



# Bio 296: Microbial Bioinformatics

## Introduction to Sequencing Technologies and DNA extraction

Tim Roloff

# Table of contents

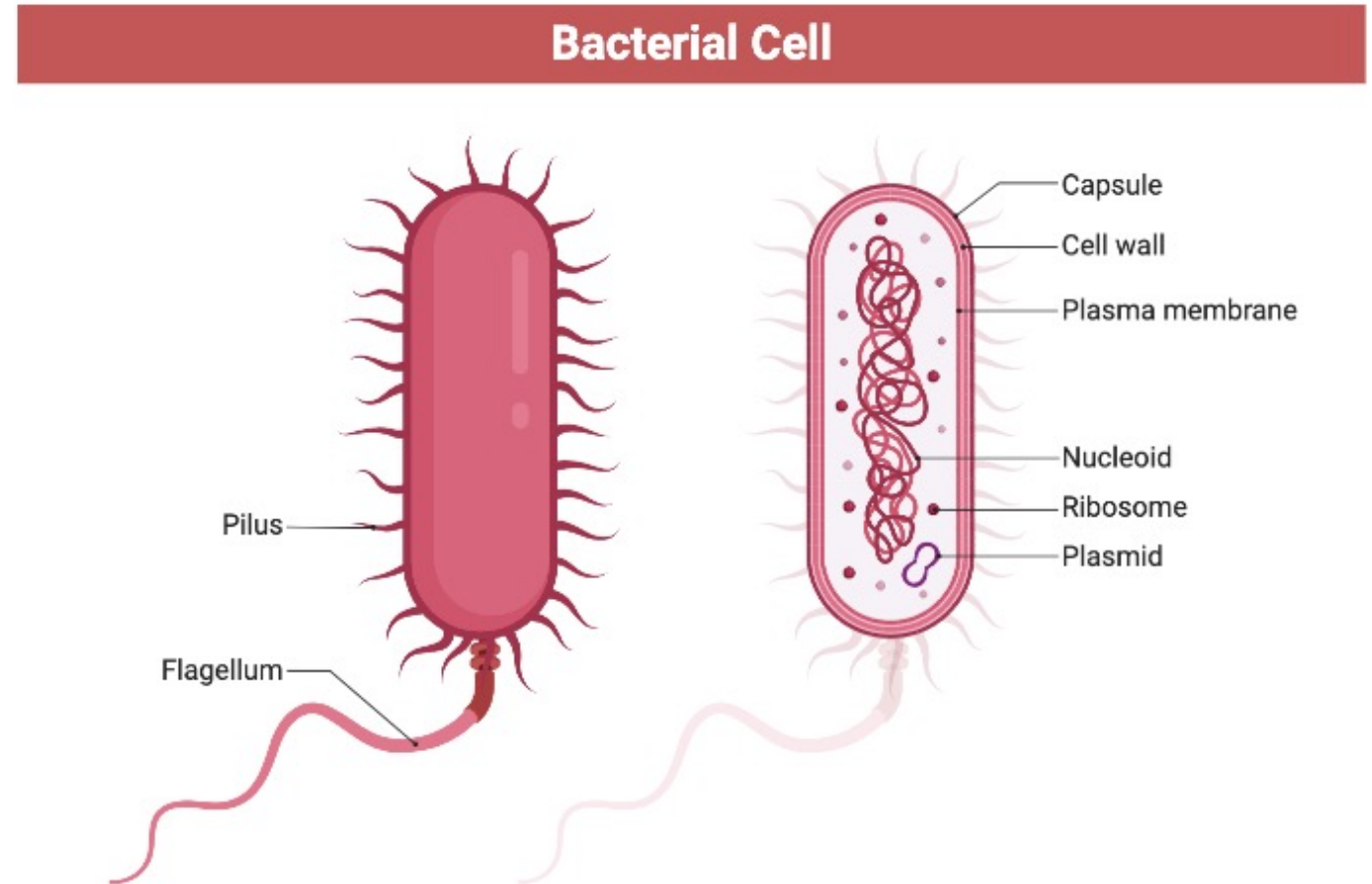


- DNA extraction
- Illumina sequencing
- Nanopore sequencing (ONT)

