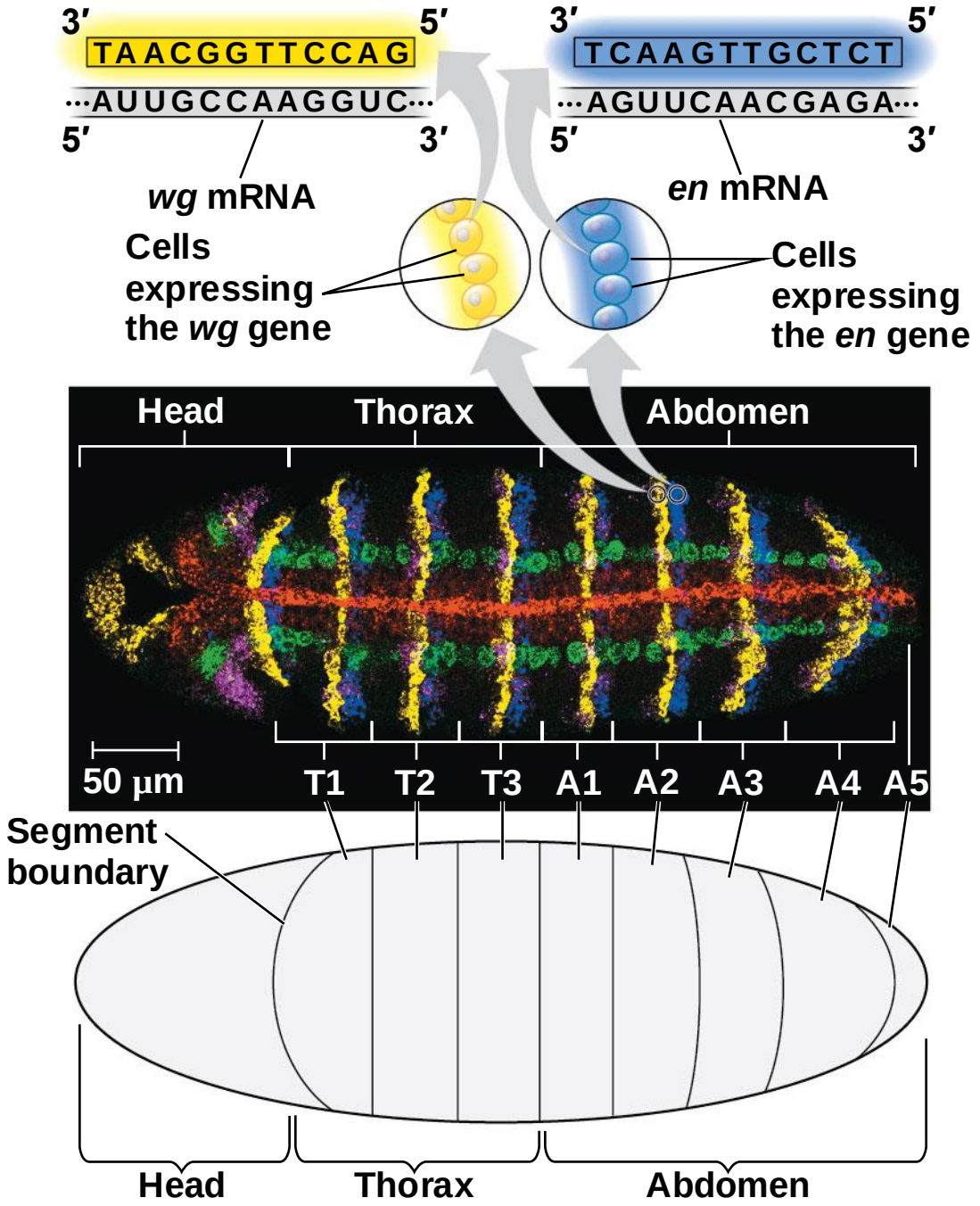
Studying the Expression of Single Genes

- mRNA can be detected by nucleic acid hybridization with complementary molecules
- These complementary molecules, of either DNA or RNA, are **nucleic acid probes**
- ***In situ* hybridization** uses fluorescent dyes attached to probes to identify the spatial location of specific mRNAs in place in the intact organism

Example of FISH

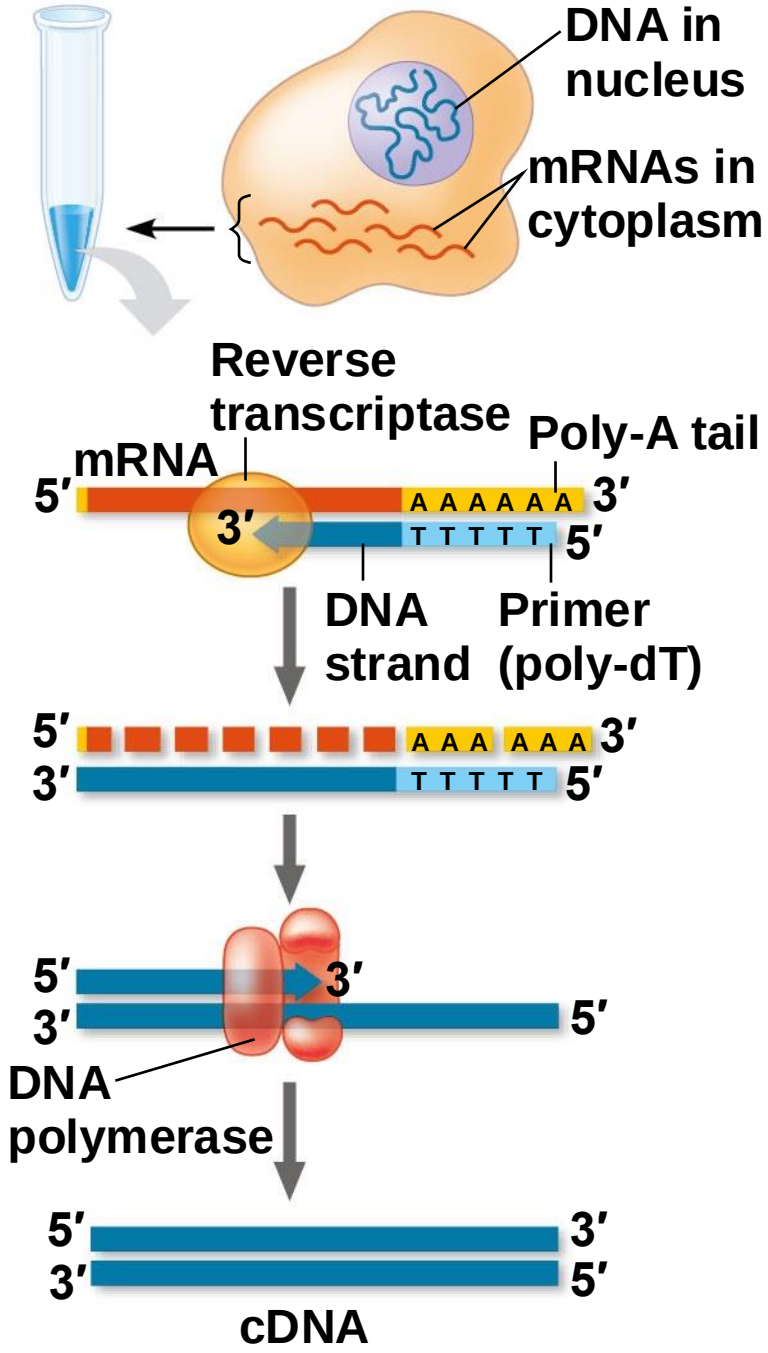
Determining where single genes are expressed by fluorescence *in situ* hybridization analysis (FISH)





- **Reverse transcriptase-polymerase chain reaction (RT-PCR)** is useful for comparing amounts of specific mRNAs in several samples at the same time
- Reverse transcriptase is added to mRNA to make **complementary DNA (cDNA)**, which serves as a template for PCR amplification of the gene of interest
- The products are run on a gel, and the complement of the mRNA of interest is identified

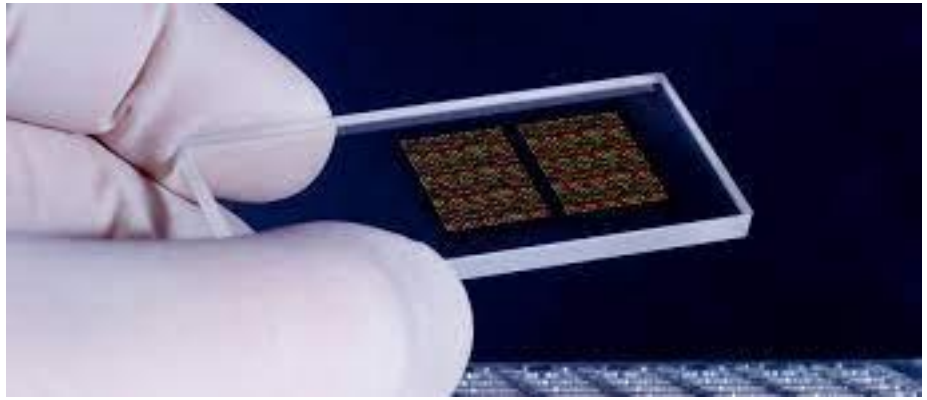
Making complementary DNA (cDNA) from eukaryotic genes

