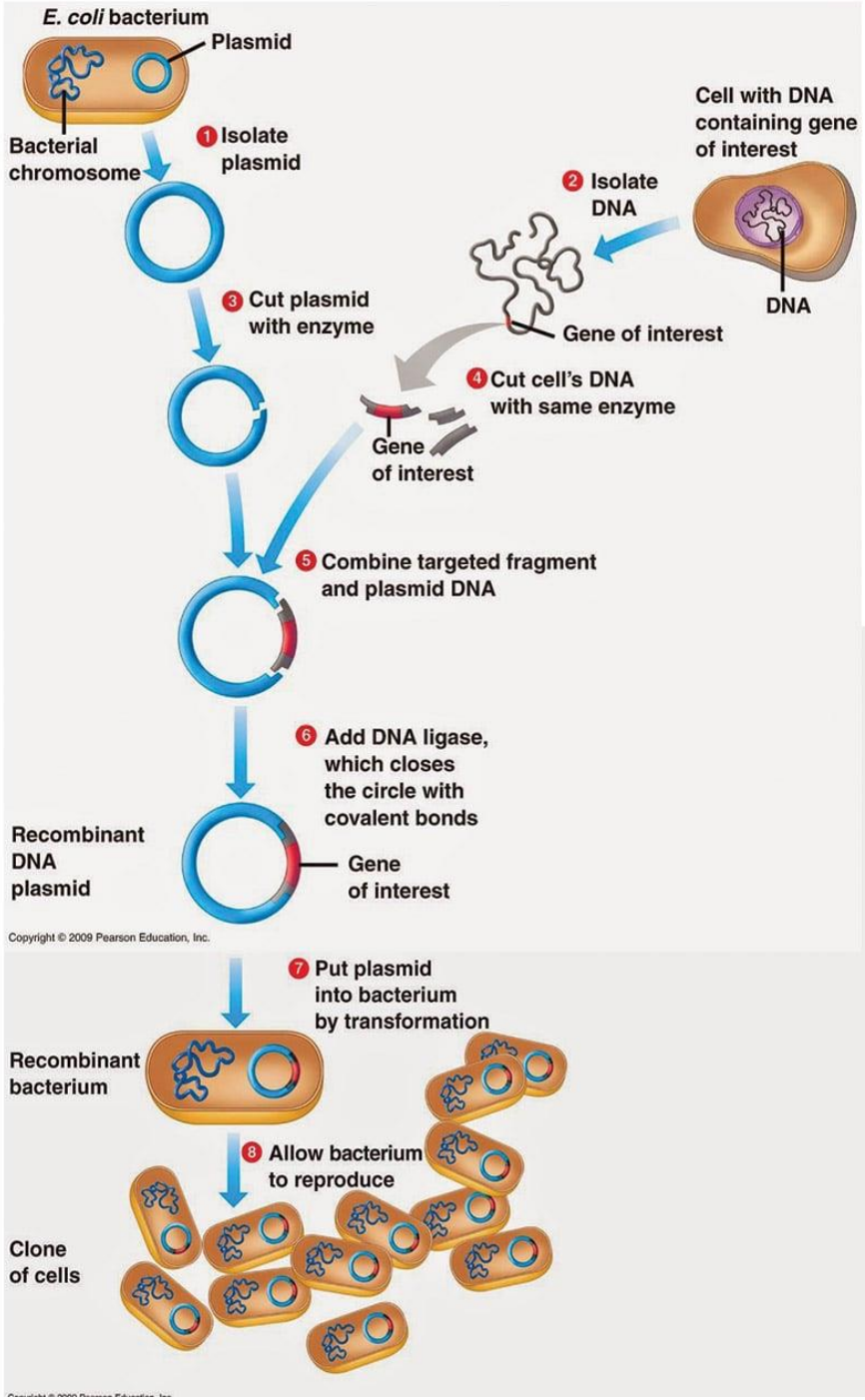
Molecular Cloning in the Laboratory

Irvine Tatenda Mutiti

Molecular Cloning

- Molecular cloning is the process of isolating a defined DNA sequence and inserting it into a vector so that it can be copied, stored, analysed, or expressed in a host cell
- At its simplest, molecular cloning allows us to:
  - Isolate or generate a gene or DNA fragment of interest
  - Insert it into a plasmid or other vector
  - Introduce that vector into a host cell
  - Select cells that carry the desired construct
  - Verify that the construct is correct
  - Use the construct for downstream experiments



Why Molecular Cloning Matters in Biology

- Molecular cloning underpins much of modern molecular and cellular biology.

  It is used to:

  - Study gene function
  - Express proteins in cells
  - Produce recombinant proteins
  - Create mutations or deletions in genes
  - Study promoter activity
  - Engineer cell lines
  - Build tools for microscopy, biochemistry, immunology, and neuroscience

The Central Logic of Molecular Cloning

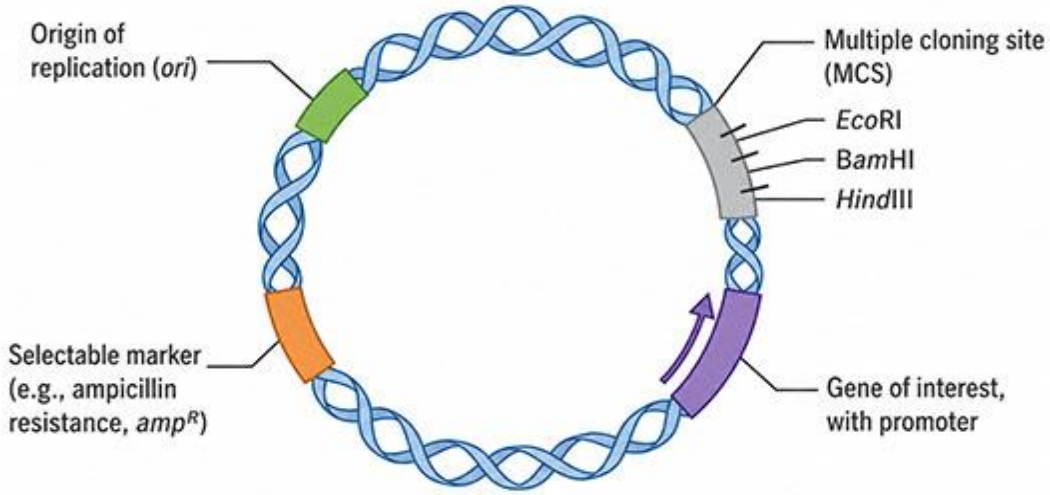
1. Define the DNA sequence of interest (target DNA)
2. Choose an appropriate vector
3. Design a cloning strategy
4. Generate compatible sticky ends
5. Join insert and vector
6. Introduce the construct (recombinant DNA) into bacteria
7. Select and screen bacterial colonies
8. Purify plasmid DNA
9. Verify the final construct
10. Use the plasmid in downstream experiments

Terminology

- **Insert**
  The DNA fragment that you want to clone
- **Vector**
  The DNA molecule that carries the insert. Most commonly this is a plasmid.
- **Plasmid**
  A circular DNA molecule that can replicate independently in bacteria.
- **Restriction enzyme**
  An enzyme that cuts DNA at specific sequences.
- **Ligase**
  An enzyme that joins DNA fragments together by forming phosphodiester bonds between nucleotides.
- **Transformation**
  The process of introducing plasmid DNA into bacterial cells.
- **Selection**
  The process of growing only bacteria that carry the plasmid, usually using antibiotic resistance.