# DNA extraction

Purify nucleic acids from bacterial cell

- Genomic DNA
- Plasmids
- RNA

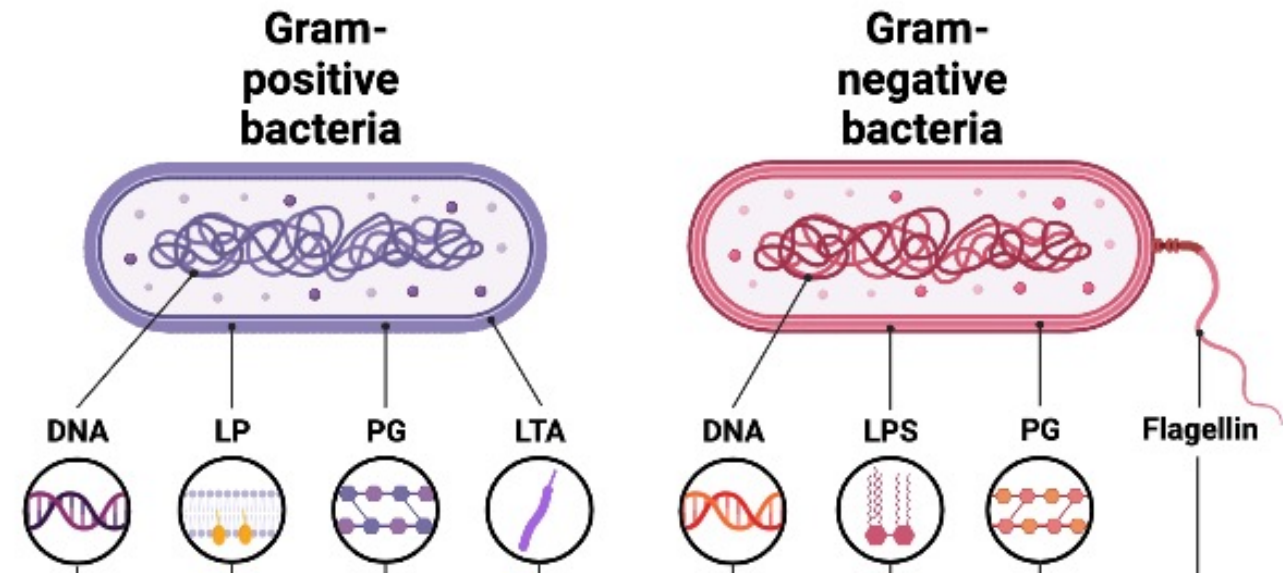


BioRender (2021). Structural Overview of a Bacterial Cell. <https://app.biorender.com/biorender-templates/figures/all/t-5ffdfb47420acf00a02c5e53-structural-overview-of-a-bacterial-cell>

# DNA extraction

- Gram positive and gram negative bacteria have different cell wall compositions
- Gram positives need harsher extraction methods
- Mechanical disruption of cell wall
  - Bead beating
- Enzymatic disruption of cell wall
  - Lysozyme (Muramidase)
  - Proteinase K
  - Lysostaphin (zink endopeptidase) for *Staphylococcus* sp.

LP - Lipopeptides  
LPS – Lipopolysaccharid  
PG – Peptidoglycan  
LTA – Lipoteichoic acid



– BioRender (2021). Recognition of Pathogen-Associated Molecular Patterns (Bacteria). <https://app.biorender.com/biorender-templates/figures/all/t-609beda351a1e400aad4193c-recognition-of-pathogen-associated-molecular-patterns-bacter>

# DNA extraction

- Manual kits or extraction robots
- Robots used at IMM
  - QIAGEN QIA Symphony
  - Promega Maxwell
  - QIAGEN QIAcube