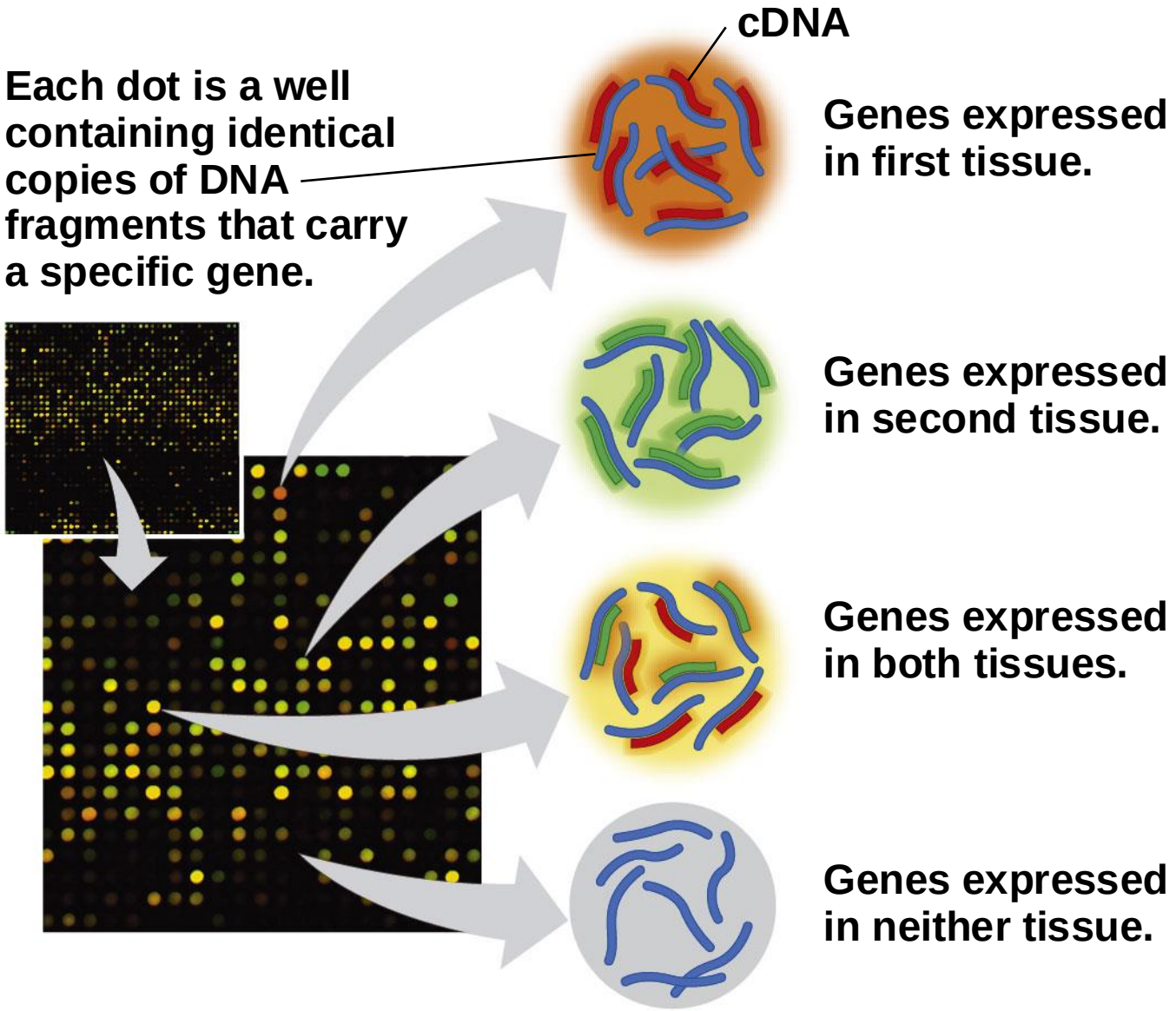


Studying the Expression of Interacting Groups of Genes

- The study of expression of thousands of genes at one time constitutes a systems approach
- One aim is to identify networks of gene expression across an entire genome
- **DNA microarray assays** compare patterns of gene expression in different tissues at different times or under different conditions
- A microarray consists of tiny amounts of many single-stranded genes fixed to a glass slide

Microarrays



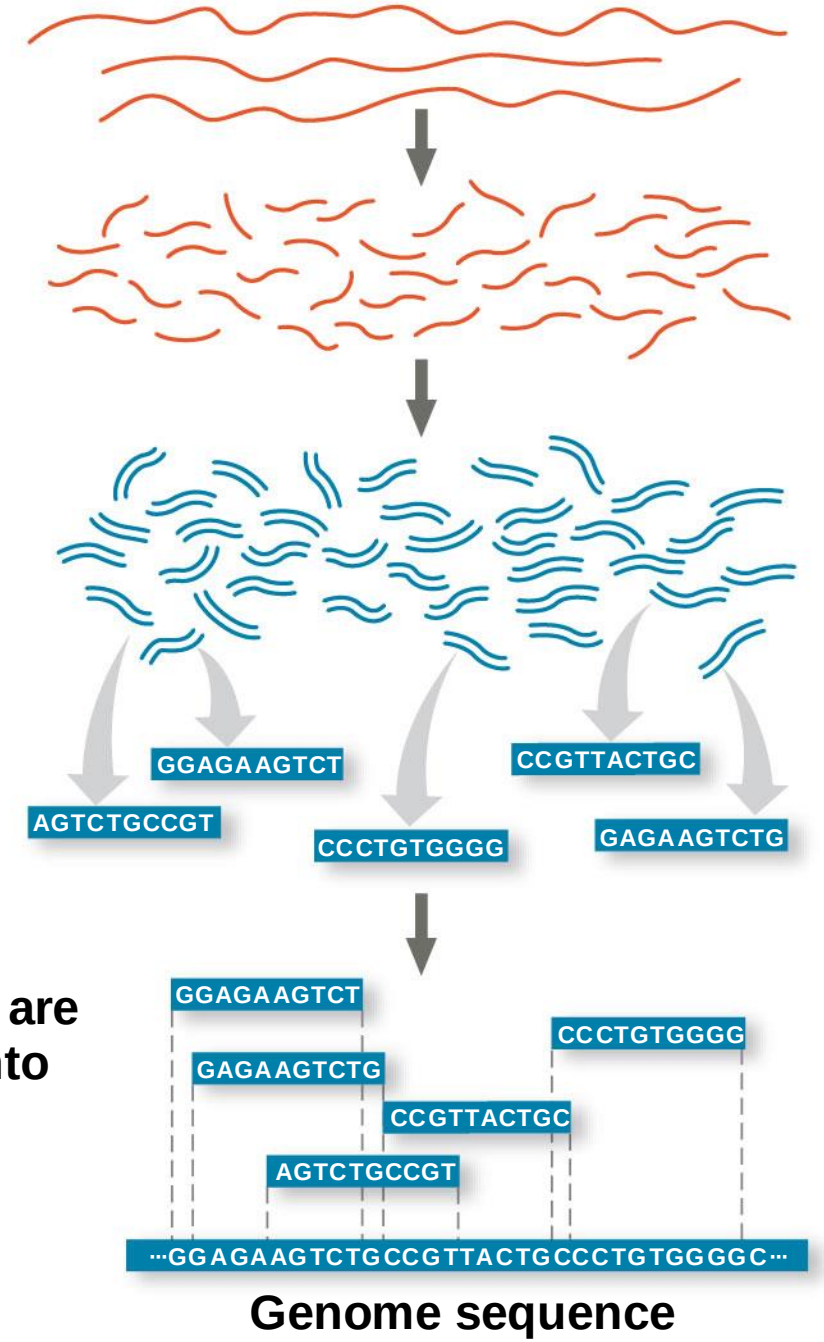


RNA sequencing

- With rapid and inexpensive sequencing methods, researchers can now just sequence cDNA samples from different tissues or embryonic stages to determine the gene expression differences between them
- This approach is called **RNA sequencing** or **RNA-seq**
- RNAs are isolated, cut into short, similar-sized fragments, converted into cDNAs, and sequenced

RNAseq

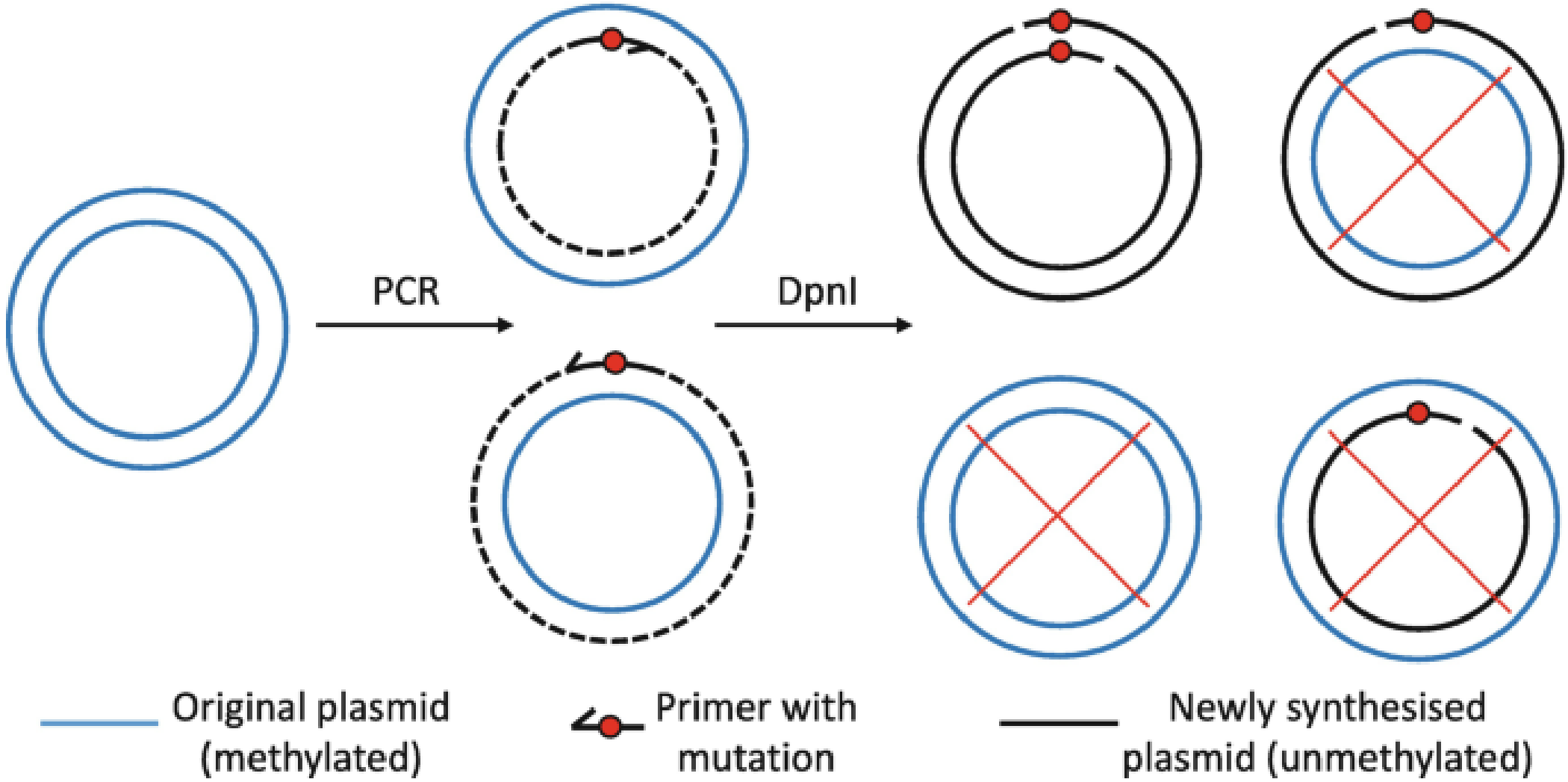
1. mRNAs are isolated from the tissue being studied.
2. mRNAs are cut into similar-sized, small fragments.
3. mRNAs are reverse-transcribed into cDNAs of the same size.
4. cDNAs are sequenced.
5. The short sequences are mapped by computer onto the genome sequence.