Plasmids

- A plasmid is a small, circular DNA molecule used to carry and amplify DNA sequences.
- A typical plasmid contains:
  - **Origin of replication** - allows plasmid replication in bacteria
  - **Transcription termination/polyadenylation signals -** important in mammalian expression vectors
  - **Antibiotic resistance gene** - allows selection of bacteria carrying the plasmid
  - **Multiple cloning site** - region containing restriction enzyme sites
  - **Promoter** - drives gene expression, if the plasmid is an expression vector
  - **Tag or reporter sequence** - for detection, purification, or localisation

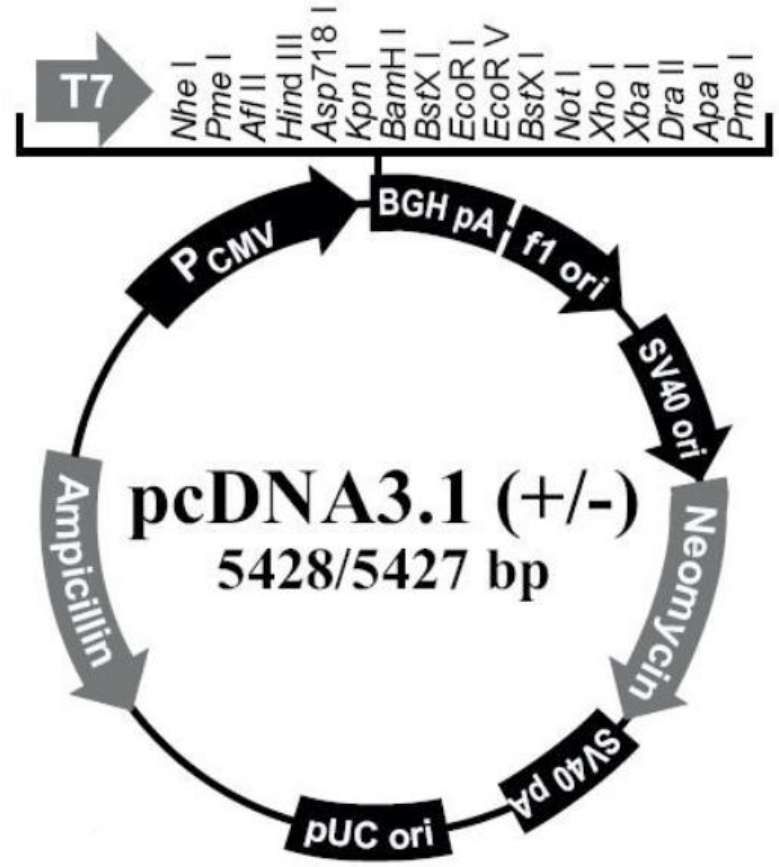


Choosing the Right Vector

- The choice of vector depends on the purpose of the experiment.
- Try to answer the following questions before getting started:
  - Do you want to store the DNA or express it?
  - Will expression occur in bacteria, mammalian cells, yeast, or another system?
  - Do you need a fluorescent tag, epitope tag, or purification tag?
  - Is antibiotic selection required?
  - Does the insert need to be in a specific reading frame?
  - Is the promoter suitable for the host cell?
  - Is the plasmid size manageable?

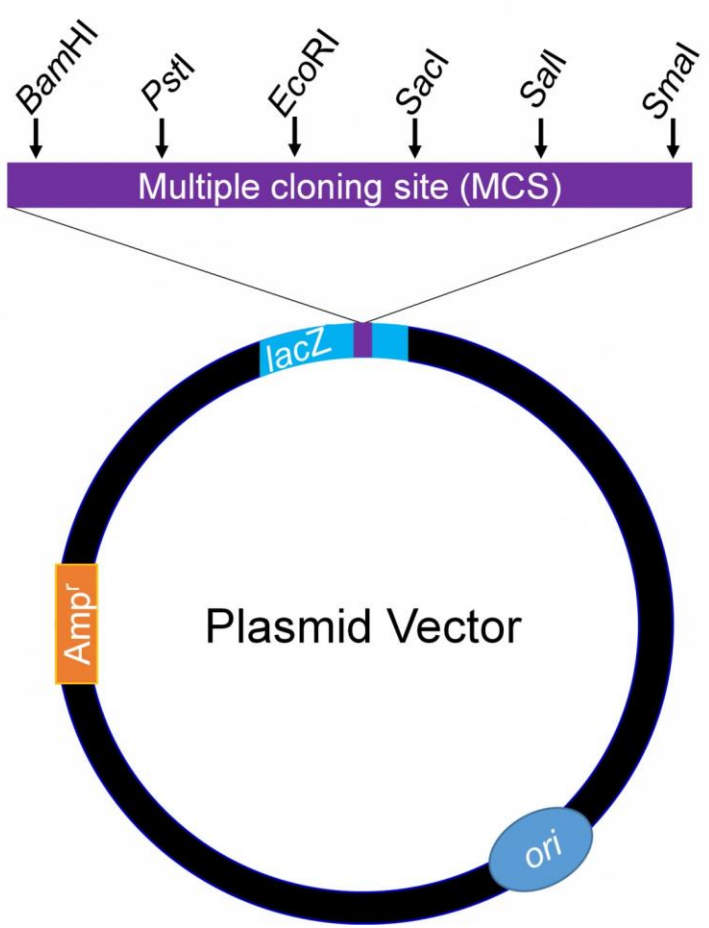
Examples of vector types:

- Cloning vectors
- Bacterial expression vectors
- Mammalian expression vectors
- Reporter vectors
- Viral vectors
- CRISPR/Cas-related vectors



The Multiple Cloning Site

- The multiple cloning site, or MCS, is a short region of a plasmid containing several unique restriction enzyme sites.
- It allows you to insert DNA fragments using different enzyme combinations.
- When using an MCS, consider the following:
  - Which sites are unique in the vector
  - Whether those sites are absent from the insert
  - Whether digestion preserves the correct reading frame
  - Whether the insert orientation is controlled
  - Whether tags or promoters are upstream or downstream



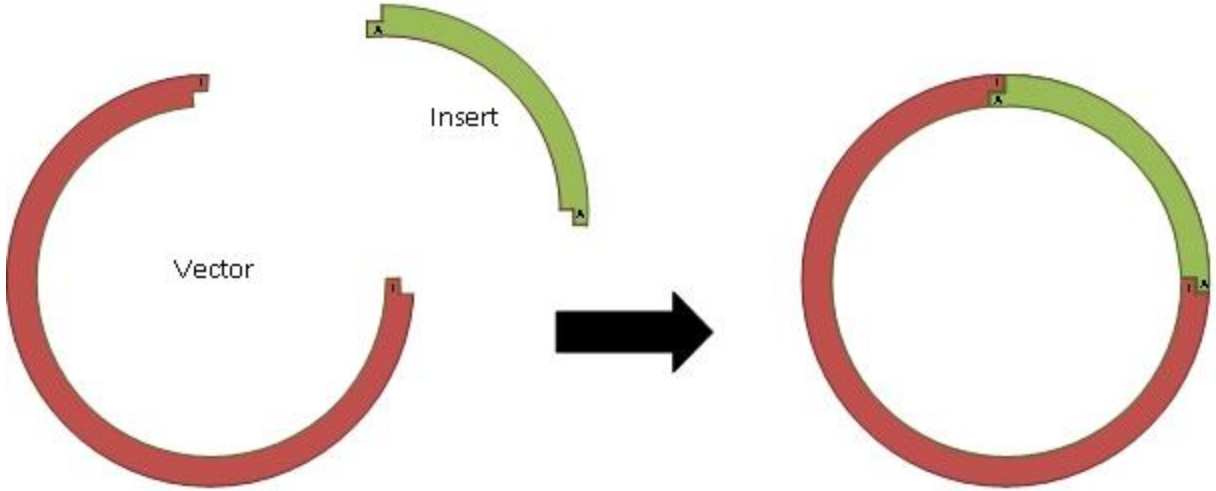
The Insert: Defining the DNA Fragment of Interest

The insert may be:

- A full-length coding sequence
- A gene fragment
- A promoter region
- A regulatory element
- A mutated sequence
- A synthetic DNA fragment

Define these before cloning:

- The exact start and end of the insert
- Whether introns are included or excluded
- Whether the sequence is genomic DNA, cDNA, or synthetic DNA
- Whether stop codons should be included or removed
- Whether tags are placed at the N-terminus or C-terminus