Different robots support different protocols

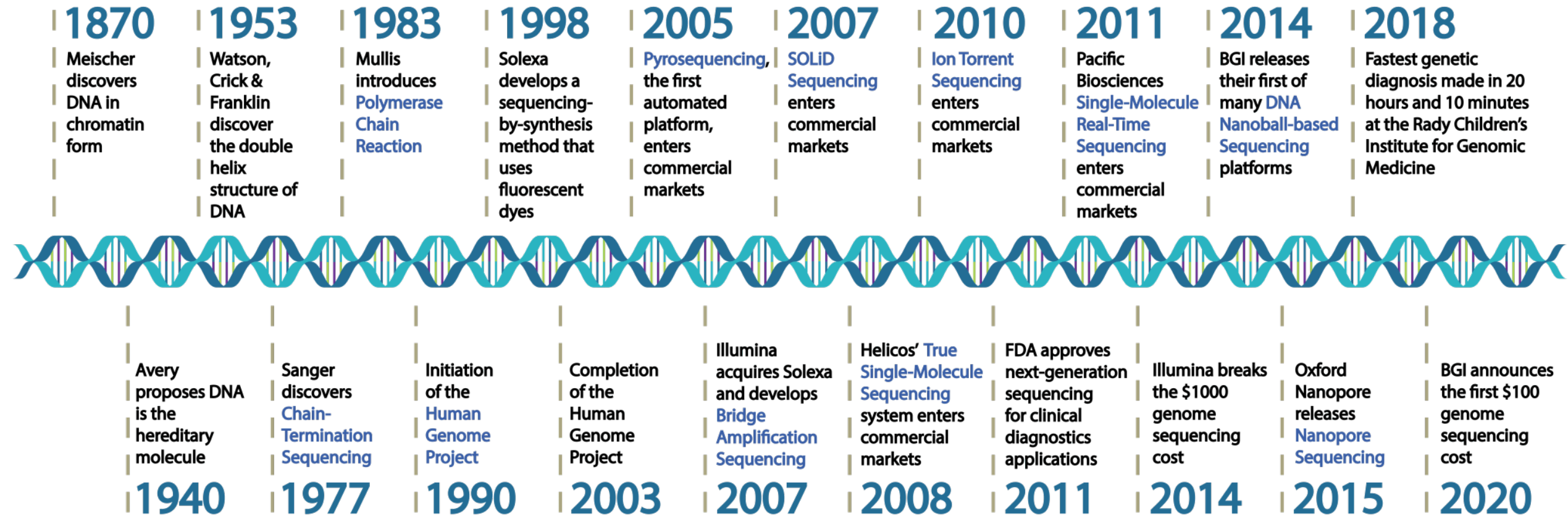
- Input type
- Throughput
- Manual intervention



## Further considerations:

- Amount and concentration of DNA needed
  - Input e.g. for library prep
- Fragment length needed e.g. for long-read sequencing
  - Bead beating will reduce fragment size
- DNA or total nucleic acid (TNA)
- Purity required for subsequent steps
- Number of samples to be extracted

# A history of DNA sequencing



# Sequencing technologies over time

## First generation



Sanger sequencing  
Maxam and Gilbert  
Sanger chain termination

Infer nucleotide identity using dNTPs,  
then visualize with electrophoresis

500–1,000 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

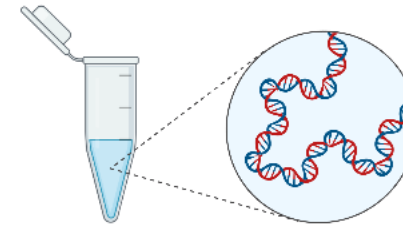
Short-read sequencing

Long-read sequencing

# Illumina sequencing workflow

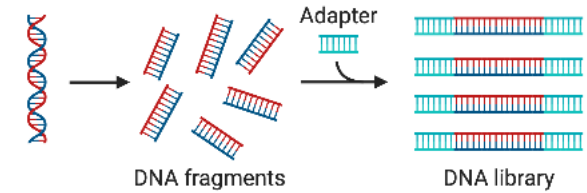
- Starting from pure nucleic acids
- Library preparation
- Sequencing
- Data analysis

**Step 1:**  
DNA extraction



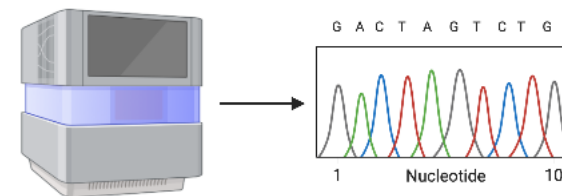
**Step 2:**

Library preparation

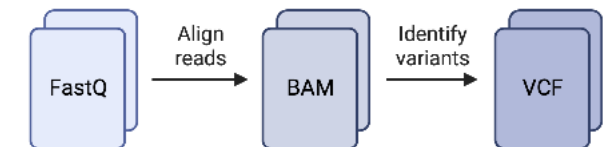


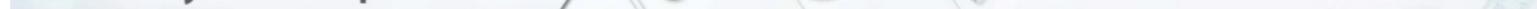
## Next Generation Sequencing Workflow

**Step 3:**  
Sequencing



**Step 4:**  
Analysis





For all you seq...

