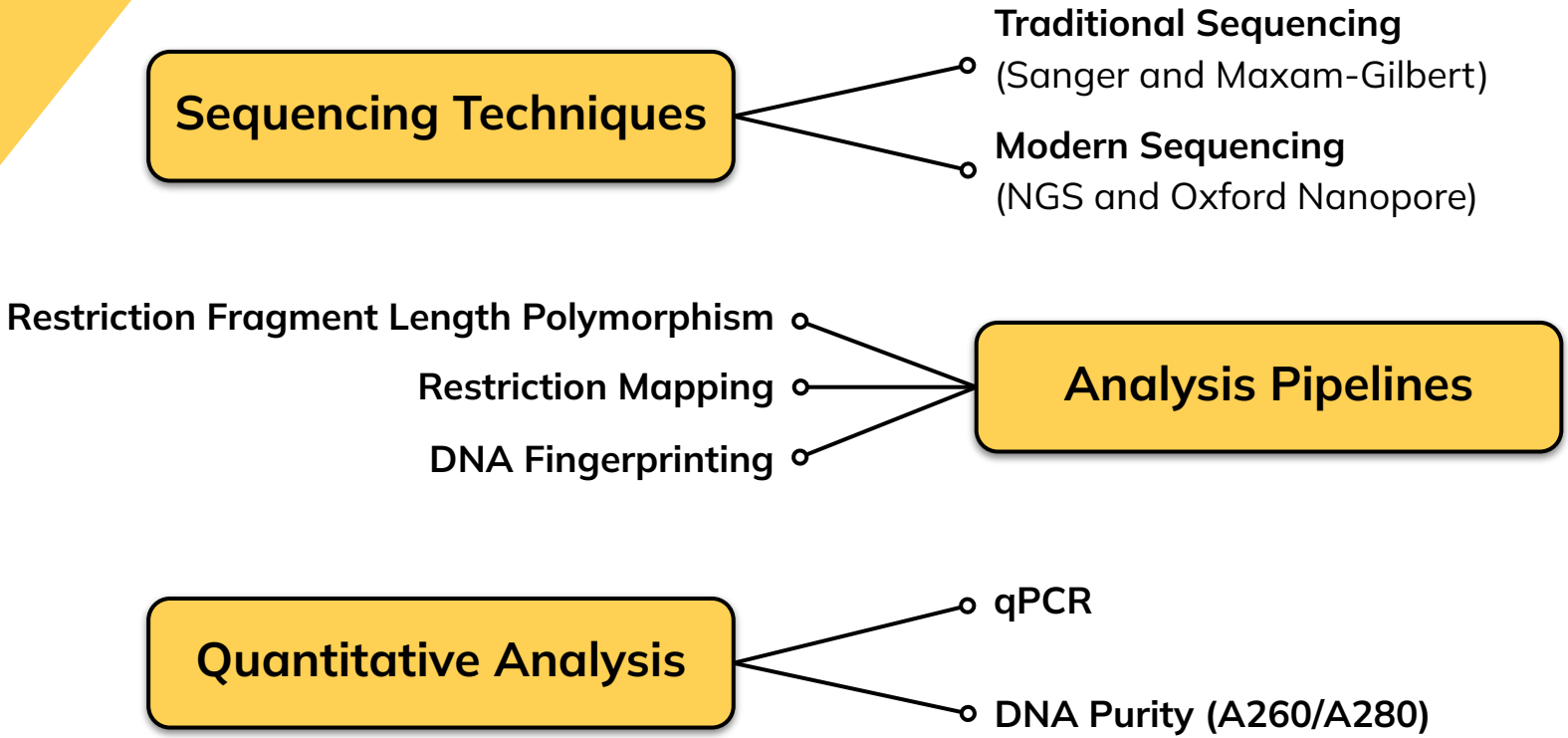
IMPORTANT INFO

After each concept is covered, a youtube link can be found in the slides that explains the concept more clearly (also in Additional Resources document)

We also STRONGLY RECOMMEND reading up / googling each concept on your own beyond the resources we provided, if you want to know their intricacies.

DNA Analysis Techniques: Lesson Overview

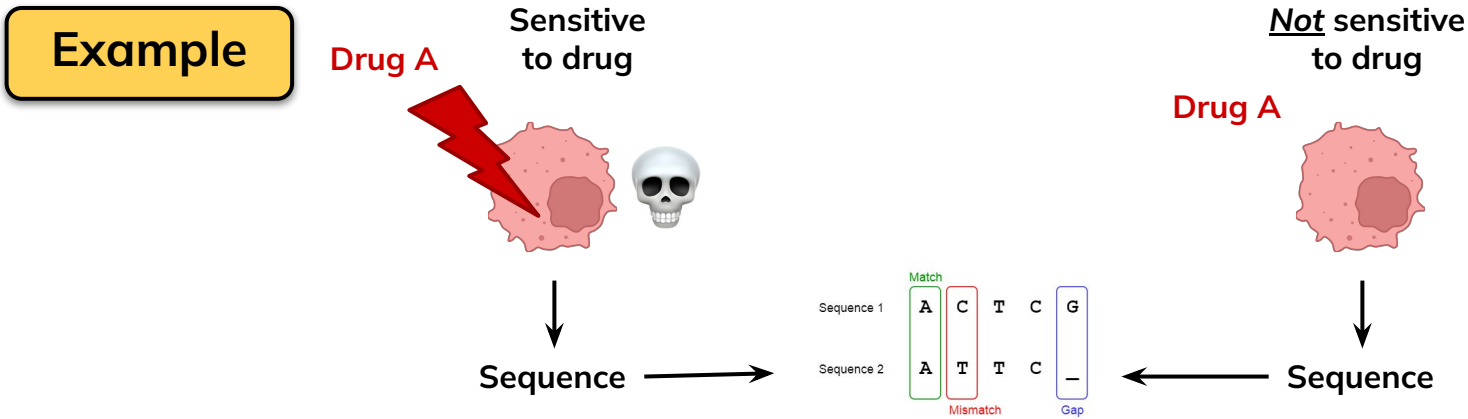


DNA Sequencing

Objective: Confirm that there are **no deleterious mutations** in our cloned insert (following the molecular cloning pipeline).

OR

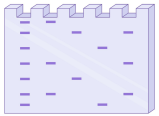
Obtain sequence from subject of interest to **analyse features**.