DNA sources

Genomic DNA

- Contains introns and exons
- Useful for studying genomic regions, promoters, or regulatory elements
- May be large and difficult to clone

cDNA

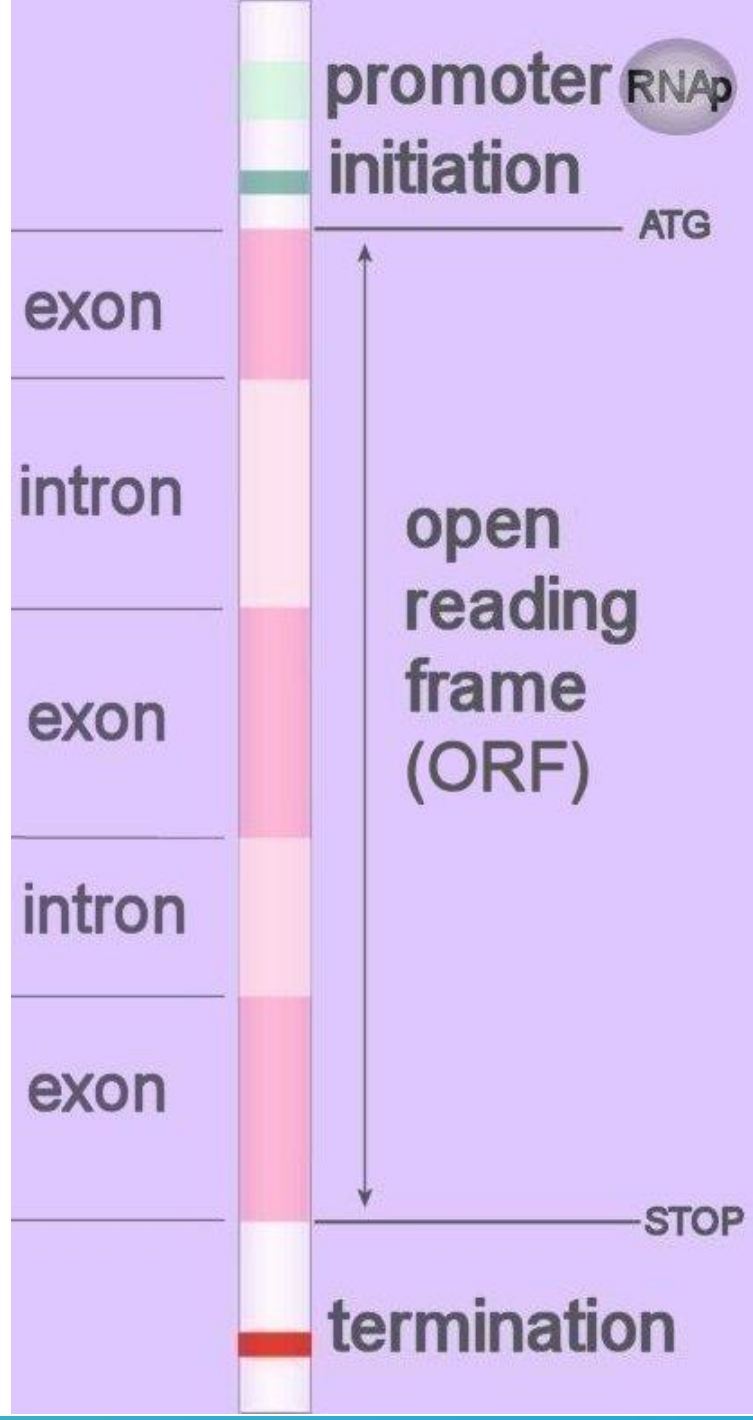
- Generated from mRNA by reverse transcription
- Contains expressed coding sequence without introns
- Commonly used for protein expression

Synthetic DNA

- Designed and ordered commercially
- Can include codon optimisation, mutations, tags, or restriction sites
- Useful when templates are unavailable or sequence design is complex

Reading Frames

- When cloning coding sequences, the reading frame must be preserved.
- A reading frame is determined by how nucleotides are grouped into codons.
- If an insert is cloned out of frame:
  - The protein sequence will be incorrect
  - Premature stop codons may appear
  - The tag may not be translated properly
  - The expressed protein may be non-functional

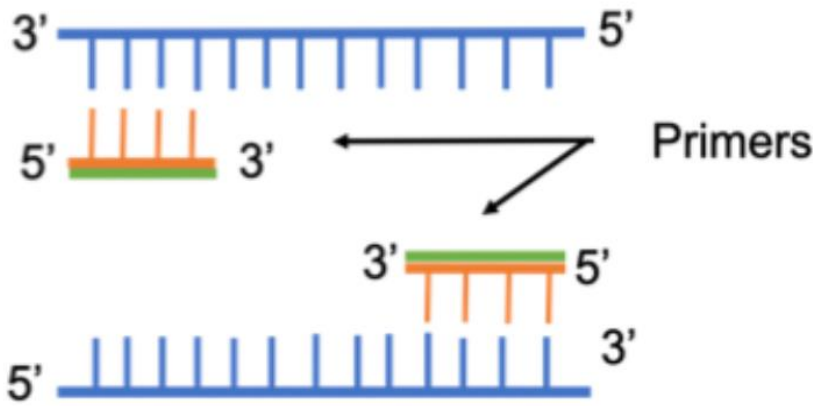


Primer Design

- Poor primer design is one of the most common reasons a cloning PCR fails.
- The part of the primer that binds to the template is the most important part for PCR performance. In most cases, this region is around 18–25 bases long.
- Good primers usually have:
  - 40-60% GC content
  - similar melting temperatures, ideally within 2-3°C
  - a clean, specific 3' end
  - no hairpins
  - no primer-dimers
  - no long runs of one base, such as AAAAA or GGGGG
  - no second binding site in the template
- The 3' end matters because this is where the polymerase starts extending. If that end binds in the wrong place, the PCR can amplify the wrong product.

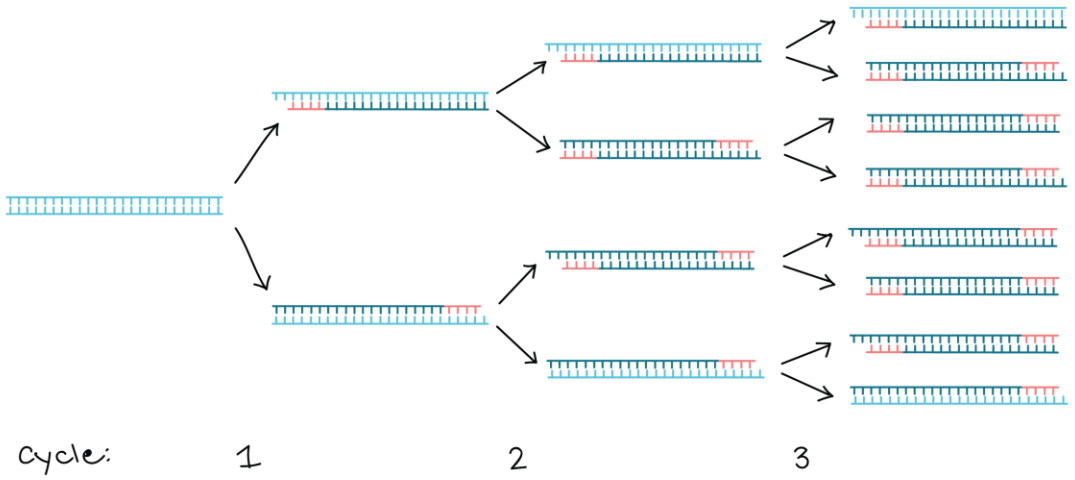
Primer design for Cloning

- Cloning primers often contain extra sequence at the 5' end.
- This extra sequence may not bind to the original template in the first PCR cycles, but it becomes part of the final PCR product.
- For restriction enzyme cloning, the 5' end of the primer may include:
  - 4- 6 extra bases to help the enzyme cut efficiently
  - a restriction enzyme site
  - then the gene-specific sequence
- For Gibson Assembly or In-Fusion the primer may include:
  - around 20 - 40 bases of overlap with the vector or neighbouring fragment
  - followed by the gene-specific sequence
- For expression constructs, the primer design may also need to deal with:
  - the start codon
  - whether to keep or remove the stop codon
  - the reading frame
  - a Kozak sequence for mammalian expression
  - a linker sequence
  - an N-terminal or C-terminal tag



PCR Amplification of the Insert

- PCR is often used to amplify the insert before cloning.
- PCR should be performed with a high-fidelity DNA polymerase, because ordinary Taq polymerase has a higher error rate and may introduce mutations into the insert.
- Common PCR design and reaction mixtures include:
  - Template DNA: 1-10 ng plasmid DNA, or 10 -100 ng genomic DNA/cDNA
  - Primer concentration: 0.2- 0.5 µM each
  - dNTP concentration: 200 µM each
  - Annealing temperature: set around 3 - 5°C below the primer melting temperature, or optimised using a gradient PCR
  - Typical annealing range: commonly 55 - 65°C, but high-Tm primers for Gibson or restriction-site cloning may require 65–72°C
  - Polymerase: high-fidelity enzymes such as Q5, Phusion, or PrimeSTAR
  - Extension time: often 15- 30 seconds per kb for many high-fidelity enzymes
  - Cycle number: usually 25- 35 cycles; fewer cycles reduce mutation risk
  - Final extension: 2- 5 minutes, depending on amplicon size and polymerase