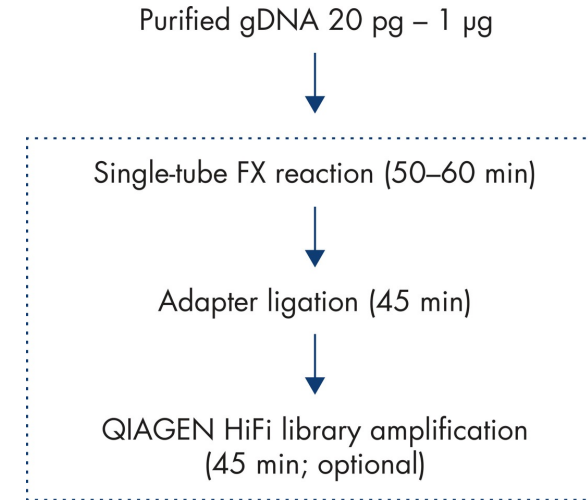
DNA

- [illegible]

# Library prep method used at IMM

## QIAseq FX workflow:

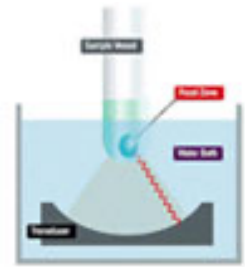
- Fragment DNA (enzymatic)
- End repair
- Adapter ligation / barcoding
- Amplification
- Quality control
  - Qubit / Fragment analyzer
- Pooling



# Library prep

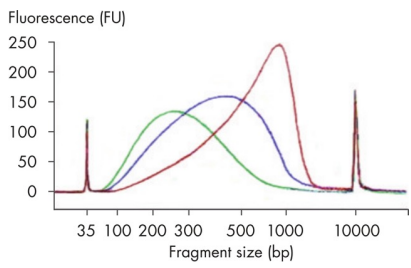
Fragmentation – ideally uniform and sequence independant

Mechanical shearing (Covaris)

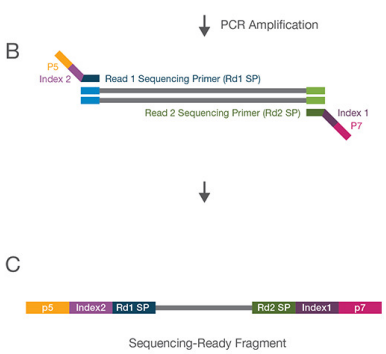
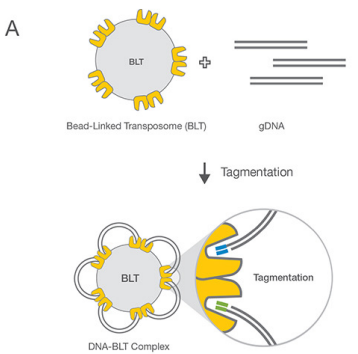