Determining Gene Function

- One way to determine function is to disable the gene and observe the consequences
- Using ***in vitro* mutagenesis**, specific mutations are introduced into a cloned gene, altering or destroying its function
- When the mutated gene is returned to the cell, the normal gene's function might be determined by examining the mutant's phenotype

In vitro site directed mutagenesis

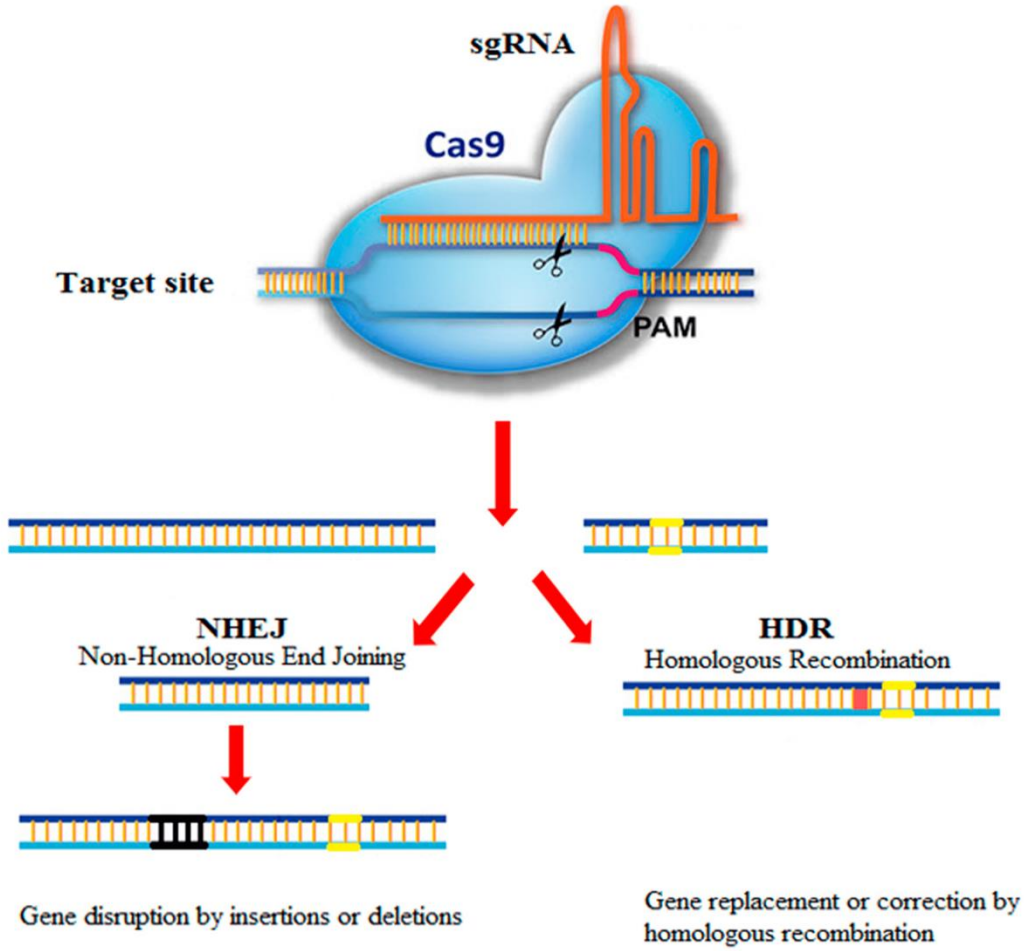


Gene editing: CRISPR-Cas9

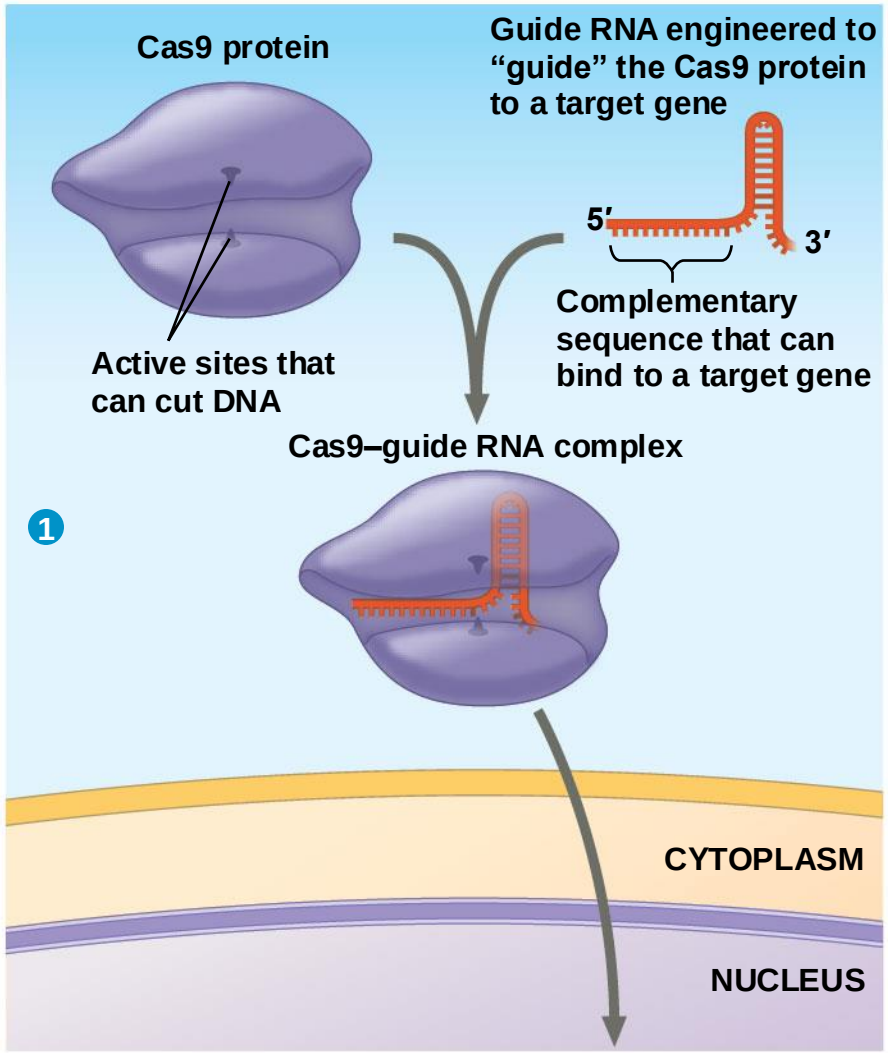
- The **CRISPR-Cas9 system** is a powerful new technique for gene editing in living cells and organisms
- Cas9 acts together with a “guide RNA” made from the CRISPR region of the bacterial system
- It will cut both strands of any DNA sequence complementary to the guide RNA
- If the guide RNA is engineered to be complementary to the target gene, the target DNA will be cut

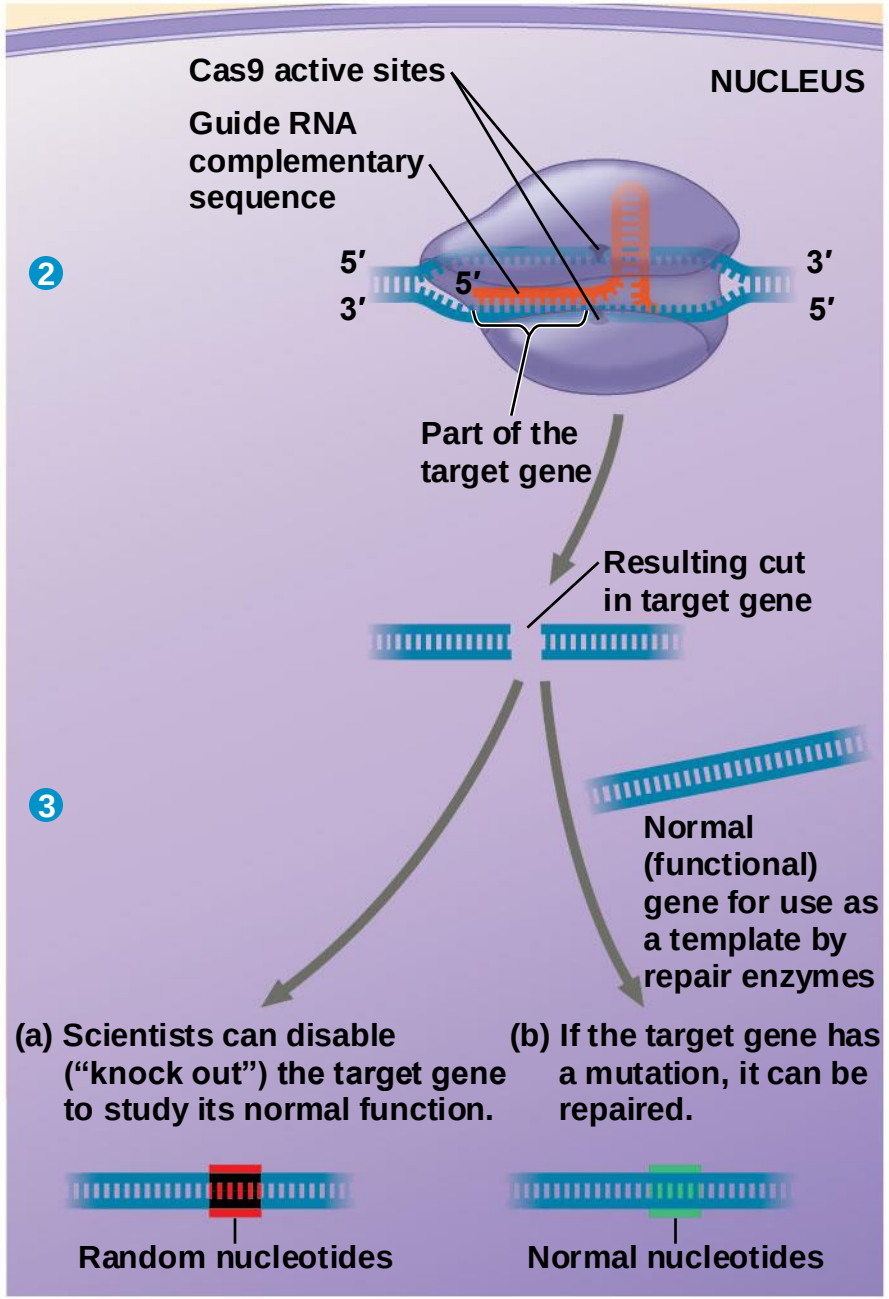
Gene editing: CRISPR-Cas9

- When the cut DNA is repaired, nucleotides may be introduced or removed, causing the gene to be inactivated
- Researchers have also modified the technique so a gene with a mutation in it can be repaired
- This is done by introducing a segment of the wild-type gene, which may be used as a template to repair the target DNA