Checking the PCR Product using Agarose Gel

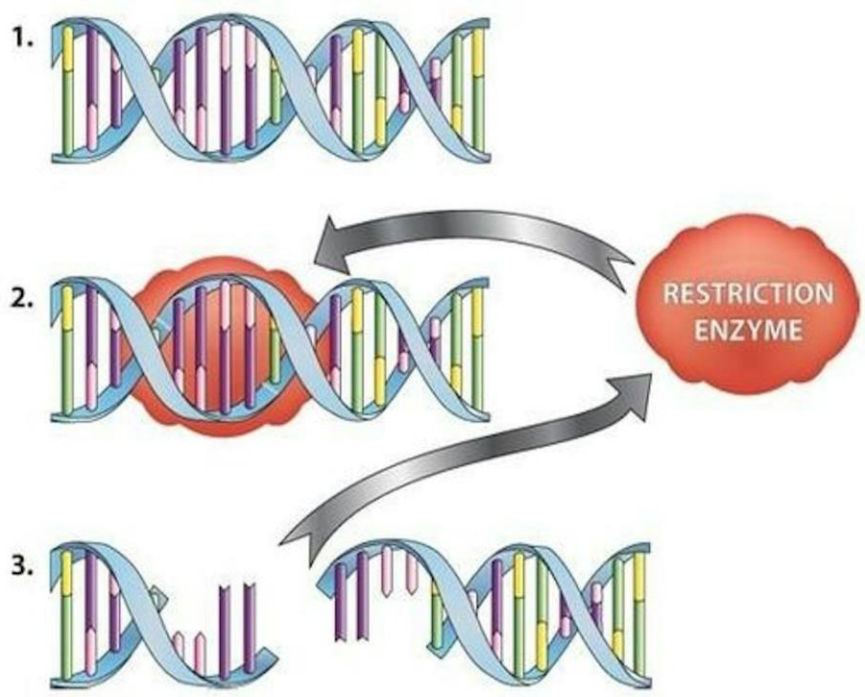
- Agarose gel electrophoresis is used to check whether PCR or restriction digestion produced DNA fragments of the expected size
- Gel conditions usually include:
  - 0.8 - 1% agarose gel for larger DNA fragments, such as 2 - 10 kb
  - 1.5 - 2% agarose gel for smaller fragments, such as 100 bp - 1 kb
  - DNA ladder chosen to match the expected fragment size
  - Loading dye added to samples before loading
  - DNA stain such as SYBR Safe, GelRed, or ethidium bromide
  - Electrophoresis run around 80 - 120 V, depending on gel size and apparatus

- A single band at the expected size supports successful amplification or digestion.
- However, gel size alone does not prove sequence correctness.
- A gel tells you whether the DNA is approximately the right size. Sequencing tells you whether it is actually the right DNA.

Restriction Enzyme Cloning: The Classical Approach

Restriction enzyme cloning uses enzymes that cut DNA at defined sequences.

1. Digest the vector with one or more restriction enzymes
2. Prepare the insert with compatible ends, either by digestion or PCR primer design
3. Purify the digested DNA fragments
4. Ligate insert and vector
5. Transform bacteria
6. Select and screen colonies
7. Verify the construct



Restriction Enzymes and DNA Ends

Restriction enzymes generate different types of DNA ends.

Sticky ends

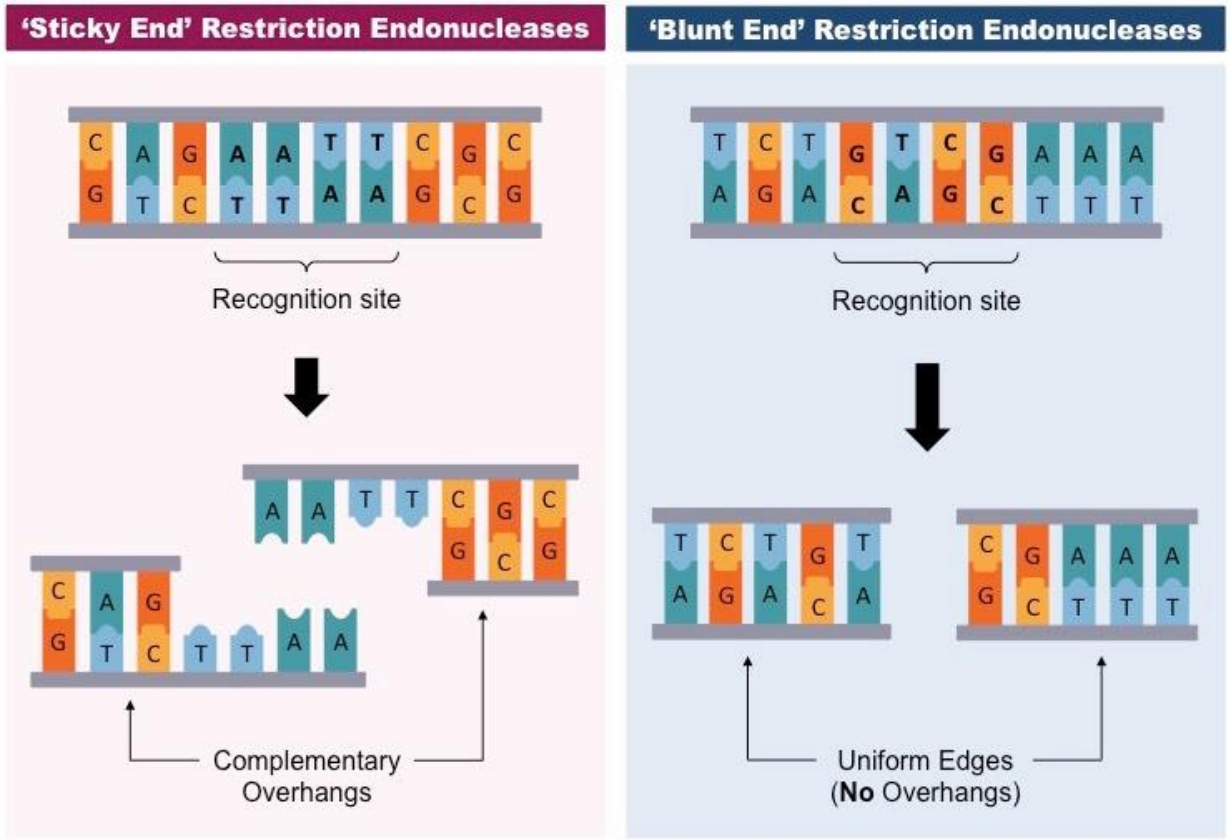
- Have short single-stranded overhangs
- Can base-pair with complementary ends
- Usually improve ligation efficiency

Blunt ends

- Have no overhang
- Can ligate to any other blunt end
- Often less efficient and less directional

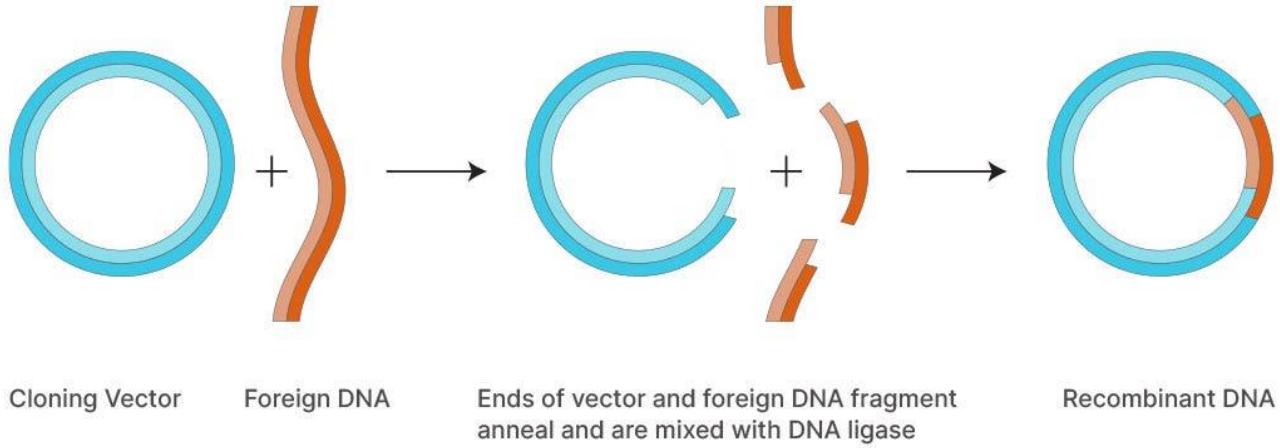
Directional cloning

- Uses two different restriction enzymes
- Forces the insert to enter the vector in one orientation
- Reduces background colonies
- Helps preserve reading frame



DNA Ligation

- DNA ligase joins DNA fragments by forming phosphodiester bonds between compatible DNA ends.
- Common practical conditions include:
  - Vector DNA: around 50-100 ng per ligation reaction
  - Sticky-end ligation: often performed at room temperature for 10-30 minutes
  - Blunt-end ligation: usually less efficient and often benefits from longer incubation, higher DNA concentration, or overnight ligation at 16°C
  - Insert-to-vector molar ratio: 3:1 as a starting point
  - Reaction volume: often 10-20 µL
  - Ligase: T4 DNA ligase
  - Buffer: must contain ATP, because ligation is ATP-dependent



Modern Assembly Methods

- Many cloning workflows now use assembly-based methods rather than classical restriction digestion.
- A few common examples:
  - Gibson Assembly
  - In-Fusion cloning
  - Golden Gate cloning
  - Gateway cloning
  - TOPO cloning
  - Ligation-independent cloning
- These approaches are usually faster, more flexible, and better suited to multi-fragment assembly.
- However, they still require careful design, clean DNA, and verification.