Enzymatic

- Time dependant



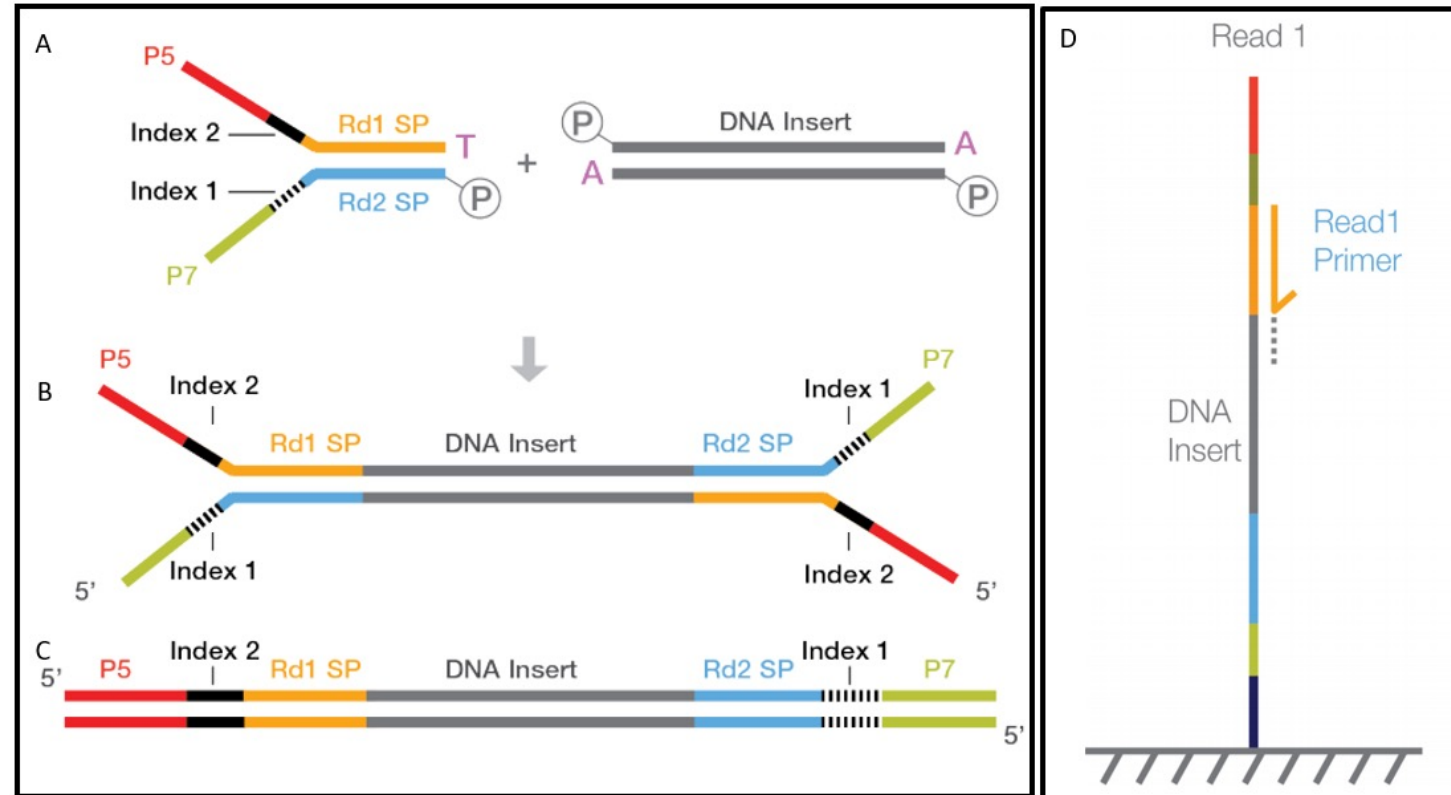
Tagmentation