A Brief History of Sequencing Technologies

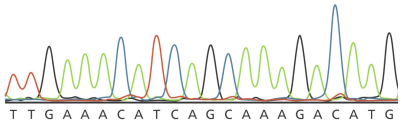
Image credit: Laura Olivares Boldú / Wellcome Connecting Science

1977



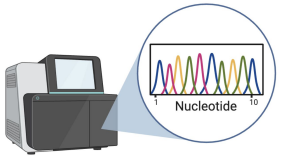
Sanger Sequencing
Manual, gel-based analysis

1990



Capillary Sequencing
Automated capillary-based sequencing with coloured bases

2000



Next-Gen Sequencing
Capillary sequencing with coloured bases

2010

Oxford NANOPORE Technologies™

Third-Gen Sequencing
Capillary sequencing with coloured bases

Information adapted from https://www.yourgenome.org/theme/timeline-the-past-present-and-future-of-sequencing-technologies/

Comparing Sequencing Methods

	1970s	1990s	2000s	2010s	2020s
Generation	1st generation with gel-based methods	1st generation with capillary methods	Next generation	Third generation	Future generations
Technologies	Sanger sequencing (manual)	Sanger sequencing (automated)	Illumina, Roche 454, Ion Torrent	PacBio SMRT, ONT	Genapsys, MGI
Breakthrough	Gel-based analysis	Capillary analysis	High throughput	Long reads	
Pros	Accuracy	Higher scale, Lower cost, Accuracy	Long reads, Ultra-fast, Portability		Aiming for cost effectiveness, accuracy
Cons	Labour intensive, High cost	Short reads make analysis more difficult		A reduction in accuracy	

Table adapted from https://www.yourgenome.org/theme/timeline-the-past-present-and-future-of-sequencing-technologies/

Question Walkthrough

Frederick Sanger (1918-2013) invented protein, RNA and DNA sequencing, and Sir Shankar Balasubramanian (1966-present) invented high-throughput DNA sequencing. The National Health Service is sequencing an unprecedented 100 000 genomes from rare-disease patients, but different sequencing technologies have different merits for this purpose, as described below.

Technology	Maximum length of sequence fragments which can be read	Error rate	Total number of bases sequenced per sample per day
Sanger sequencing	900 bp	1 in 1000 bp	900 bp (1 fragment)
Illumina machines	200 bp	1 in 100 bp	$3x10^{11}$ bp (>$1.5x10^{9}$ fragments)
PacificBiosciences machines	5000 bp	1 in 10 bp	$4x10^{8}$ bp (>80 000 fragments)

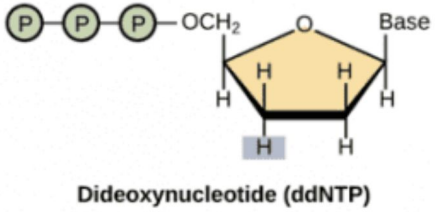
	True	False
Illumina technology is best for finding new Single Nucleotide Variations (mutations to a single base) in the patient genomes.		
PacificBiosciences technology is best for assessing transcriptional changes by RNA sequencing.		
PacificBiosciences technology is best for finding rearrangements of chunks of DNA in the patient genomes.		
Sanger sequencing is best for validating sequencing results before using patients' genetic information to guide clinical interventions.		

Images adapted from IBO 2020 Theory 1: "Sequencing" Problem

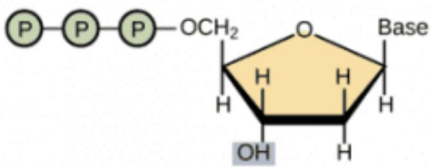
Sanger Sequencing: Chain-termination PCR Principles

Mostly similar to normal **PCR**, except for a few innovations:

- Only **one primer** is used → ensures that sequencing results are read in one direction.
- Inclusion of **ddNTPs** → creates fragments of variable lengths with coloured bases at the 3’ end.



Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Image credit: OpenStax Biology