Gene editing: CRISPR-Cas9

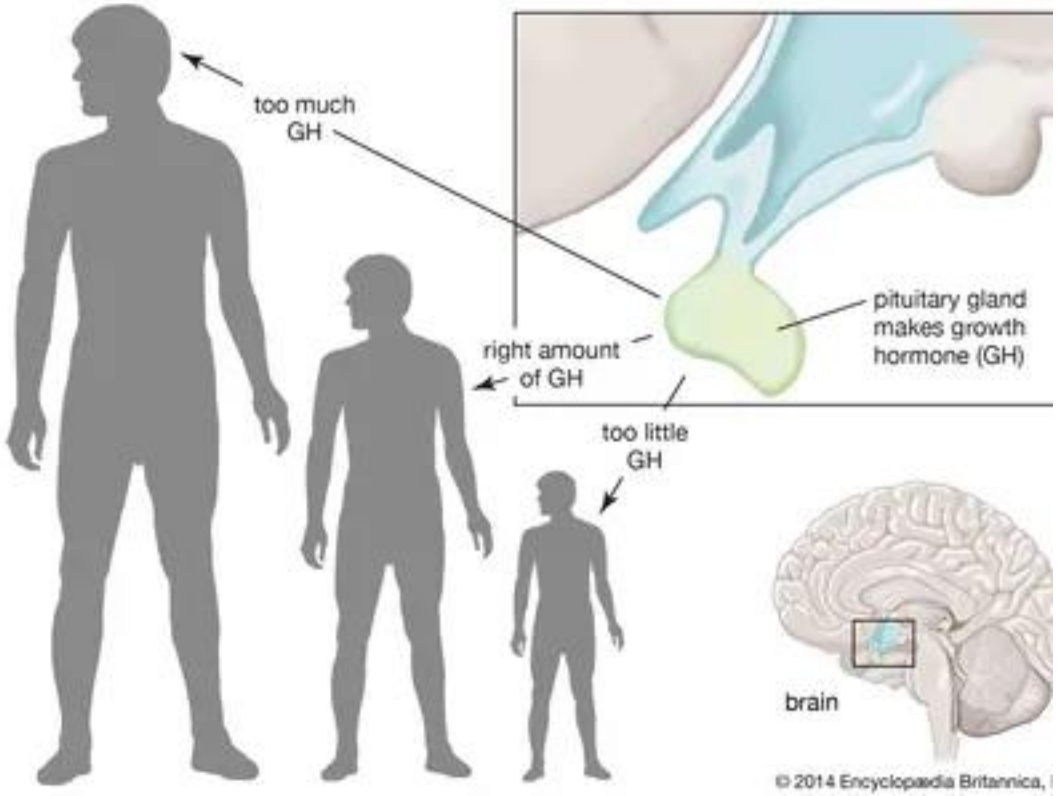




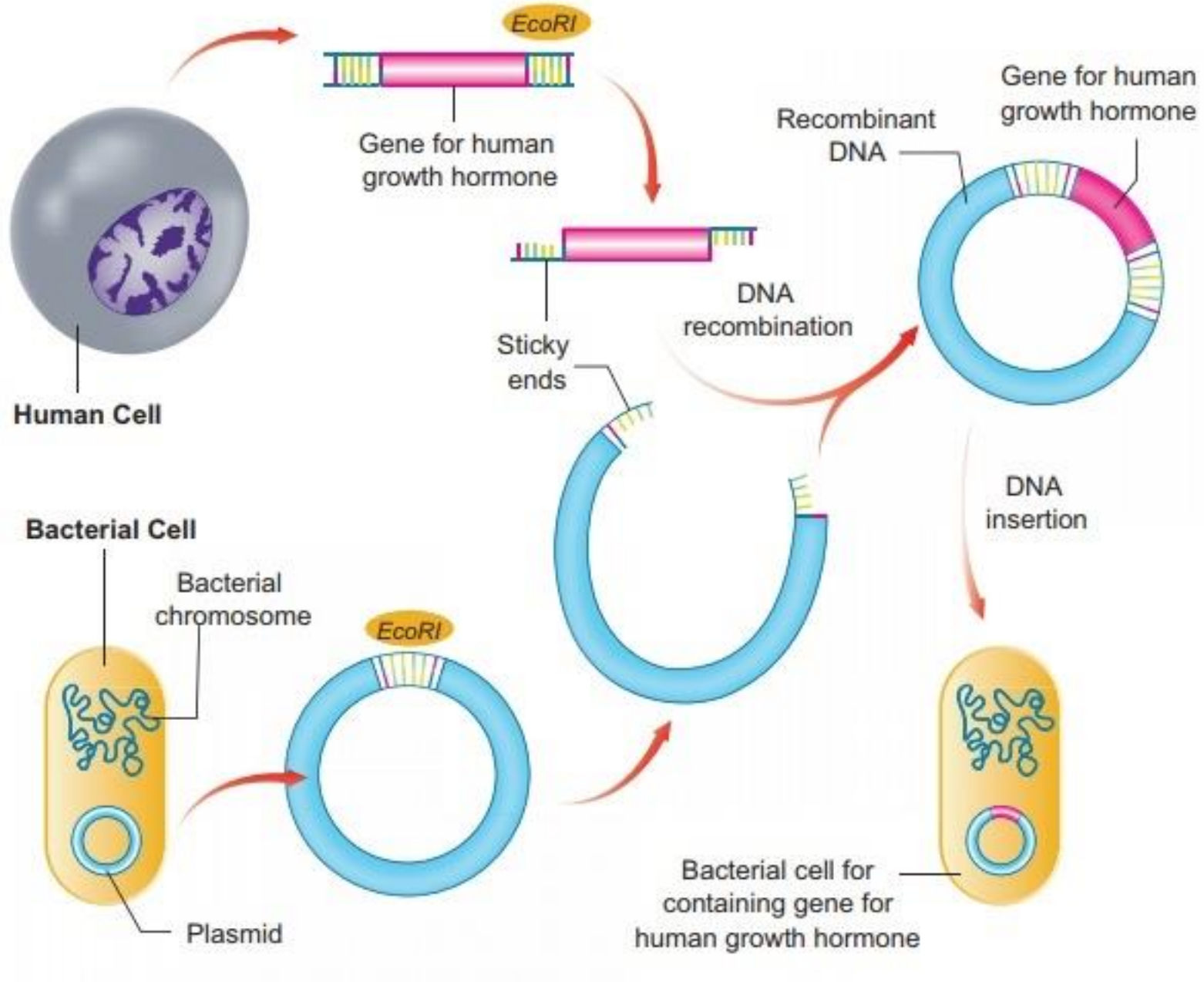
Biomedical applications

Protein Production in Cell Cultures

- Pharmaceutical products are commonly synthesized on a large scale using cell cultures
- The host cells can be engineered to secrete a protein as it is made, simplifying the task of purifying it
- Human insulin and human growth hormone (HGH) are among the first such products made in this way



Production of human growth hormone

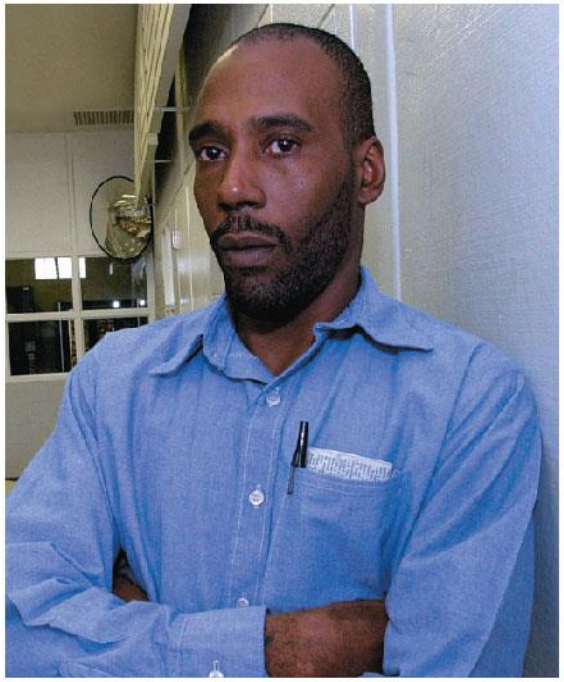


Forensic Evidence and Genetic Profiles

- An individual's unique set of genetic markers, or **genetic profile**, can be obtained by analysis of tissue or body fluids
- DNA testing can identify individuals with a high degree of certainty
- Genetic profiles are currently analyzed using genetic markers **called short tandem repeats (STRs)**

- STRs are variations in the number of repeats of specific DNA sequences
- PCR and gel electrophoresis are used to amplify and then identify STRs of different lengths
- The probability that two people who are not identical twins have the same STR markers is exceptionally small
- As of 2016 more than 340 innocent people have been released from prison as a result of STR analysis of old DNA evidence

(a) Earl Washington just before his release in 2001, after 17 years in prison

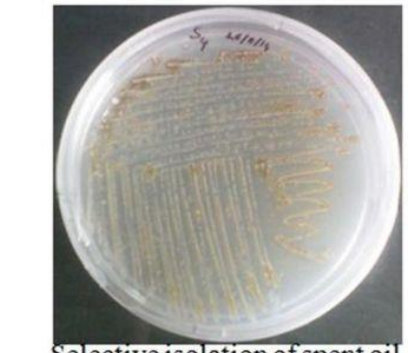


Source of sample	STR marker 1	STR marker 2	STR marker 3
Semen on victim	17,19	13,16	12,12
Earl Washington	16,18	14,15	11,12
Kenneth Tinsley	17,19	13,16	12,12

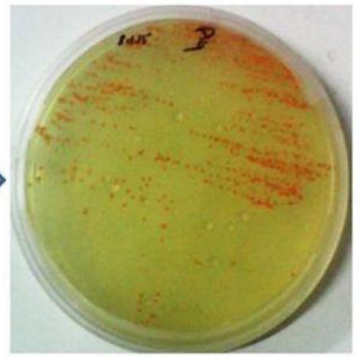
(b) These and other STR data (not shown) exonerated Washington and led Tinsley to plead guilty to the murder.

Environmental Cleanup

- Genetic engineering can be used to modify the metabolism of microorganisms
- Some modified microorganisms can be used to extract minerals from the environment or degrade potentially toxic waste materials