# Library prep

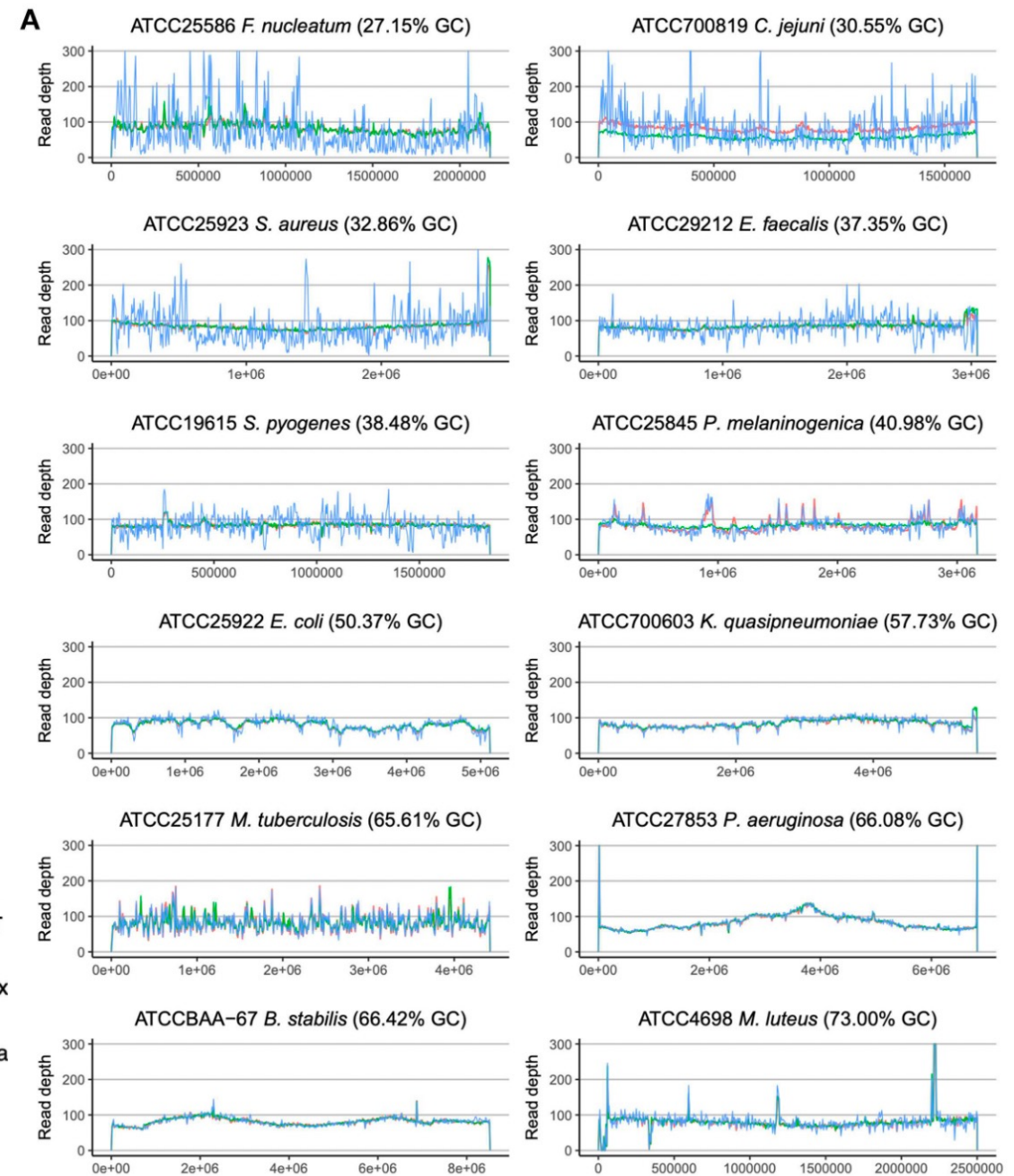
Final library contains:

- DNA insert
- P5 and P7 adapter
- Indices (Index 1 / 2)
- Sequencing primer binding sites (Rd1/Rd2 SP)

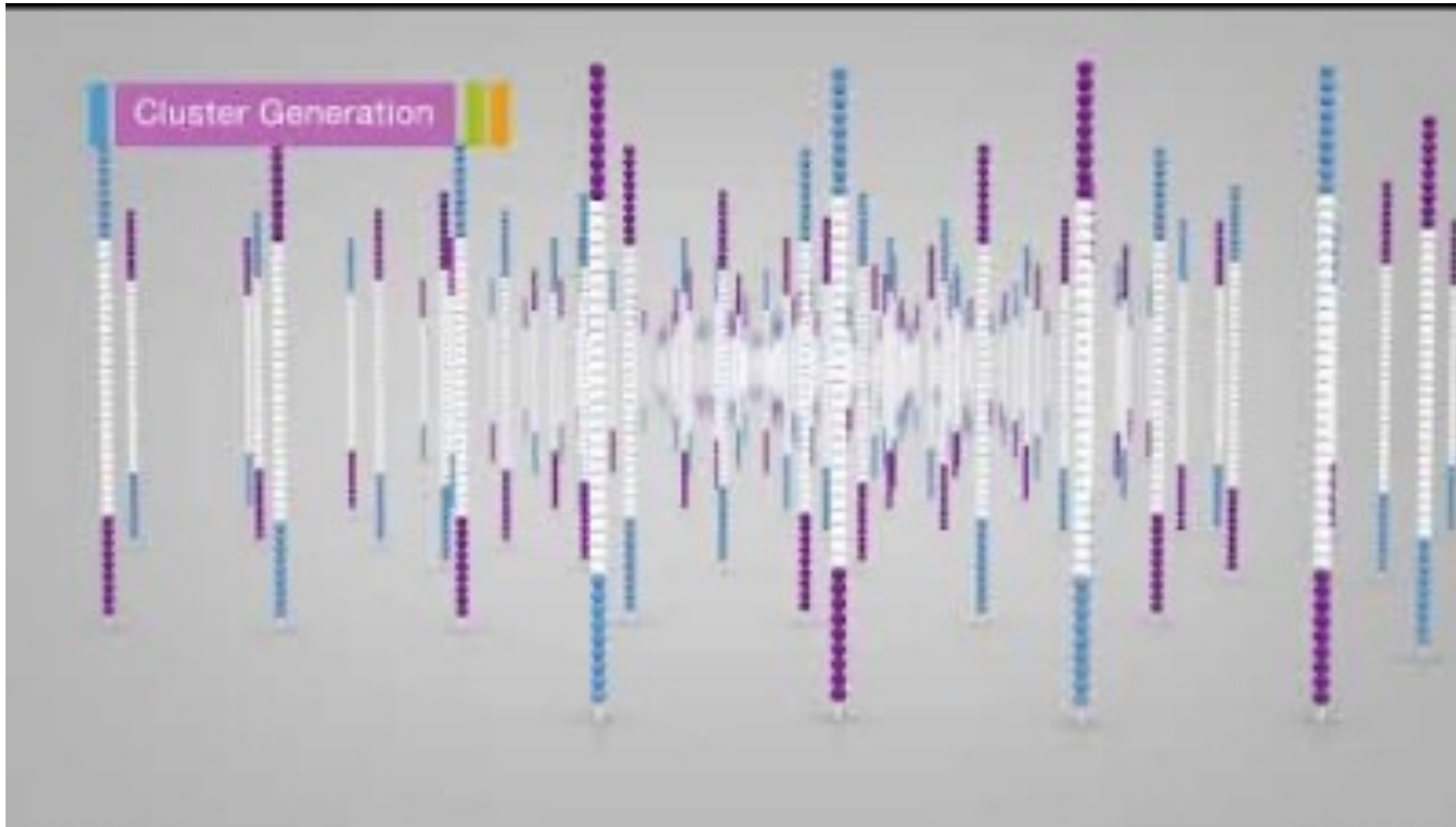


# Library prep - find the best kit

- Panel of 12 ATCC strains with complete genomes published:
  - variable genome sizes
  - variable %G+C
  - plasmid content
  - ARG content
- Tested:
  - 3 kits: Illumina DNA Prep (flex), QIAseq Fx (Qia) and Illumina Nextera XT (XT)
  - evenness of coverage
  - insert size distribution
  - base composition of reads
  - Assemblies with subsampling
  - QC, kmer content, core genes
- 50x acceptable for flex and Qia; 100x for XT