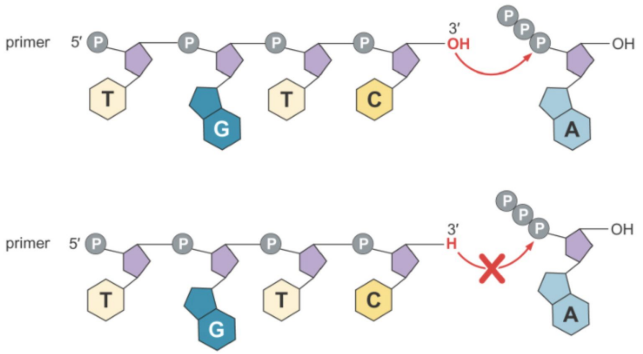


Image credit: https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/

Sanger Sequencing: Chain-termination PCR Principles

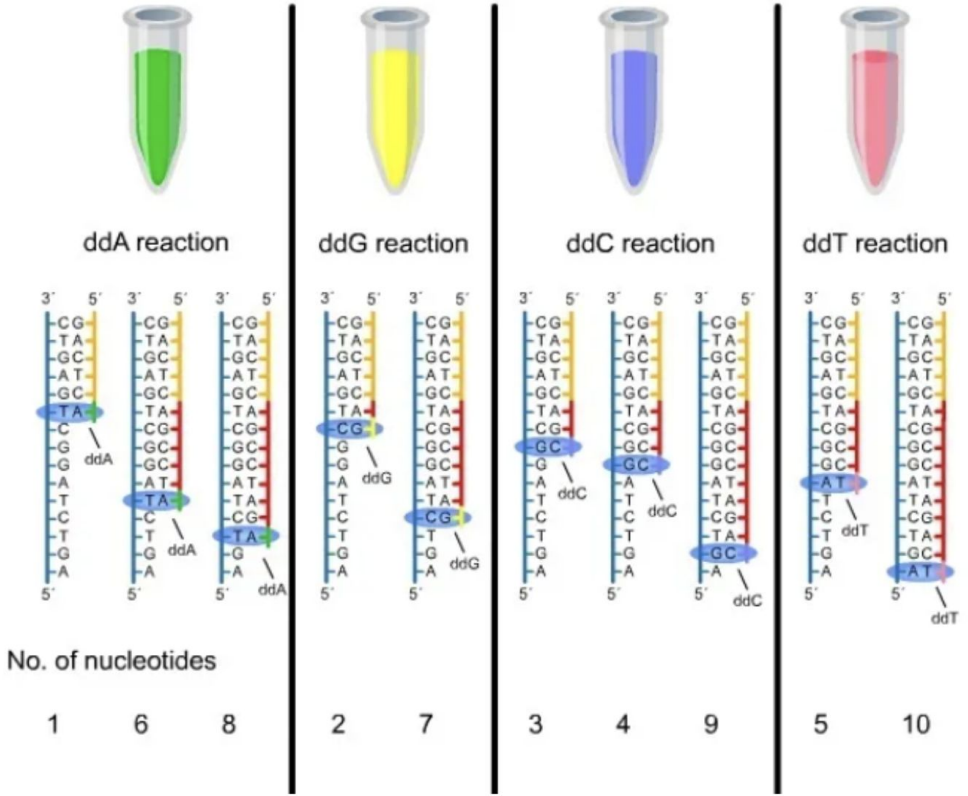


Image credit: https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/

Sanger Sequencing Workflow

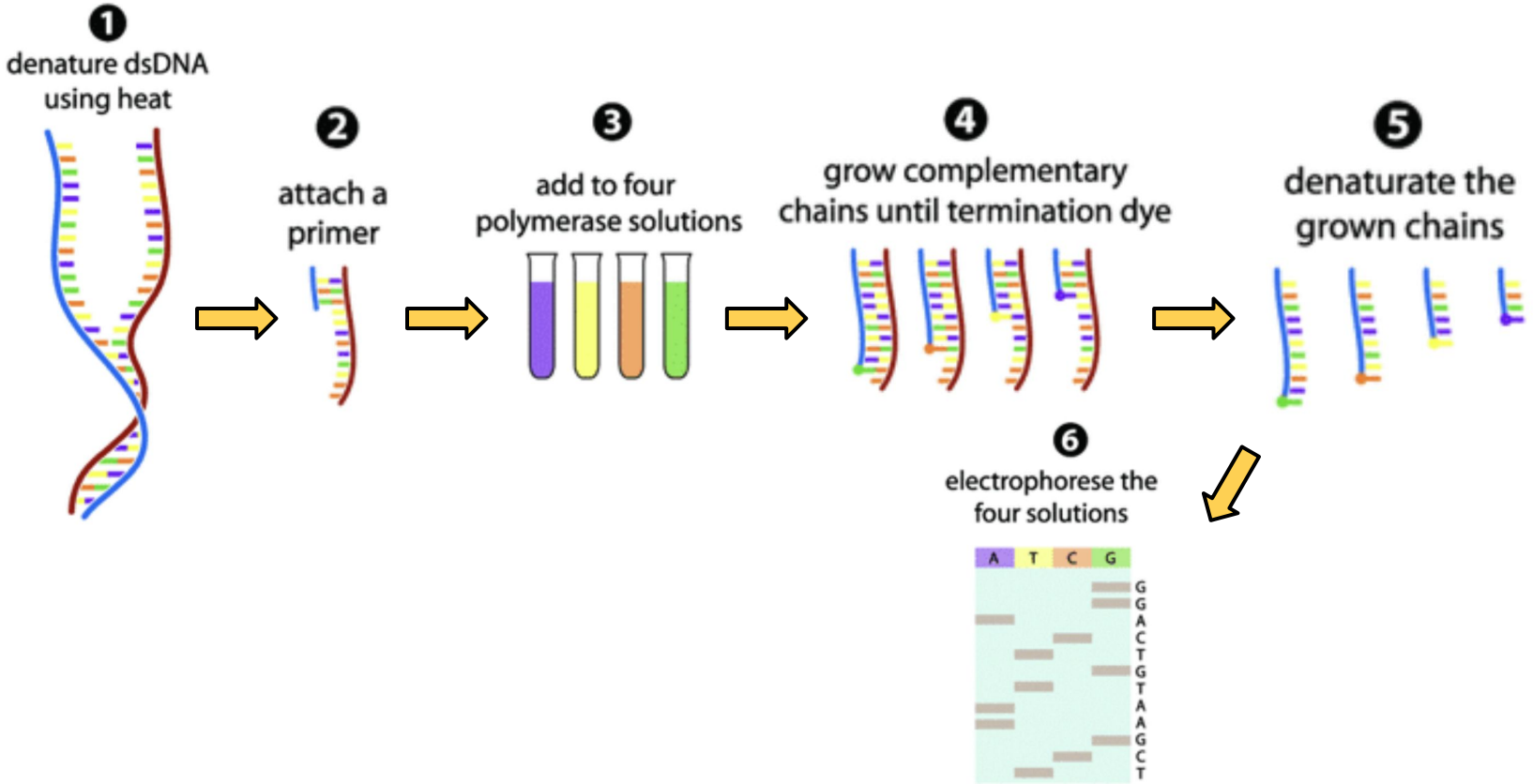


Image credit: https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/

Reading Sequencing Results from a Gel

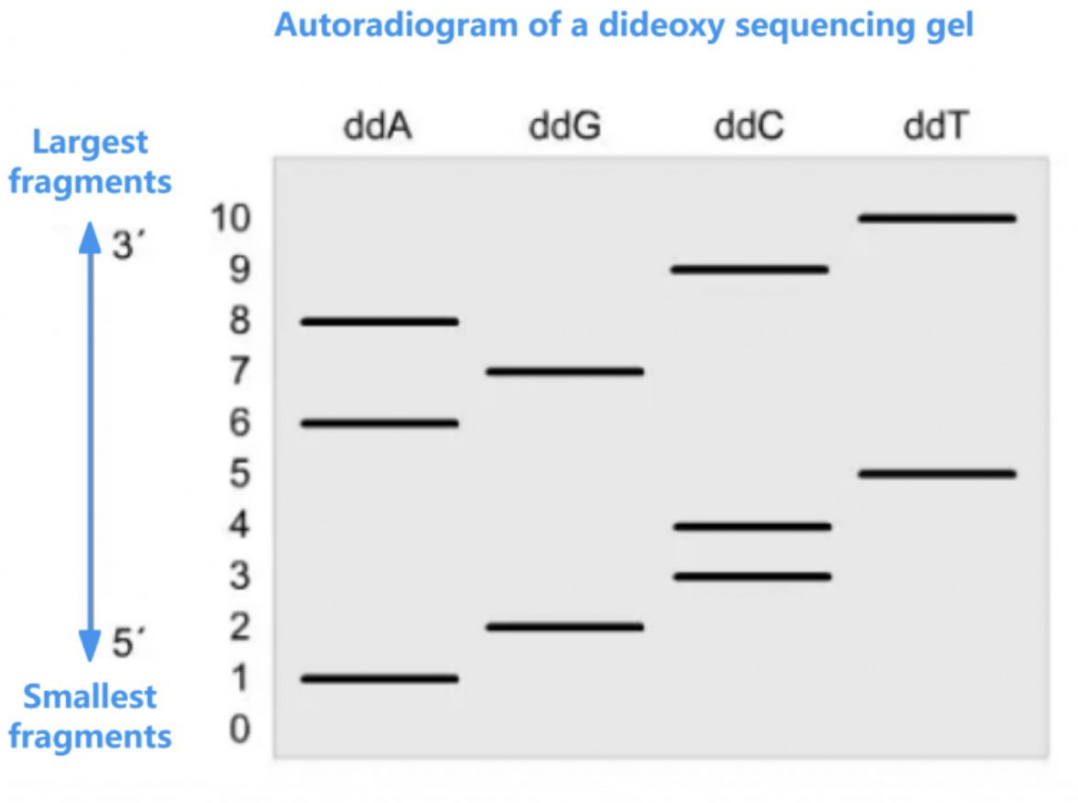


Image credit: Watson et al., MBG, 2008

Useful Videos for Sequencing

Sanger Sequencing Principles: https://youtu.be/KTstRrDTmWI

Capillary Electrophoresis: https://youtu.be/K_YtRgnM4nQ

History of Sanger Sequencing: https://youtu.be/X9566yI2cBo

Maxam-Gilbert Sequencing

- **5’ Radiolabeling**
- **Chemical cleavage in four separate reactions:**
  - **A+G**: Formic acid
  - **G**: Dimethyl sulfate
  - **C+T**: Hydrazine
  - **C**: Hydrazine + Piperidine
- **Autoradiography**