Gibson Assembly

Gibson Assembly joins DNA fragments that share overlapping ends.

The reaction typically uses:

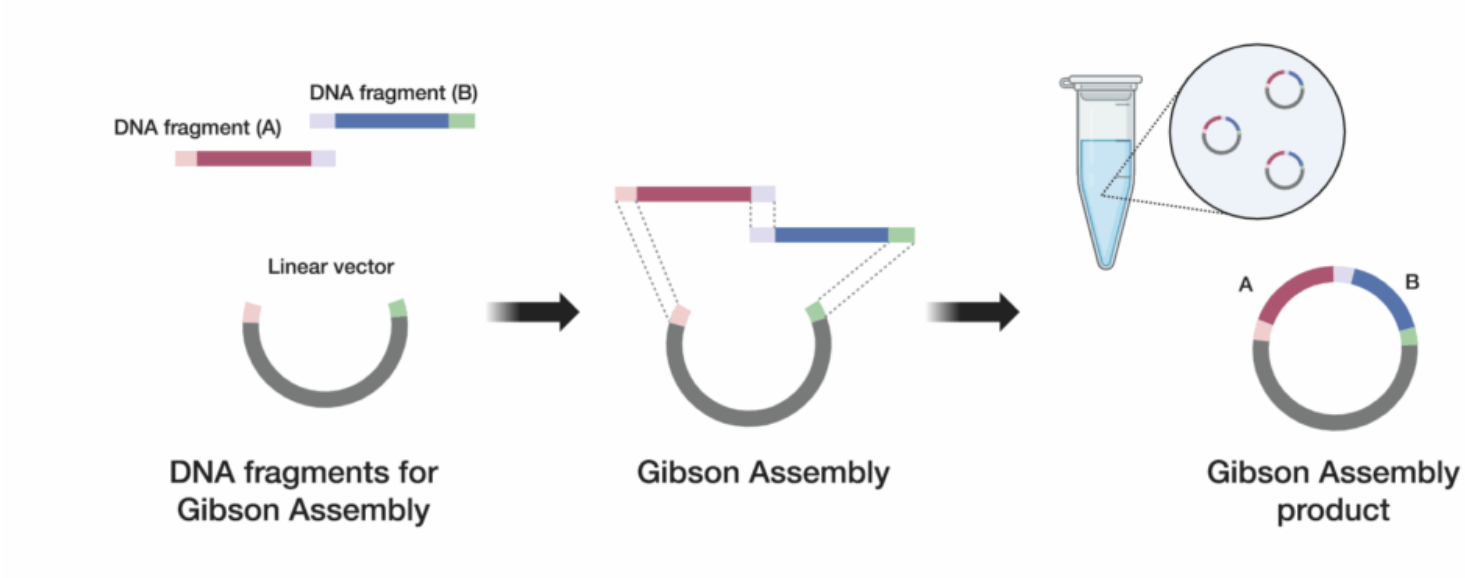
- Exonuclease activity to create single-stranded overlaps
- DNA polymerase to fill gaps
- DNA ligase to seal nicks

Advantages:

- Seamless cloning
- No need for restriction sites at junctions
- Can assemble multiple fragments
- Useful for complex constructs

Design requirements:

- Overlapping sequences between fragments
- Correct fragment order
- Accurate primer design
- High-quality PCR products



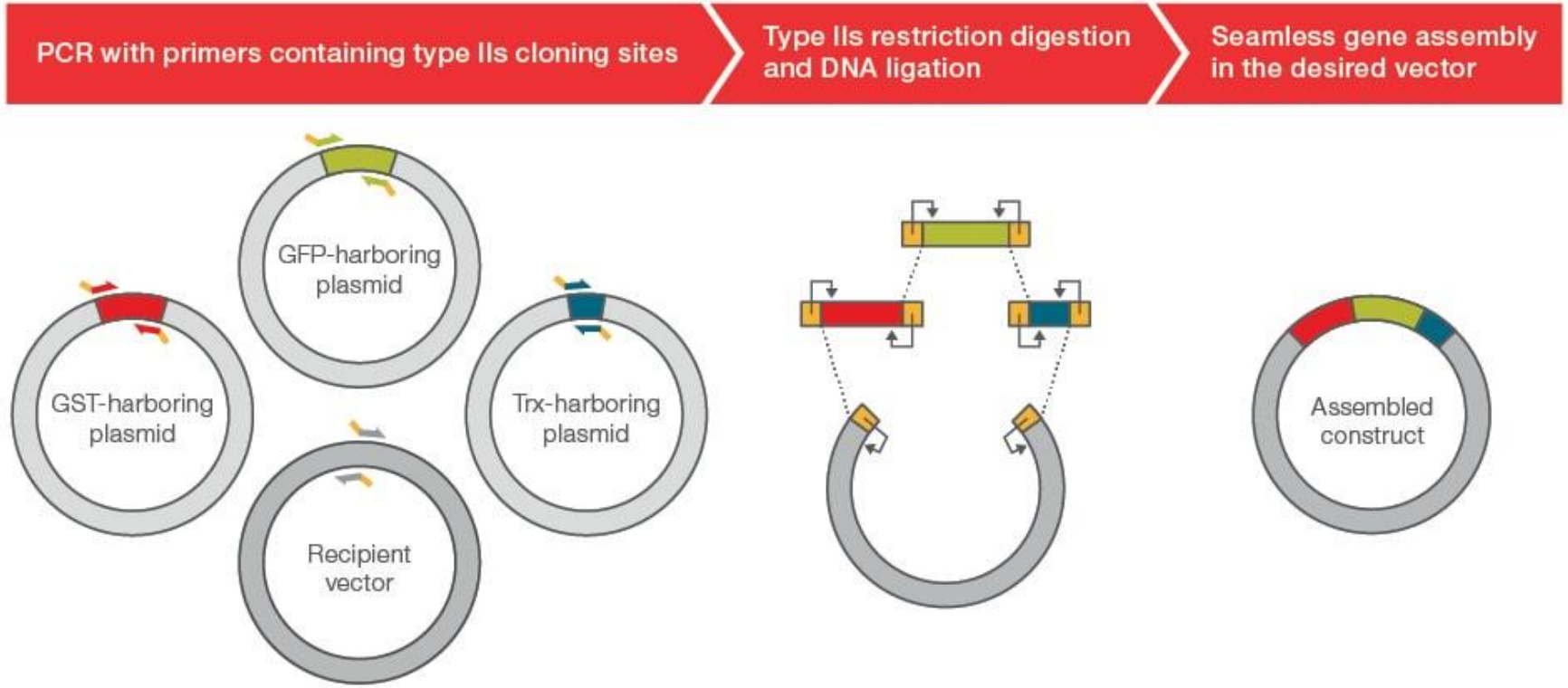
Gibson Assembly is powerful because the junctions are designed by you rather than dictated by restriction sites

Golden Gate Cloning

- Golden Gate cloning uses Type IIS restriction enzymes, which cut outside their recognition sequence.
- This allows researchers to design custom overhangs that guide DNA fragments into a defined order.
- Advantages:
  - Directional assembly
  - Multi-fragment cloning
  - Scarless or near-scarless junctions
  - Useful for modular cloning systems