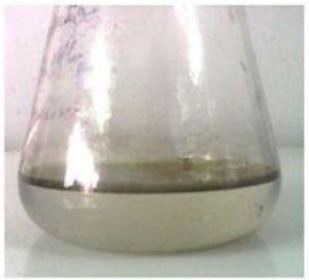


Selective isolation of spent oil degrading bacteria in BH agar media



Purified colonies of *G.terrae*

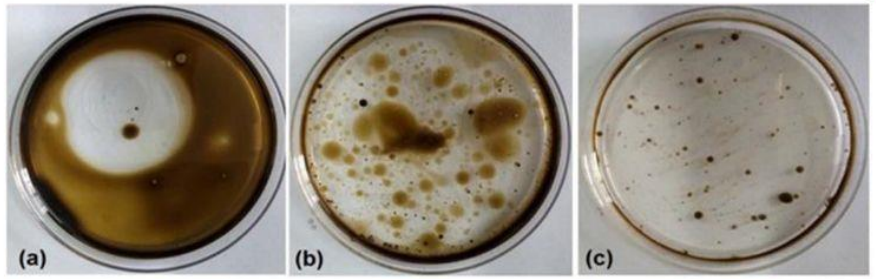
Slant culture of *G.terrae* for storage



Before degradation of spent oil

After degradation

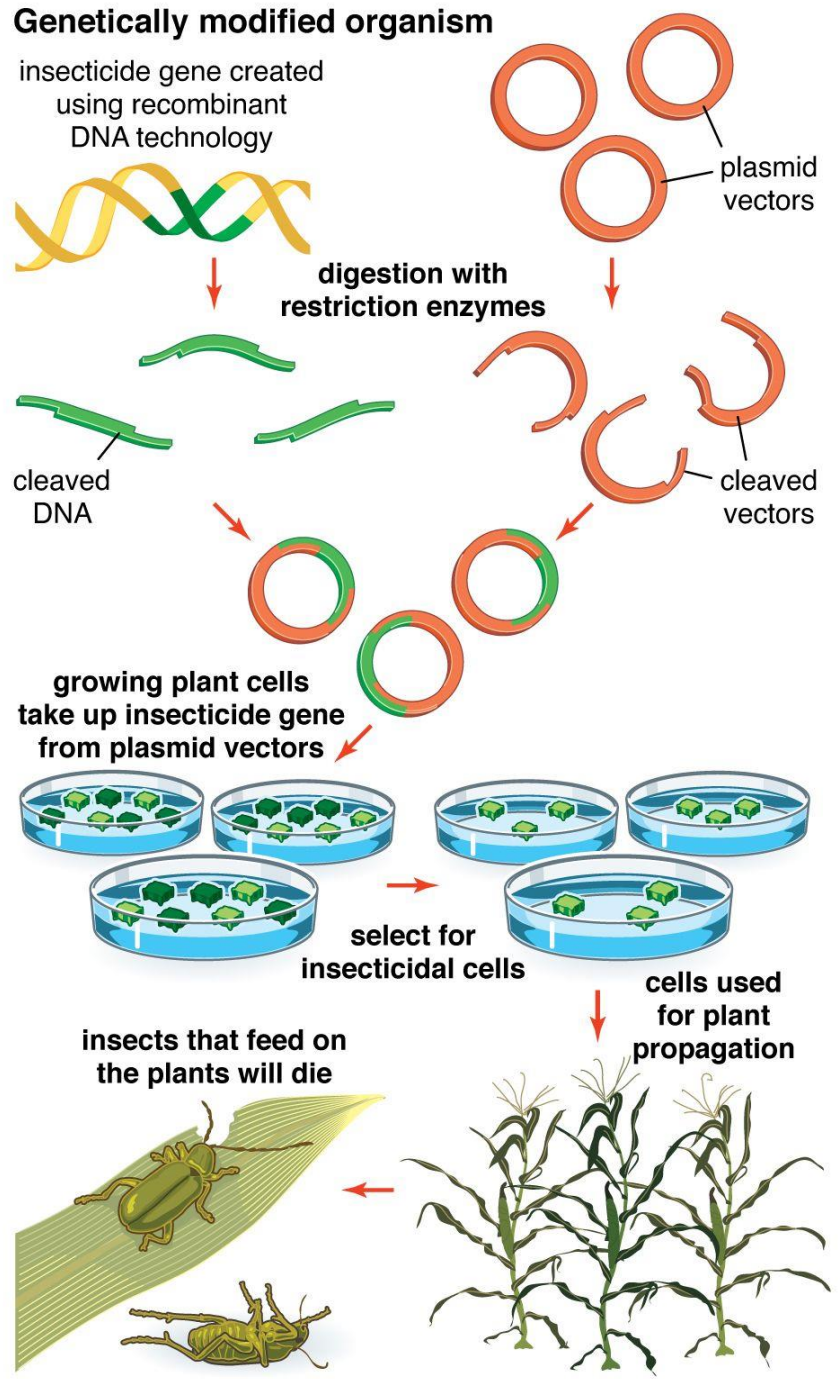
Liquid-liquid solvent extraction of residual oil for gravimetric analysis after degradation



Oil displacement due to biosurfactant production in the course of free (a), coarse (b) and micro-emulsified oil degradation

Agricultural Applications

- DNA technology is being used to improve agricultural productivity
- Agricultural scientists have endowed a number of crop plants with genes for desirable traits
- Genetic engineering in plants has been used to transfer many useful genes, including those for herbicide or pest resistance, increased resistance to salinity, and improved nutritional value of crops

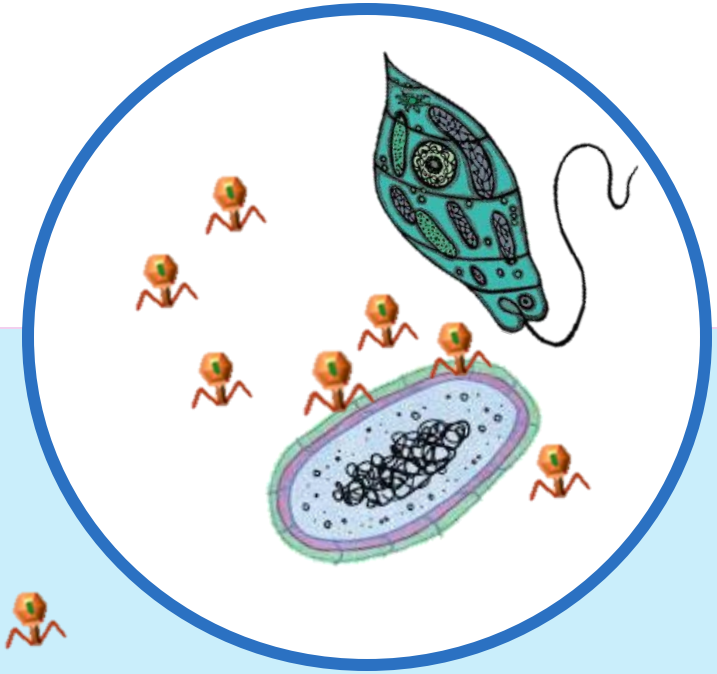


Extra slides: not to be evaluated

Case study: investigating bacteria communities in the aquatic ecosystems

Microbial communities are the most abundant life forms in arctic lakes

10^9-10^{10} phages/liter
10^7-10^9 bacteria
10^2-10^4 protists

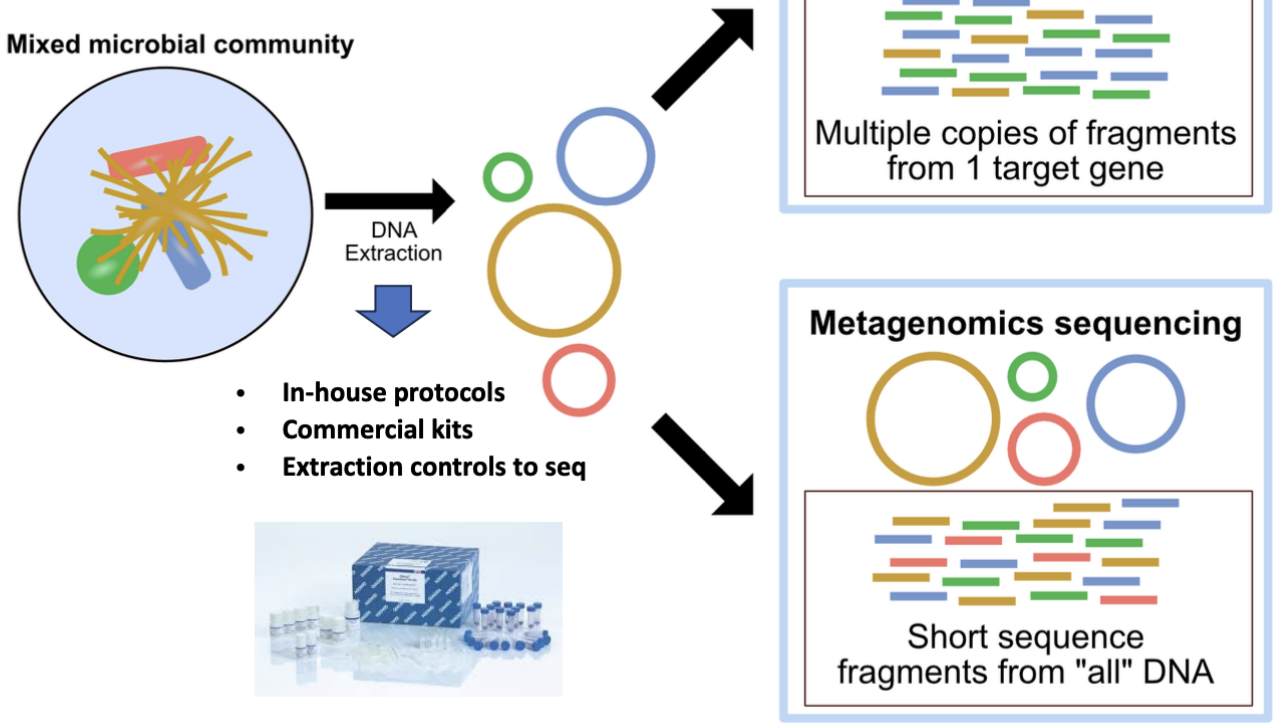


Biodiversity in alpine and arctic lakes adapted to:

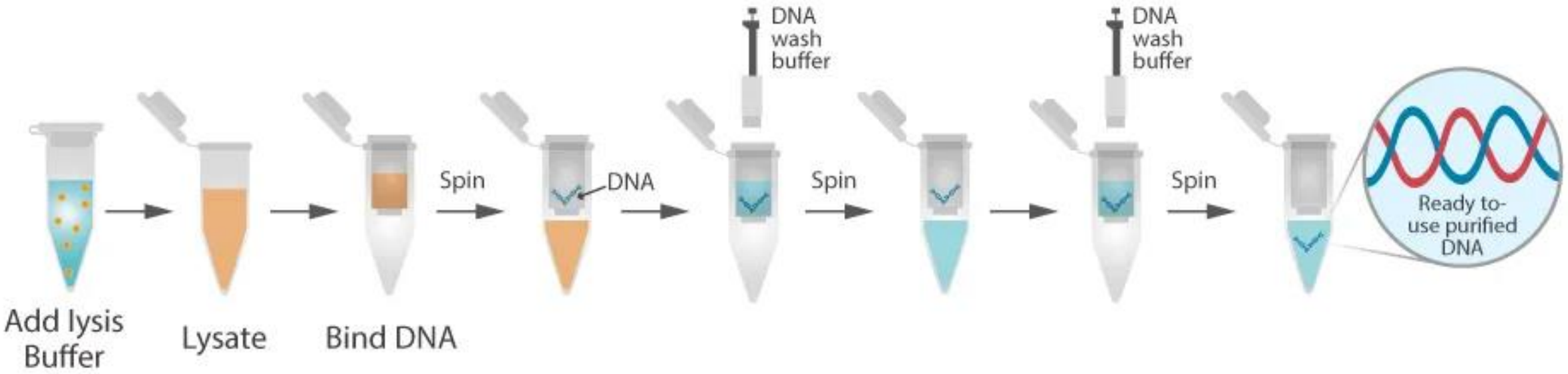
- High UV radiation exposure
- Extreme and quick environmental changes
- Ice cover for several months during the year
- Low organic inputs, low nutrients

Sample pre-processing

- Filtering (0.8, 0.45, 0.22 µm…)
- Fixation
- Culturing, enrichments…
- Measuring environmental parameters as metadata
- Appropriate sample conservation

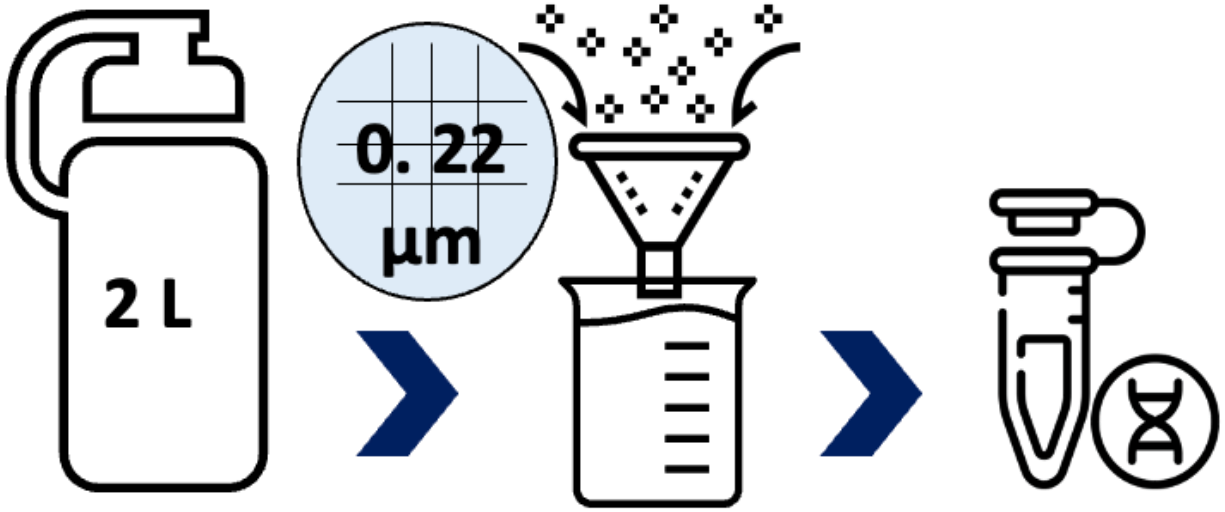


Nucleic acid extractions



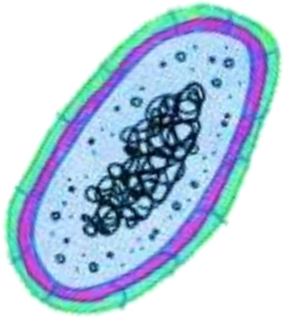
Methodology

- **PCR**
- **DNA Sequencing**



16/18S metagenomics Illumina NextSeq

qPCR (16s rRNA, mcyA, 18s…)