Video Explanation
https://youtu.be/cl2s-ZMmcbc

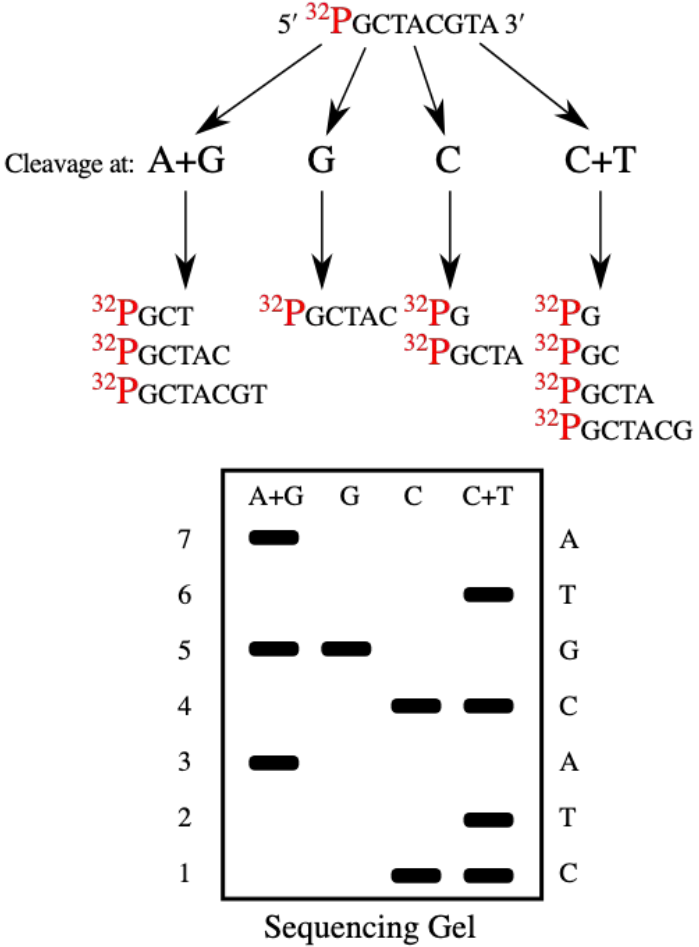


Image credit: Wikimedia Commons, Maxam-Gilbert Sequencing

Gel from M-G Sequencing

Sequence:

ATAGATGAGGCCTCG

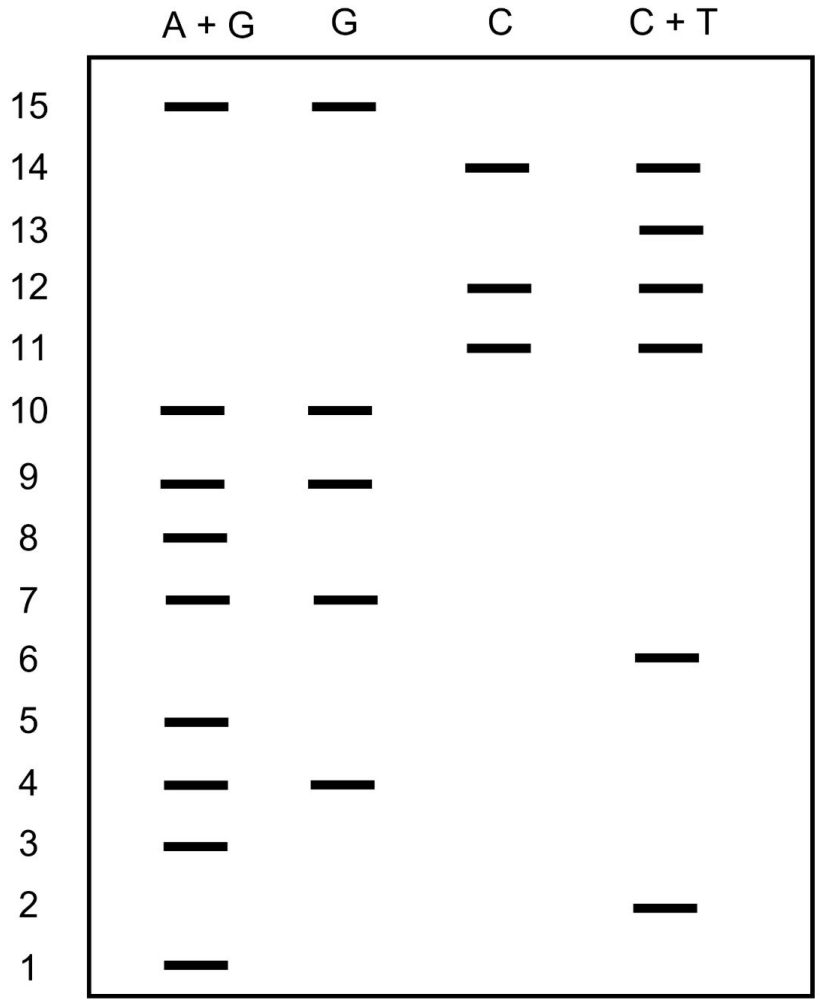


Image credit: https://bitesizebio.com/36696/maxam-gilbert-sequencing/

Next-Generation Sequencing

- **Library Prep**: Fragment DNA and ligate adapters
- **Bridge Amplification**: Create clusters of identical fragments
- **Sequencing**: Live sequencing using labelled nucleotides
- **Data Analysis**: Generate contiguous sequence

Video Explanation
https://youtu.be/CZeN-IgjYCo

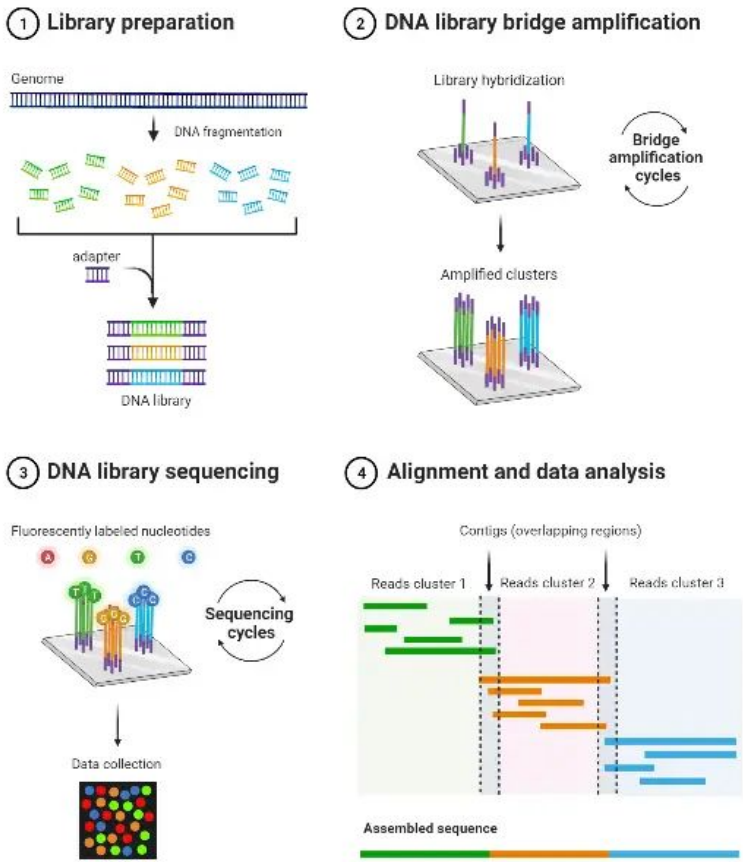


Image credit: https://microbeonline.com/next-generation-sequencing/

Nanopore Sequencing (3rd Gen)