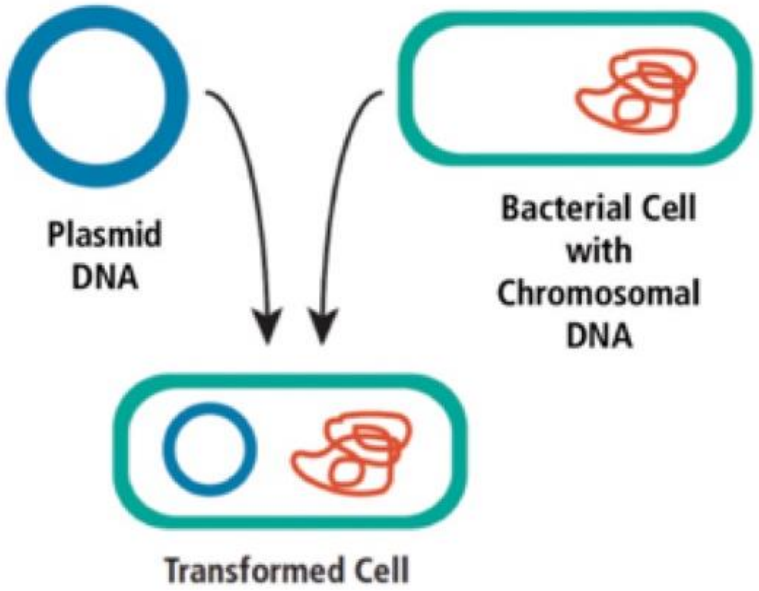
Examples of Type IIS enzymes include:

- BsaI
- BsmBI
- BbsI



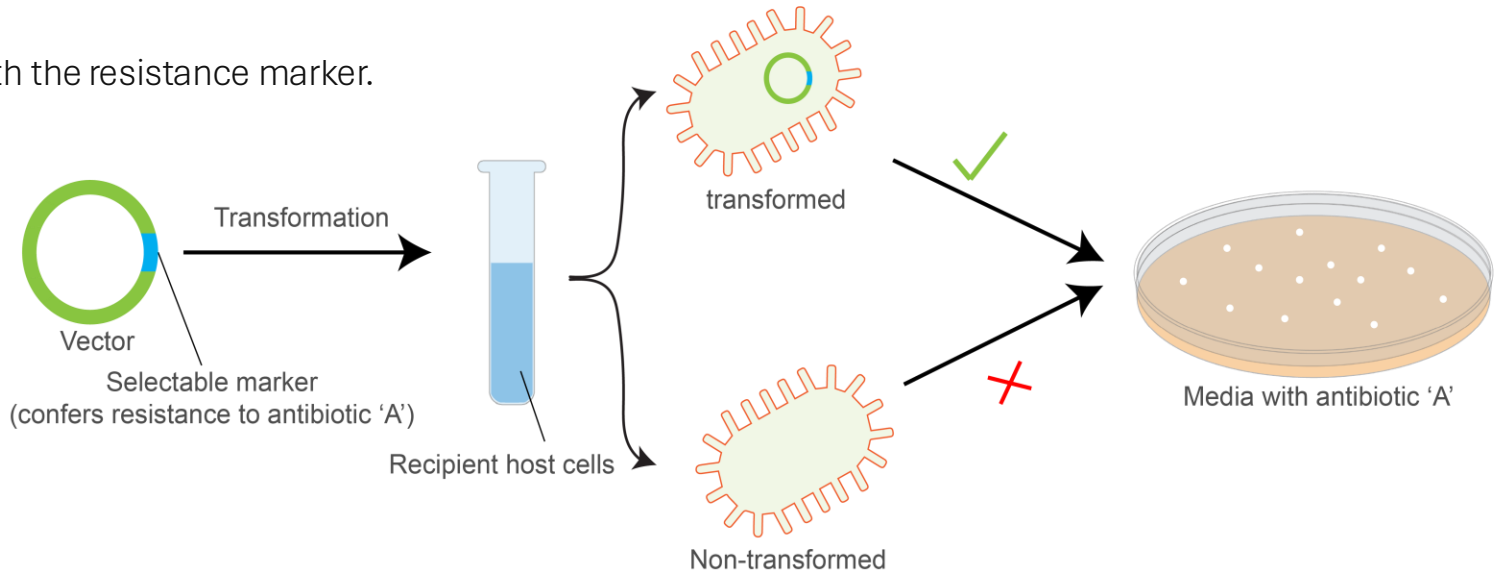
Transformation

- After the plasmid is assembled, it must be introduced into bacterial cells.
- This process is called transformation.
- Common methods are:
  - Chemical transformation
  - Electroporation
- The bacteria are usually plated on agar containing an antibiotic.
- Only bacteria that have taken up a plasmid carrying the resistance gene should grow.
- Common host strains include laboratory strains of *E. coli*, such as DH5α or TOP10.



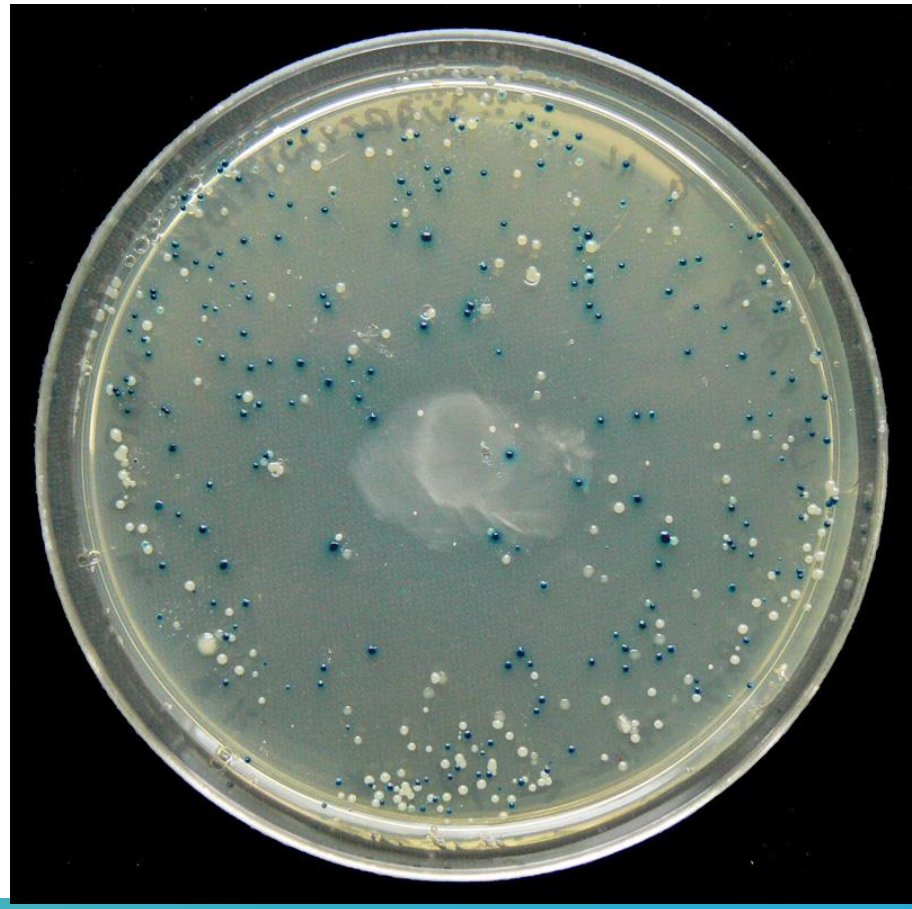
Antibiotic Selection

- Plasmid vectors carry an antibiotic resistance gene.
- This allows selection of bacteria that contain the plasmid.
- Common antibiotic resistance markers include:
  - Ampicillin resistance
  - Kanamycin resistance
  - Chloramphenicol resistance
  - Spectinomycin resistance
- Selection does not prove that the insert is correct.
- It only indicates that bacteria likely carry a plasmid with the resistance marker.



Colony Screening

- After transformation, individual colonies must be screened.
- Common screening methods are:
  - Colony PCR
  - Restriction digest screening
  - Blue-white screening
  - Diagnostic PCR
  - Small-scale plasmid preparation followed by analysis
  - Sanger sequencing
- Colony PCR can quickly indicate whether an insert of the expected size is present.
- Restriction digest can test whether the plasmid has the expected pattern.
- Sequencing is required to confirm the exact DNA sequence.