Taxonomic abundance (reads/total sequencing reads)

Diversity indices (alpha as Chao1, beta as Bray-Curtis dissimilarity distance)

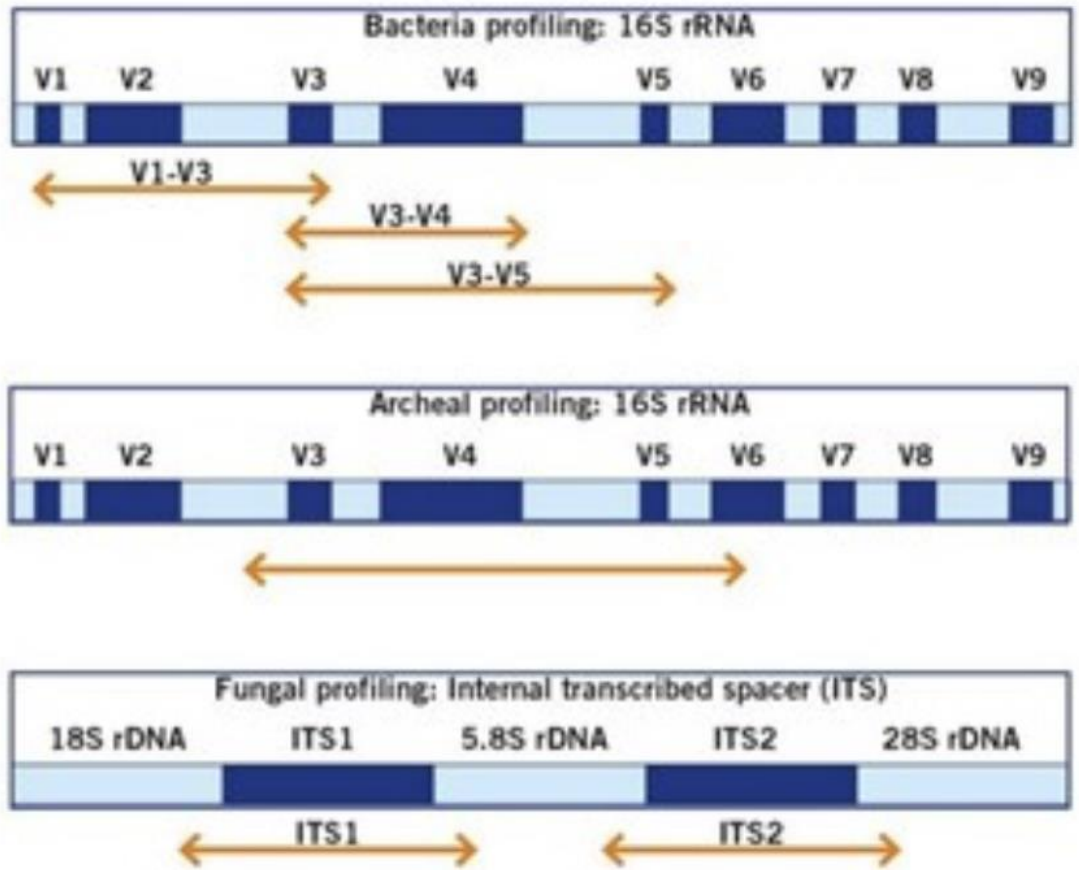
Functional predictions (Picrust2)

Identification of biomarkers (LefSe)

Correlation and redundancy analysis (RDA plots, environmental associations)

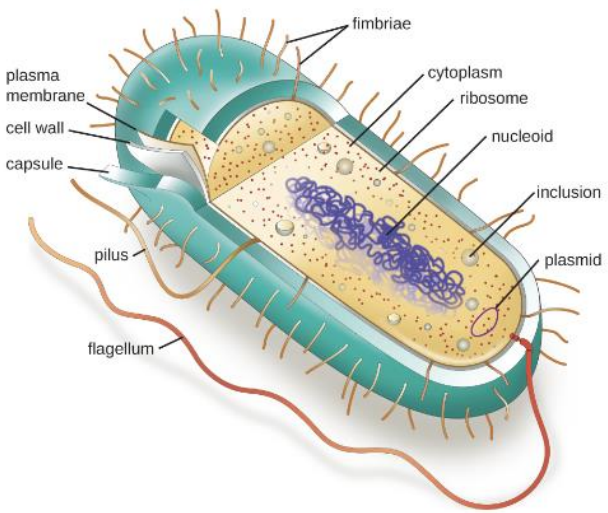
Amplicon sequencing:

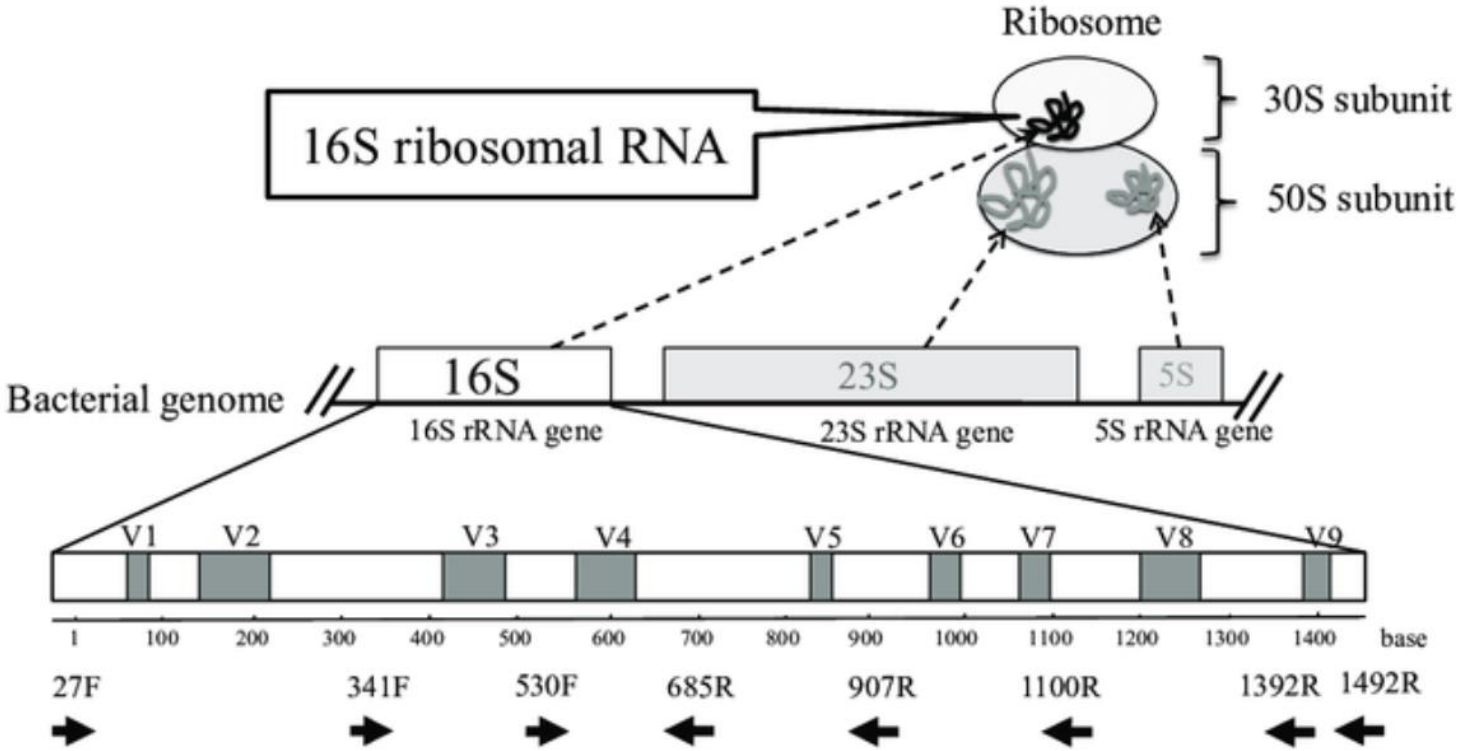
- Sequence a well-conserved gene
- Based on "universal" (-ish) primers for **PCR**
- Gene sequence is a barcode for species
- Usually, ribosomal RNA
- Bacteria/Archaea: 16S rRNA
- Fungi: Internal Transcribed Spacer (ITS)
- Other genes occasionally used, e.g. COI



What is 16S rRNA?