Video Explanation
https://youtu.be/CyJWeVMFPSU

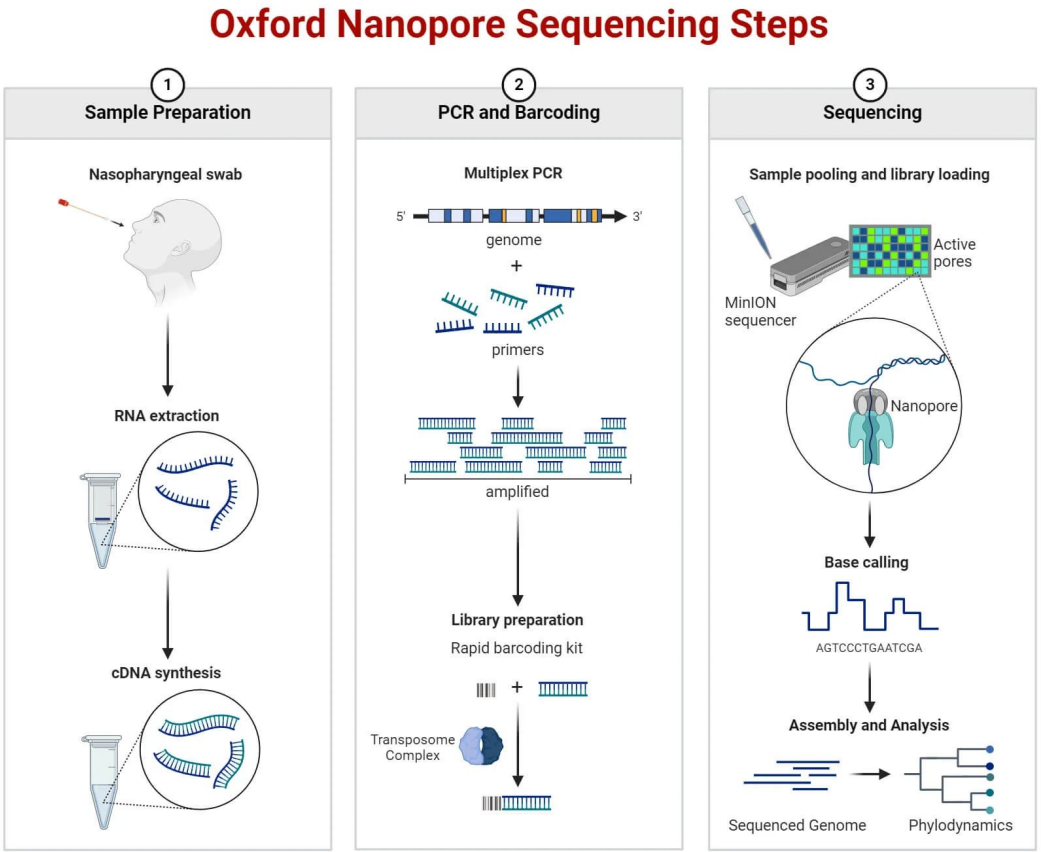


Image credit: https://microbenotes.com/oxford-nanopore-sequencing/

DNA Sequence Analysis

- Mostly **following the question** and relatively uncomplicated.

- Common **<u>concepts</u>** used:
  - **Feature** Analysis of DNA Sequences
  - Types of **mutations** (Prerequisite Info)
  - Inferring **effects on protein function** after **identifying mutations**, which may require knowledge of **amino acid properties** (Prerequisite Info – sometimes given in a graph)

Types of DNA Mutations and Their Effects

	Point Mutations (A single nucleotide mutating into another) in a Gene		
	Silent Mutation	**Missense Mutation**	**Nonsense Mutation**
A.a. sequence	No change in amino acid.	Amino acid change.	Premature stop codon.
Example	GCC (Ala) → GCT (Ala)	TGC (Cys) → GGC (Arg)	TGC (Cys) → TGA (stop)
Effects on protein f(x)	**No** effects – however, may affect codon usage (Silent mutations are also called **synonymous mutations**.)	Depends on **which residue** was mutated, and whether mutation is **conservative** (similar a.a. properties) or **nonconservative**.	Usually **deleterious** since truncated proteins are synthesised and degraded via nonsense-mediated decay (NMD).

Types of DNA Mutations and Their Effects

	Insertion / Deletion (Indel) Mutations in a Gene	
	Frameshift Mutation	**In-frame Mutation**
Type	**Non-multiple of 3** indel → reading frame changes and all amino acids downstream of indel site affected	**Multiple of 3** indel → no change in reading frame
Example	**Insertion:** ATG GCT ACG ... (Met-Ala-Thr...) → ATG GCG TAC G.. (Met-Ala-Tyr...	**Deletion:** ATG GCT ACG ... (Met-Ala-Thr...) → ATG ACG ... (Met-Thr-...)
Effects on protein f(x)	**Deleterious** if it happens relatively early on in sequence, also usually results in premature stop codon	Usually **less harmful than frameshift**, depends on where the indel occurs.
Notes	Studies show that **deletion** mutations tend to be, on average, **more deleterious** than insertions.	

Types of DNA Mutations and Their Effects

	Other Mutations	
	Splice Site Mutations	**Regulatory Sequence Mutation**
Type	Mutation in the **GT+AG** splice sites flanking introns (GT-intron-AG) **Watch here for explanation:** https://youtu.be/DJUQwuwFT5A	Mutation in **promoter/enhancer/silencer** → affects transcriptional regulation
Example	Exon 1 – **GT-Intron-AG** – Exon 2 → Exon 1 – **GA-Intron-AG** – Exon 2 **Outcome**: Intron inclusion in mRNA	**TATA Box:** TATAAA → TACAAA: Transcriptional downregulation
Effects on protein f(x)	**Incorrect intron splicing** which may introduce premature stop codons/delete exons, usually **deleterious**.	**Upregulation / downregulation** of affected protein.

Types of DNA Mutations and Their Effects

	Other Mutations	
	Transition Mutations	**Transversion Mutation**
Type	**Purine ↔ Purine / Pyrimidine ↔ Pyrimidine**	**Purine ↔ Pyrimidine**
Example	ATACG → GTACG (Purine → Purine)	ATACG → TTACG (Purine → Pyrimidine)
Effects on protein f(x)	Both can be classified as traditional point mutations. Note that **transversions** have higher likelihood of **amino acid mutations** due to the wobble base effect. **Read more here:** https://en.wikipedia.org/wiki/Transversion	