Colony PCR

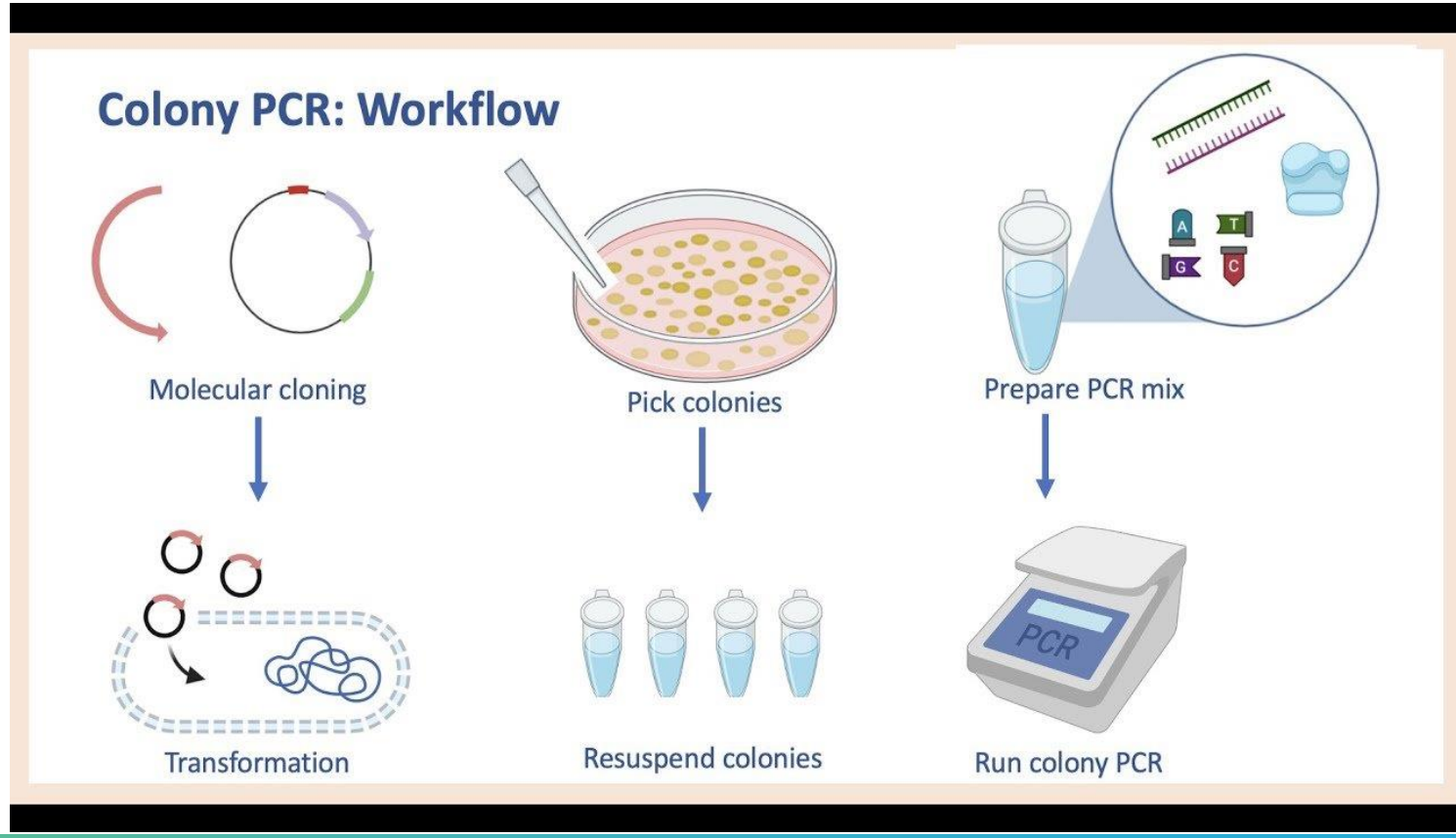
- Colony PCR allows rapid screening of bacterial colonies without first purifying plasmid DNA.
- A small amount of bacterial material is used as PCR template.
- Primer combinations may include:
  - Vector forward + vector reverse
  - Vector primer + insert primer
  - Insert-specific forward + insert-specific reverse

Colony PCR can test:

  - Whether an insert is present
  - Approximate insert size
  - Sometimes insert orientation

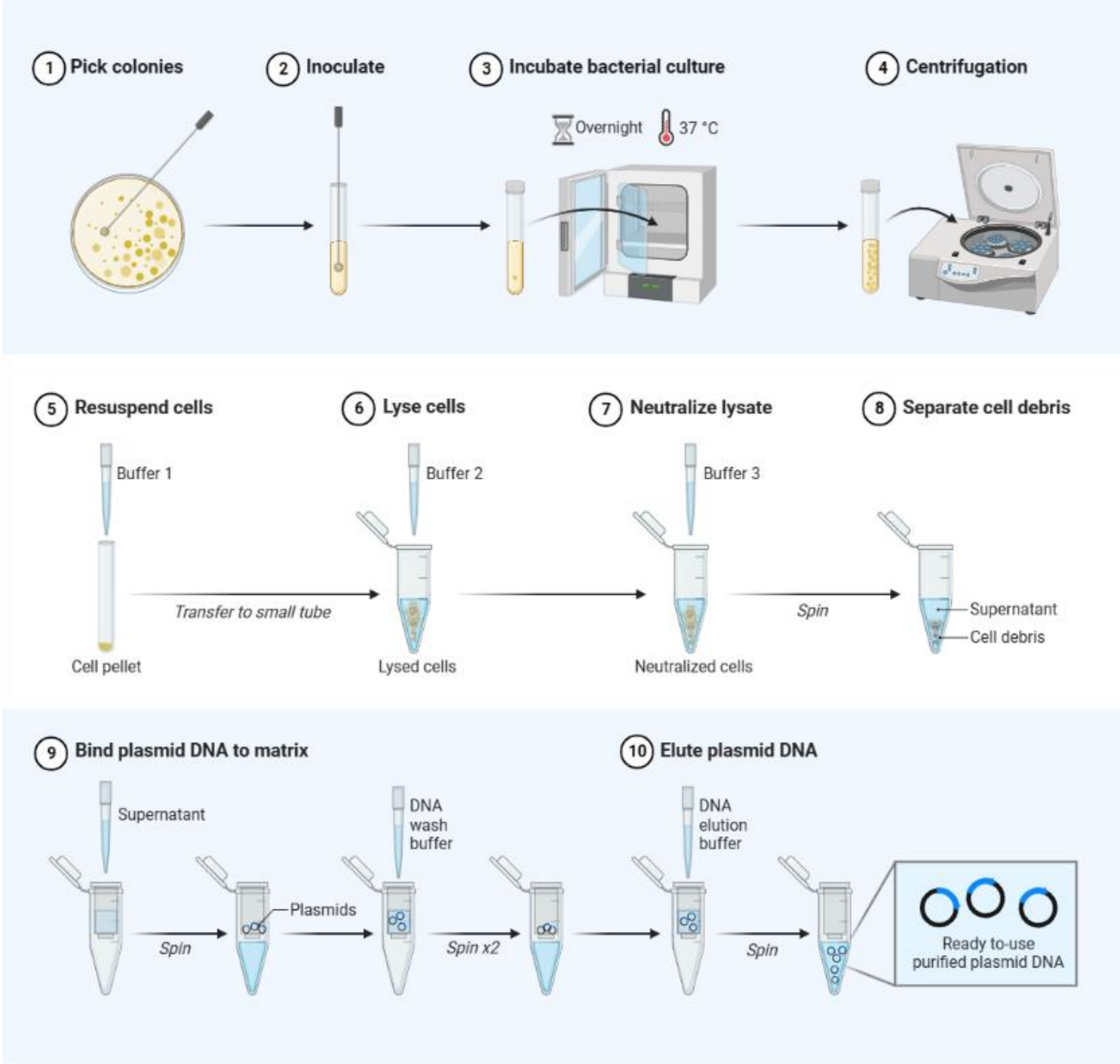
Limitations:

  - False positives can occur
  - PCR failure does not always mean the clone is wrong
  - It does not confirm exact sequence



Plasmid Miniprep

- A miniprep is a small-scale plasmid DNA purification from bacterial culture.
- The basic logic is:
  - Grow a bacterial colony in liquid culture
  - Lyse the bacterial cells
  - Remove proteins, genomic DNA, and debris
  - Bind plasmid DNA to a purification column or resin
  - Wash contaminants away
  - Elute purified plasmid DNA
- The purified plasmid can then be used for:
  - Restriction digest
  - PCR
  - Sequencing
  - Transfection
  - Further cloning