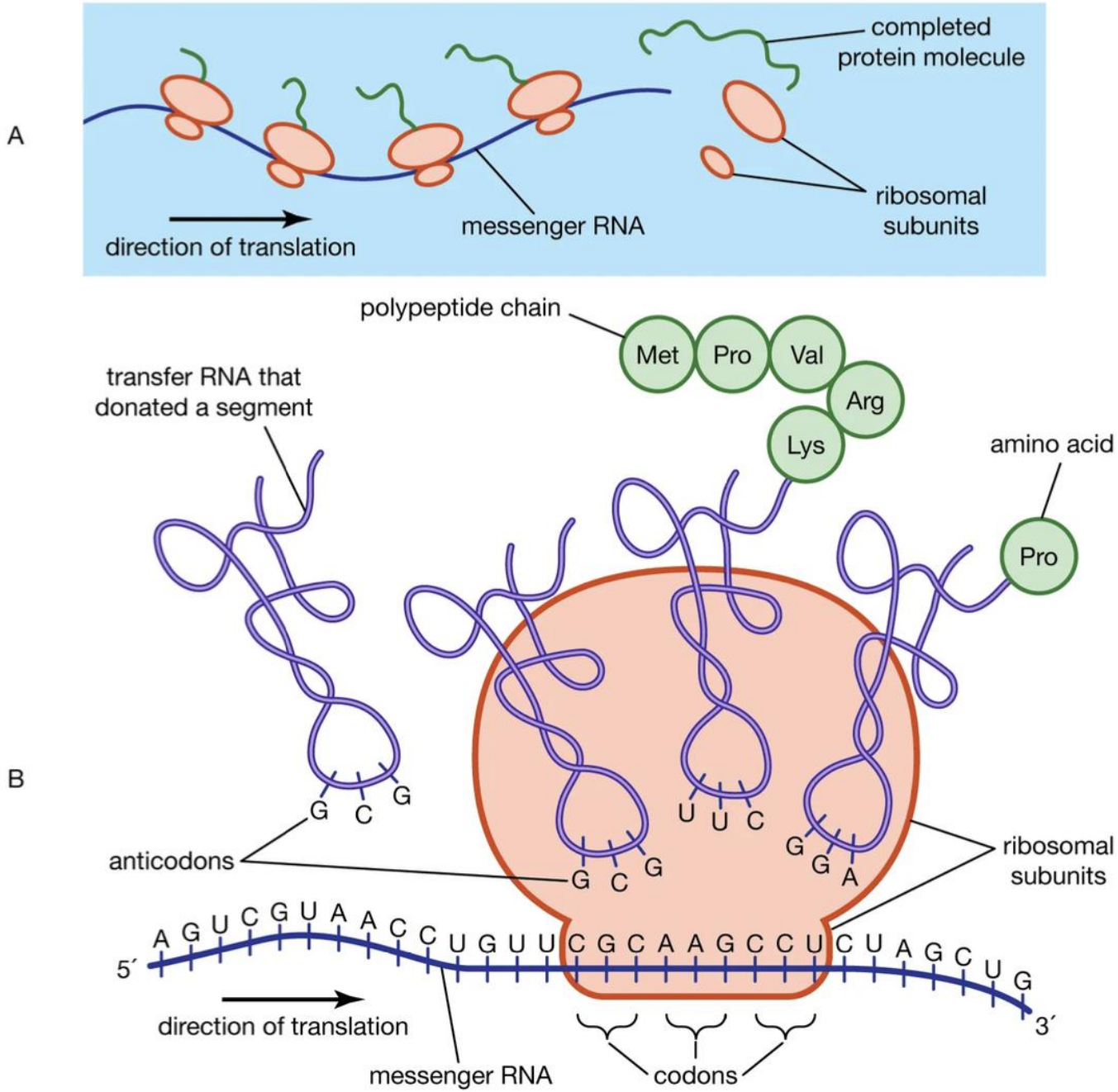
- 16S rRNA is a type of ribosomal RNA, a key structural and functional component of ribosomes—the cellular machinery responsible for protein synthesis.
- Found in the small subunit of prokaryotic ribosomes, 16S rRNA is highly conserved across different bacterial and archaeal species.



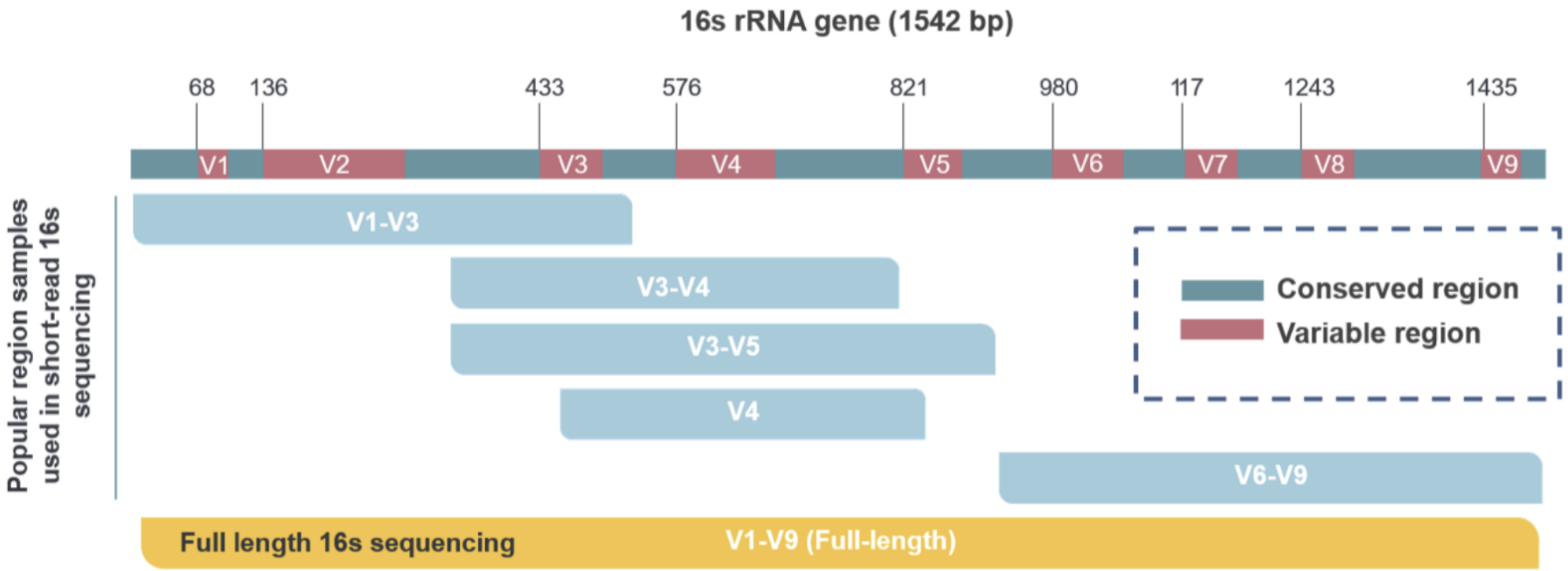


Functional significance

- The primary function of 16S rRNA is to help align and stabilize the ribosomal components during protein synthesis, facilitating the decoding of mRNA (messenger RNA) and the binding of tRNA (transfer RNA) molecules.

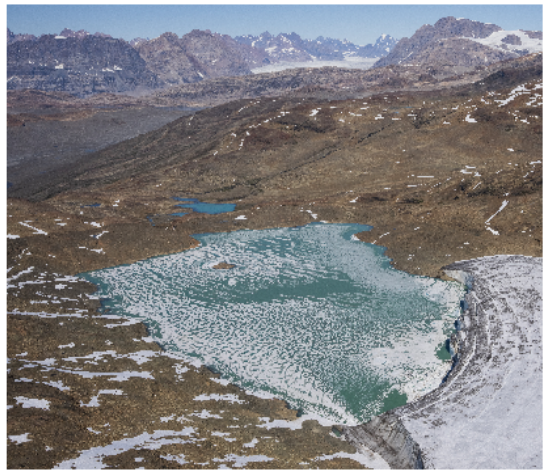


16s rRNA gene regions

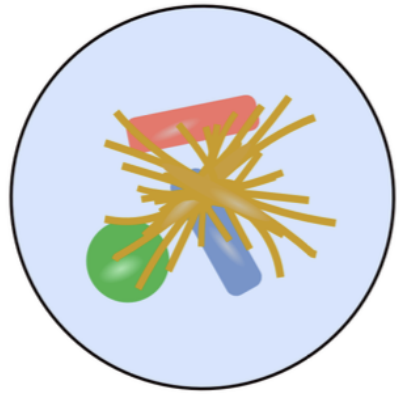


Workflow of 16S rRNA Sequencing

- **Sample collection**
- **DNA extraction**
- **16s rRNA PCR amplification**
- Library preparation
- Sequencing
- Data analysis