Amino Acid Properties – Conservative vs Nonconservative

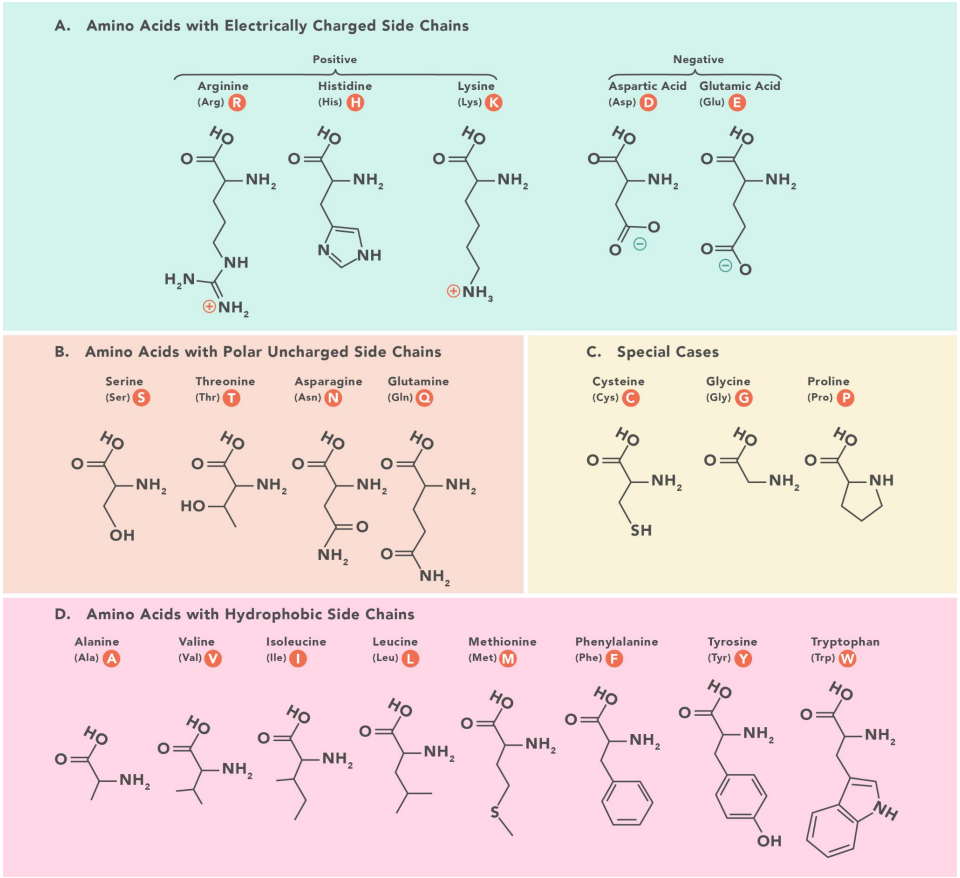


Image credit: https://www.technologynetworks.com/applied-sciences/articles/essential-amino-acids-chart-abbreviations-and-structure-324357

Amino Acid Properties – Conservative vs Nonconservative

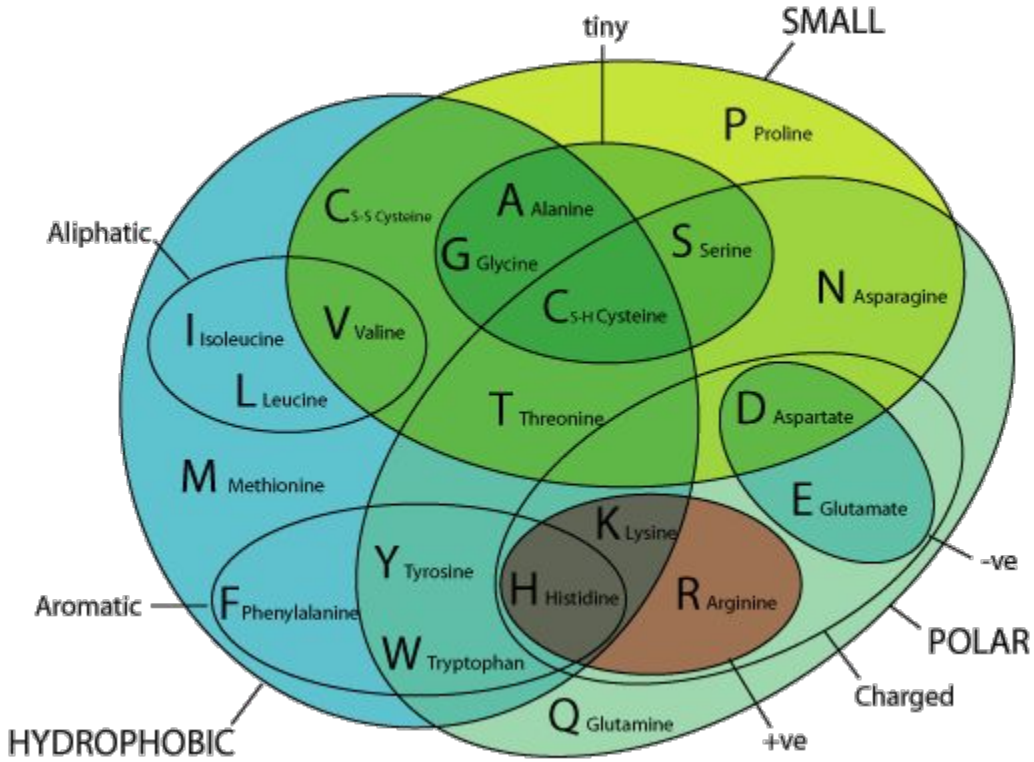


Image credit: Livingstone, C. D. and Barton, G. J. (1993)

Extracting Features and Analysing DNA Sequences

GC content will determine **melting temperature (Tm)** – temperature at which half of the DNA is single-stranded.