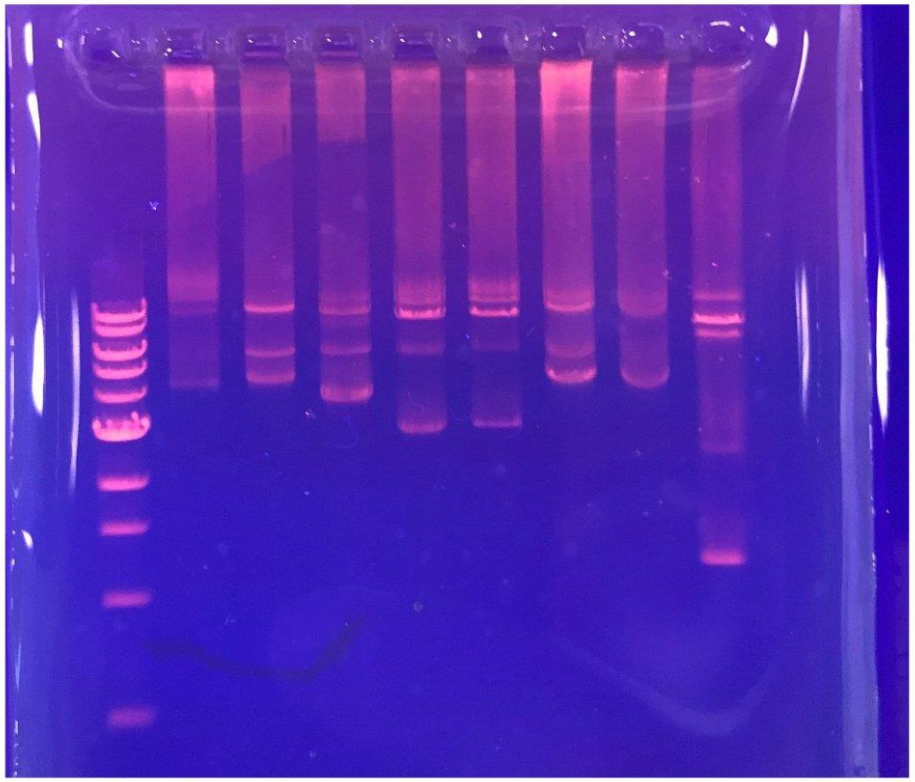


Checking the Plasmid by Restriction Digest

Restriction digest after miniprep checks whether the miniprep DNA gives the band pattern expected for the cloned plasmid.

Before running the digest, decide exactly what the digest should prove:

- Does the plasmid release an insert of the expected size?
- Does the vector backbone have the expected size?
- Can the enzyme combination distinguish correct from wrong orientation?
- Are there extra bands suggesting partial digestion, rearrangement, or contamination?
- Do the observed bands match the predicted map?
  - Digest 200 - 500 ng of miniprep DNA
  - Use one or two enzymes chosen from the plasmid map
  - Run alongside an undigested plasmid control
  - Compare bands with the predicted fragment sizes
  - Send promising clones for Sanger sequencing

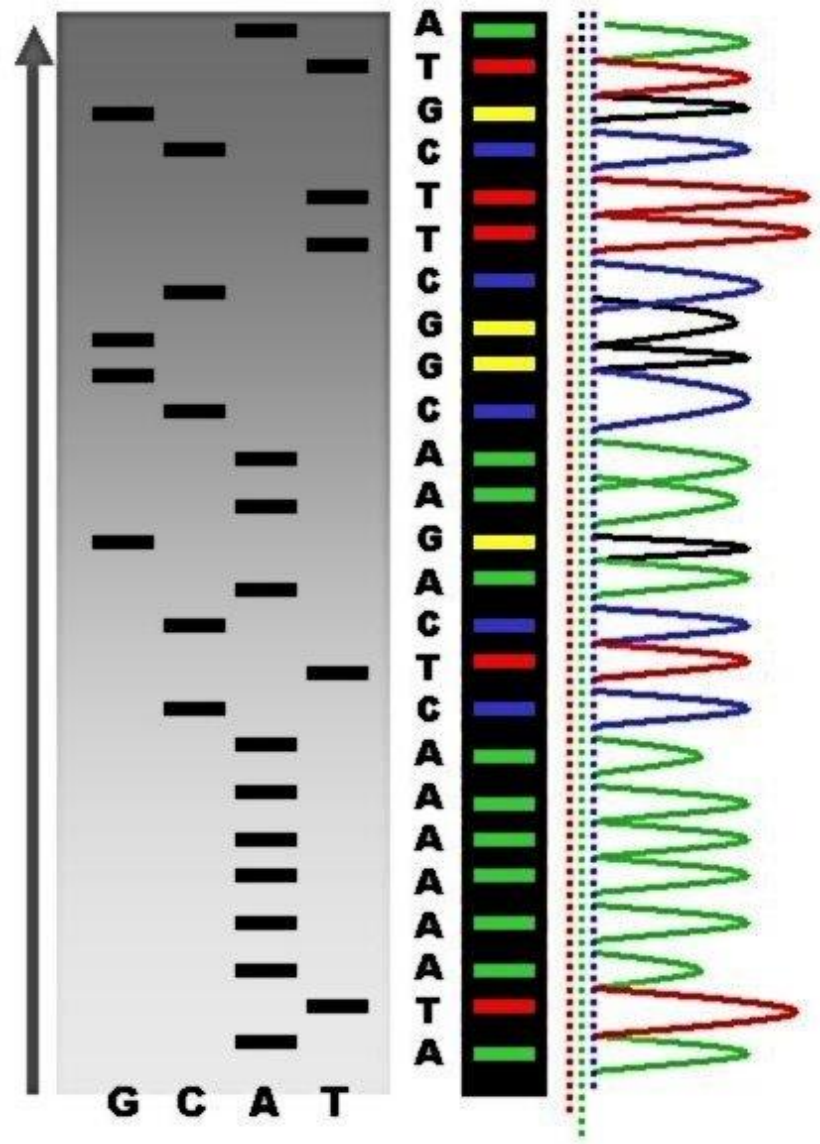


Sanger Sequencing: Final Construct Verification

- Sanger sequencing is the standard method for confirming cloned DNA sequences.
- It is used to check:
  - Insert sequence
  - Junctions between insert and vector
  - Reading frame
  - Mutations introduced during PCR
  - Correct orientation
  - Presence or absence of stop codons
  - Integrity of tags or linkers

For expression constructs, sequencing should cover:

  - The full insert
  - Both cloning junctions
  - Any engineered mutation
  - Any fusion boundary



Controls in Molecular Cloning

- Controls help distinguish between technical failure and biological absence.
- Useful controls include:
  - Undigested plasmid control
  - Digested vector-only ligation control
  - Insert-only control where appropriate
  - Positive control plasmid for transformation
  - No-template PCR control
  - Known positive PCR control
  - Empty vector control
  - Sequencing control or reference sequence

Controls answer specific questions:

- Are the competent cells working?
- Is the antibiotic plate working?
- Is the PCR contaminated?
- Is the vector self-ligating?
- Is the insert being amplified correctly?

Common Reasons Molecular Cloning Fails

- Poor primer design
- Incorrect restriction sites
- Internal restriction sites in the insert
- Incomplete vector digestion
- Vector self-ligation
- Low-quality PCR product
- Incorrect insert-to-vector ratio
- Low transformation efficiency
- Toxic insert
- Recombination or plasmid instability
- Incorrect antibiotic selection
- Wrong reading frame
- Mutations introduced during PCR

Troubleshooting Strategy

A rational troubleshooting approach asks:

Was the insert generated correctly?

- Check PCR product size
- Optimise PCR conditions
- Use high-fidelity polymerase

Was the vector prepared correctly?

- Confirm complete digestion
- Purify linearised vector
- Reduce self-ligation

Was the assembly successful?

- Optimise insert-vector ratio
- Check DNA concentration
- Include positive and negative controls

Was transformation efficient?

- Test competent cells
- Use fresh antibiotic plates
- Include control plasmid

Were colonies screened correctly?

- Use appropriate primers
- Confirm by sequencing

Summary

- Molecular cloning is a controlled way of building DNA constructs for biological experiments.
- The essential principles are:
  - Define the biological question
  - Choose the correct insert and vector
  - Design the cloning strategy carefully
  - Preserve orientation and reading frame
  - Use appropriate selection and screening
  - Confirm the final construct by sequencing
  - Maintain clear records and quality control
  - Use the verified construct to test a biological hypothesis