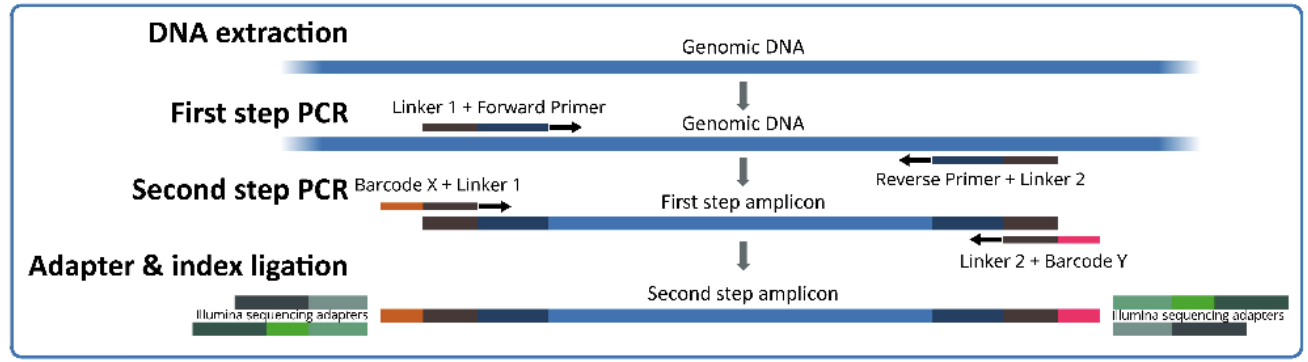


Mixed microbial community



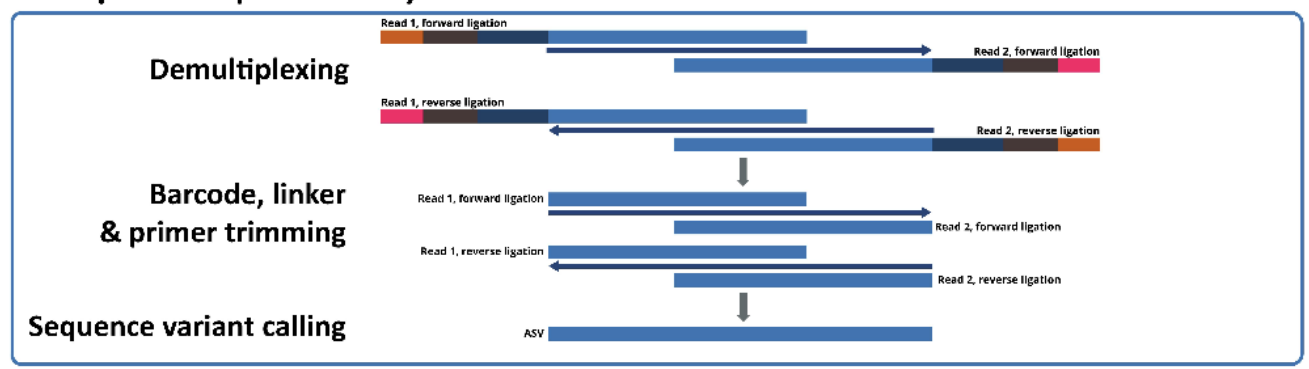
DNA Extraction

1. Amplicon generation



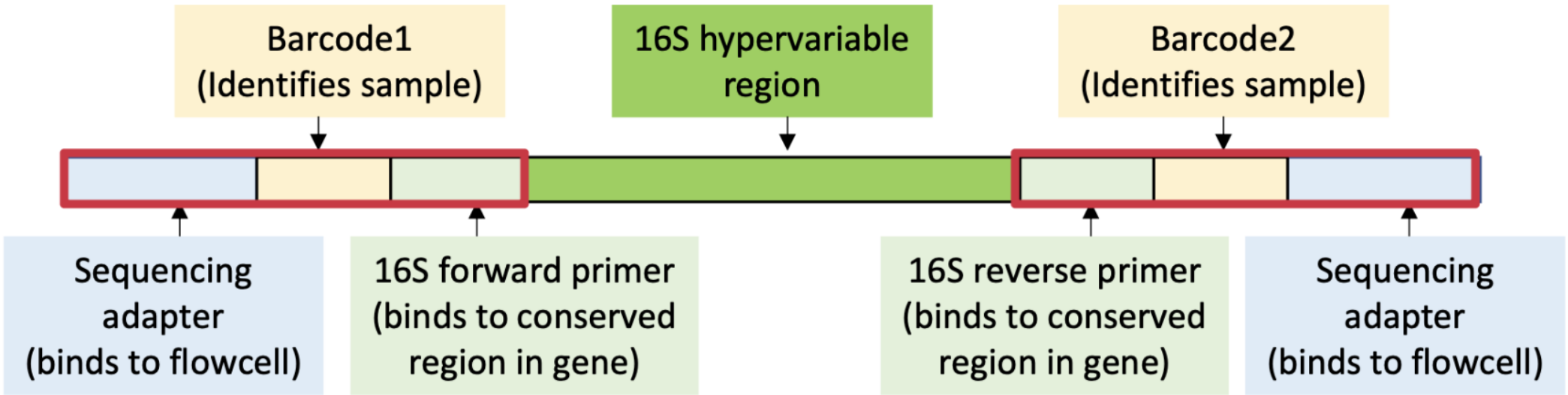
2. Sequencing

3. Amplicon sequence analysis



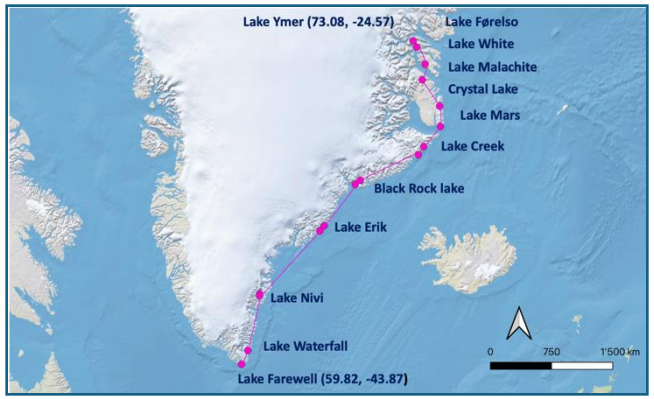
16s rRNA PCR: amplicon generation

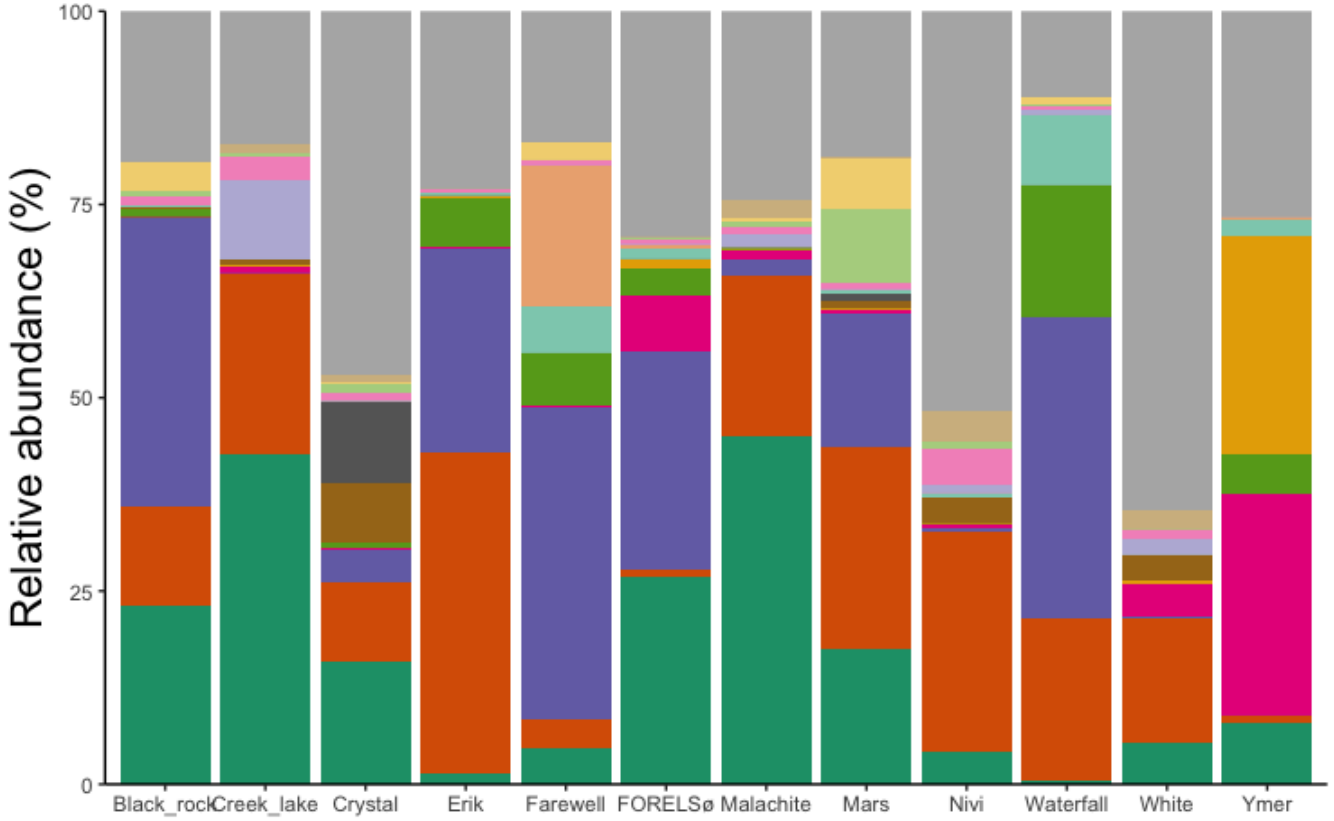
Oligos for PCR reaction



PCR + sequencing:

Bacteria community composition in East Greenland lakes

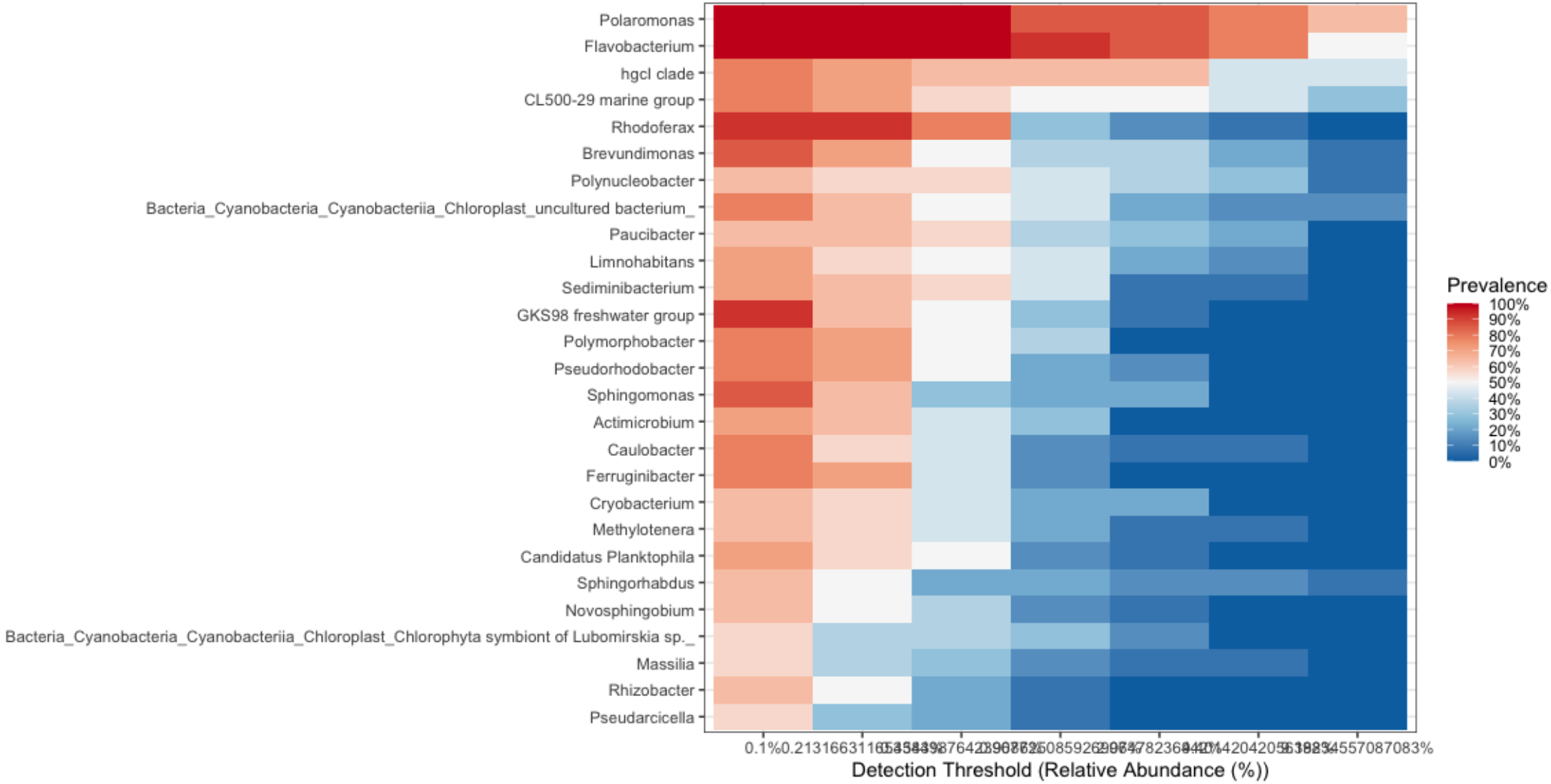




Genus

- Others
- Bacteria|Actinobacteriota|Actinobacteria|Micrococcales|Microbacteriaceae|Cryobacterium
- Bacteria|Cyanobacteria|Cyanobacteriia|Chloroplast|Chlorophyta symbiont of Lubomirskia sp.|
- Bacteria|Proteobacteria|Gammaproteobacteria|Burkholderiales|Methylophilaceae|Methylotenera
- Bacteria|Proteobacteria|Gammaproteobacteria|Burkholderiales|Comamonadaceae|Rhodoferax
- Bacteria|Actinobacteriota|Actinobacteria|Micrococcales|Micrococcaceae|Pseudarthrobacter
- Bacteria|Actinobacteriota|Actinobacteria|Corynebacteriales|Mycobacteriaceae|Mycobacterium
- Bacteria|Proteobacteria|Gammaproteobacteria|Burkholderiales|Comamonadaceae|Limnohabitans
- Bacteria|Actinobacteriota|Acidimicrobiia|IMCC26256|Actinobacteria bacterium IMCC26256|
- Bacteria|Actinobacteriota|Actinobacteria|Micrococcales|Microbacteriaceae|Parafrigoribacterium
- Bacteria|Proteobacteria|Alphaproteobacteria|Sphingomonadales|Sphingomonadaceae|Sphingorhabdus
- Bacteria|Proteobacteria|Gammaproteobacteria|Burkholderiales|Burkholderiaceae|Polynucleobacter
- Bacteria|Proteobacteria|Alphaproteobacteria|Caulobacterales|Caulobacteraceae|Brevundimonas
- Bacteria|Actinobacteriota|Actinobacteria|Frankiales|Sporichthyaceae|hgcI clade
- Bacteria|Proteobacteria|Gammaproteobacteria|Burkholderiales|Comamonadaceae|Polaromonas
- Bacteria|Bacteroidota|Bacteroidia|Flavobacteriales|Flavobacteriaceae|Flavobacterium

PCR + sequencing: Identification of core microbiome in East Greenland lakes



PCR + sequencing: estimating bacteria diversity in East Greenland lakes