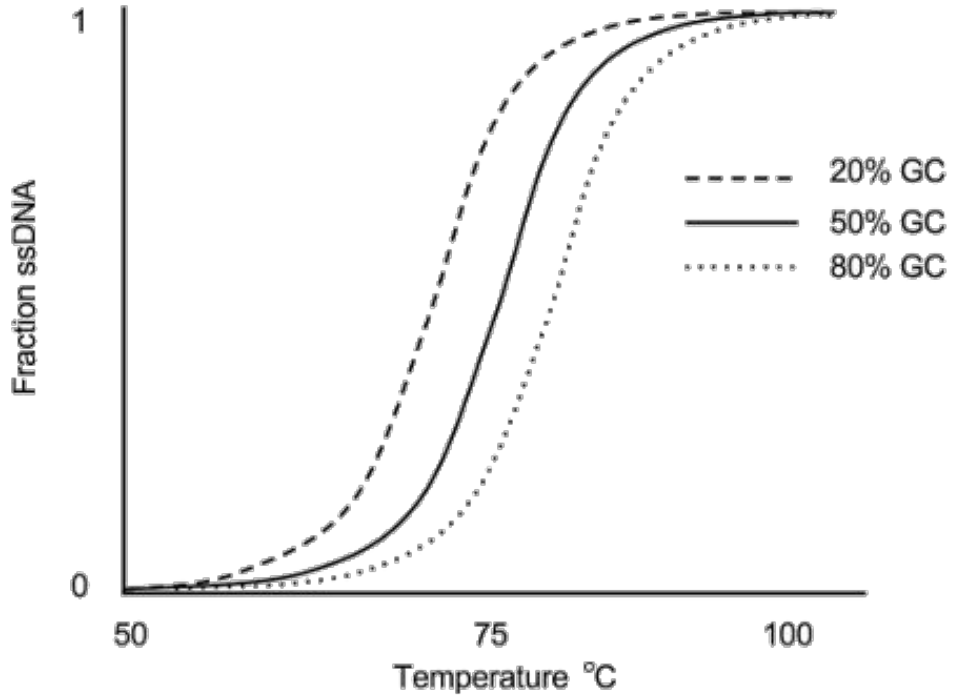


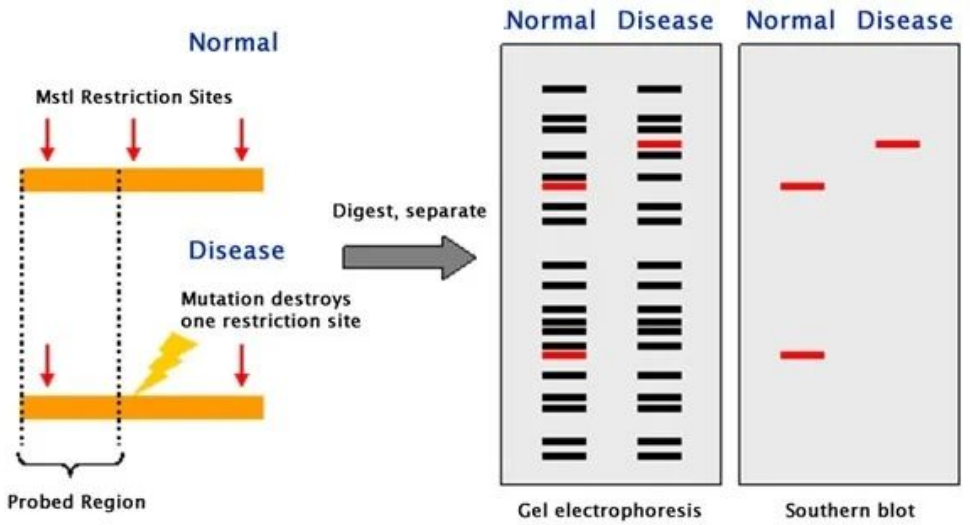
Image credit: Khan Academy

Restriction Fragment Length Polymorphism (RFLP)

Differentiation of organisms by analysing **patterns** produced through **restriction enzyme digest**.

Applications:

- Determining **disease states** of an individual
- Forensic / paternity cases to determine **DNA sample source**
- Genetic mapping



Video Explanation
https://youtu.be/swQ_pm4cR5o

Image credit: https://www.news-medical.net/life-sciences/Restriction-Fragment-Length-Polymorphism-%28RFLP%29-Technique.aspx

Restriction Mapping Pipeline

- Identifying **restriction site locations** on an unknown segment of DNA
- Usually involves **three-way digestion** in separate reactions of the DNA fragment:
  1. RE1 (e.g. SalI)
  2. RE2 (e.g. BamHI)
  3. RE1 + RE2 (e.g. SalI + BamHI)

This produces **distinct fragments** that enable where the cut sites are.

Video Explanation
https://youtu.be/qqaIgp3iKGM

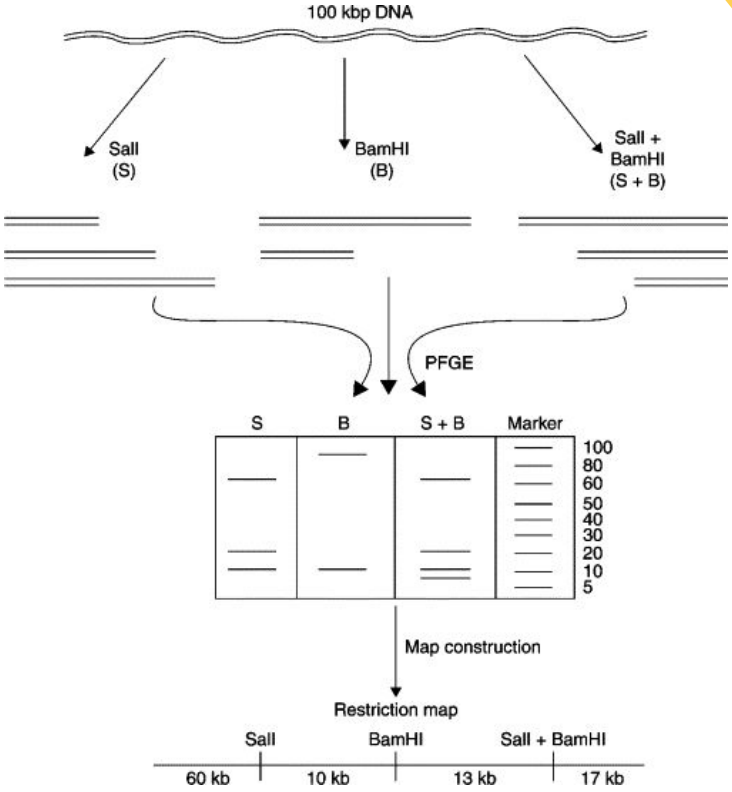


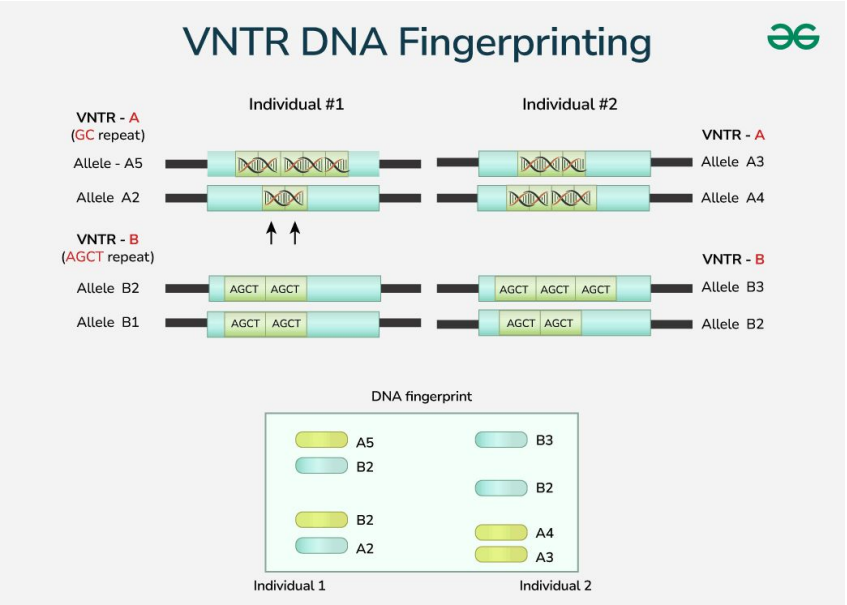
Image credit: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/restriction-map

DNA Fingerprinting

- Used in **forensics** to identify culprits based on the idea that each person has a **unique DNA fingerprint**
- DNA samples are first extracted then analysed via one of the pipelines below.

Pipelines for DNA fingerprinting:

- RFLP
- PCR → Short Tandem Repeat (STR) Analysis / Variable Nucleotide Tandem Repeat (VNTR) Analysis



Video Explanation
https://youtu.be/DbR9xMXuK7c

Image credit: https://www.geeksforgeeks.org/vntr/

Quantitative PCR (qPCR)

- **Quantify** starting amount of a specific DNA sequence
- Works through **amplification** of specific DNA sequence by PCR and **measuring DNA accumulation** during each cycle using fluorescent dyes (e.g. SYBR Green) that bind to dsDNA.

Video Explanation
https://youtu.be/iu4s3Hbc_bw

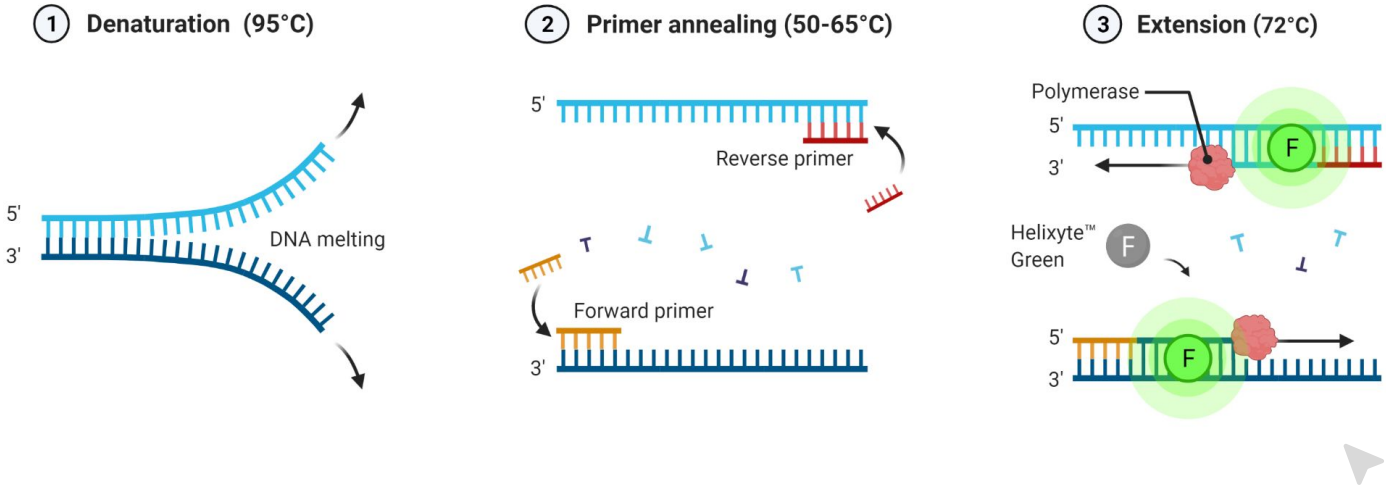
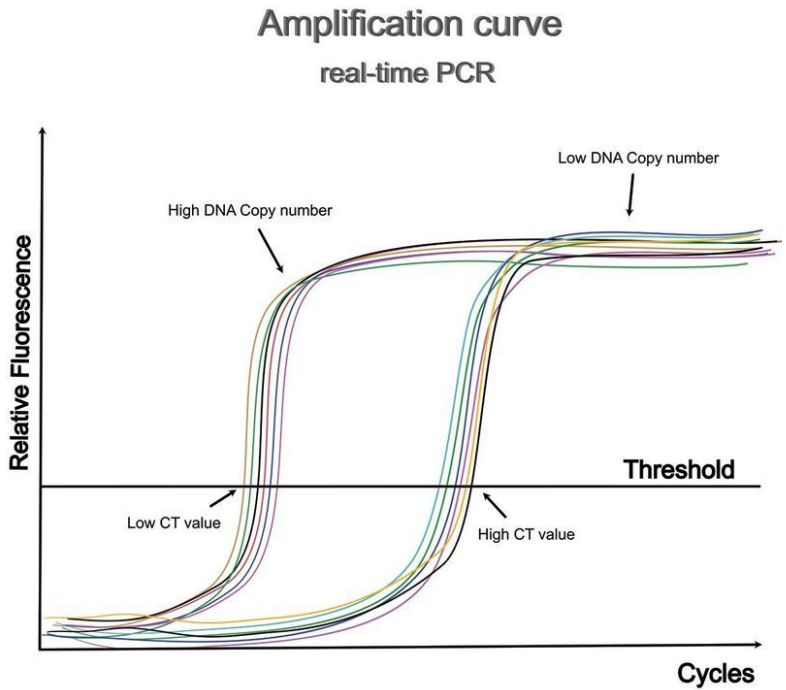


Image credit: https://www.aatbio.com/catalog/real-time-pcr-qpcr

Quantitative PCR (qPCR) Graph

- **Ct (cycle threshold) value** is the indicator for initial DNA quantity → defined as **number of cycles** required to reach **fluorescent threshold**
- **<u>High</u> Ct value → <u>less</u> starting DNA** since more amplification cycles required to reach threshold
- **<u>Low</u> Ct value → <u>more</u> starting DNA** since less amplification cycles required to reach threshold
- **Absolute** (determining exact conc.) vs **Relative** (comparing two samples w/o exact conc.) quantification depends on experimental design.



More Detailed qPCR
https://youtu.be/tH_ozcFwQ_Q

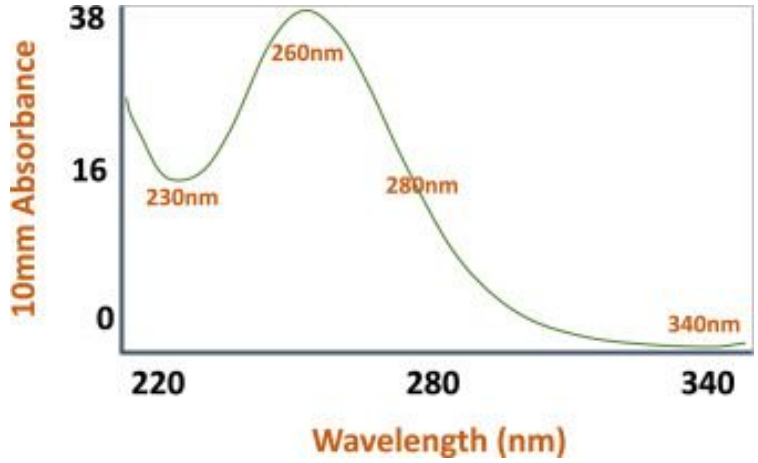
Image credit: https://www.the-scientist.com/insights-into-qpcr-protocol-detection-methods-and-analysis-71478

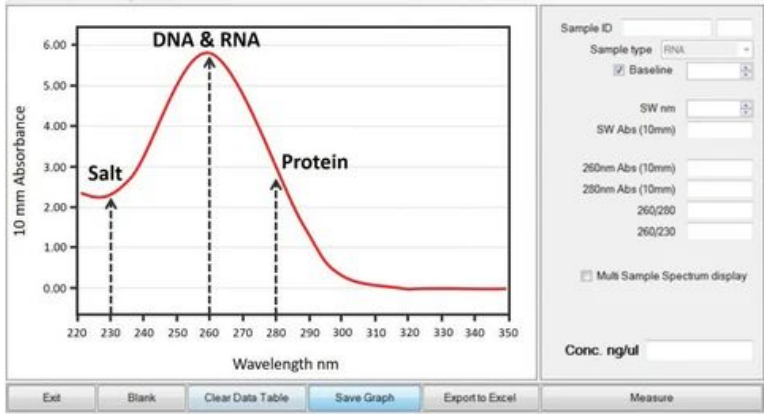
DNA Purity Ratios and Nanodrop

- Ratio of **absorbance at 260nm and 280nm (A260/A280)** gives idea of **DNA purity**.
  - A260/A280 ~ **1.8**: ‘Pure’ DNA
  - A260/A280 ~ **2.0**: ‘Pure’ RNA
- The presence of **contaminants** will **lower A260/A280 ratio** → due to protein/phenol/other contaminants that absorb strongly at 280nm
- **A260/A230** ratios may also be calculated to check for contaminants that absorb at 230nm
- Nanodrop measures DNA concentration using **A260**. Read more here: https://assets.thermofisher.com/TFS-Assets/MSD/Technical-Notes/TN52810-nucleic-acid-measurements-260-nm.pdf

Video Explanation
https://youtu.be/6i6PKyGIY3w

DNA Absorbance Graph





Thank you!