# Illumina: Sequencing by Synthesis



# Illumina: the sequencer portfolio



iSeq 100



MiniSeq



MiSeq Series +



NextSeq 550 Series +



NextSeq 1000 & 2000



NovaSeq 6000 Series +

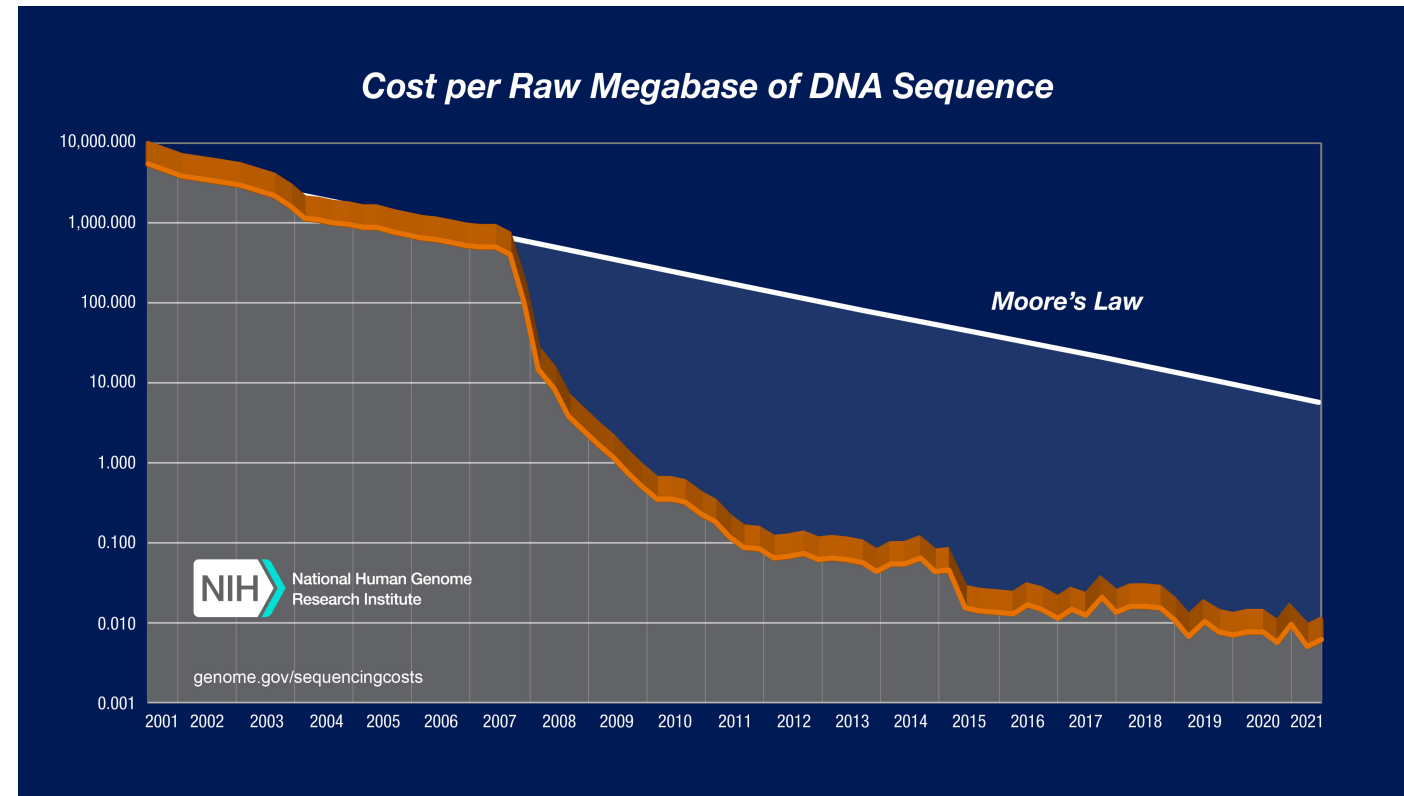


NovaSeq X Series

1.2 Gb	7.5 Gb	15 Gb	120 Gb	360 Gb	6 Tb	16 Tb
2 x 150 bp	2 x 150 bp	2 x 300 bp	2 x 150 bp	2 x 150 bp (2 x 300 bp)	2 x 250 bp	2 x 150 bp
19 h	24 h	55 h	30 h	48 h	44 h	48 h
20k CHF	50k CHF	125k CHF		200k CHF		985k CHF
	4 h 1 x 100bp	Long fragments		Long fragments		
Test libraries	Few bacterial genomes	16S amplicon		Transcriptomes Eukaryotic genomes	Transcriptomes Eukaryotic genomes	Human genomes

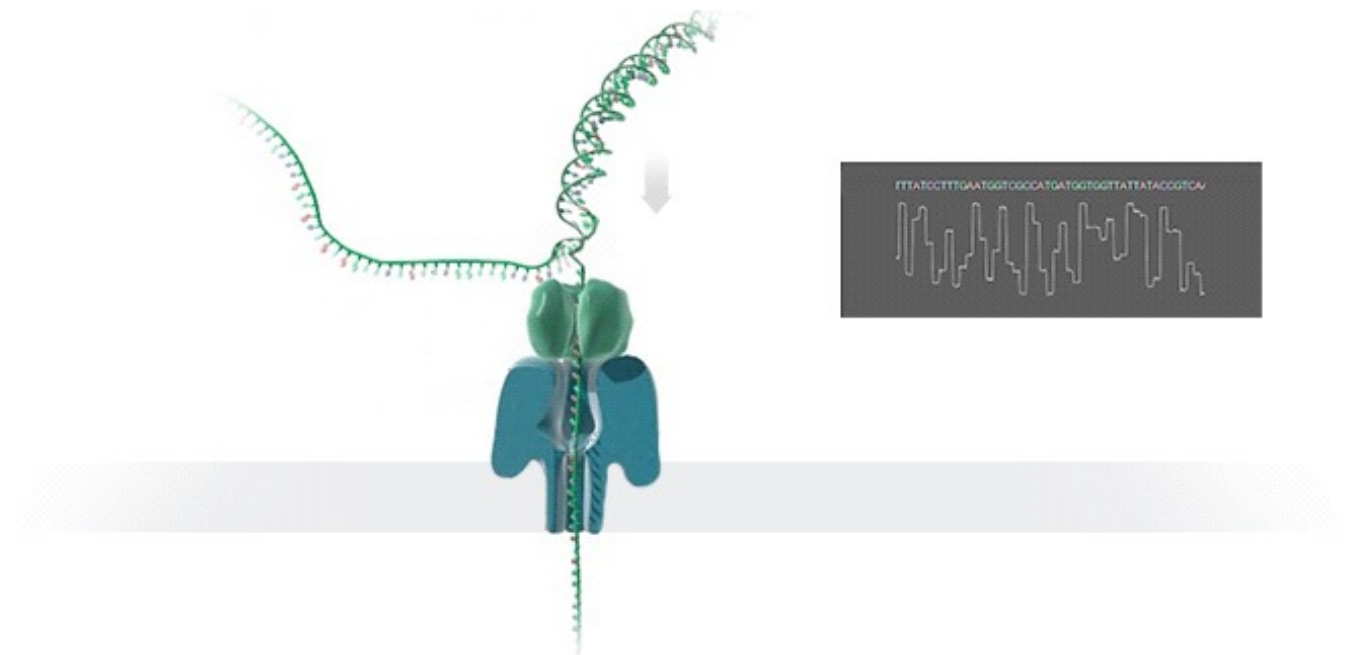
# Illumina sequencing cost

- Sequencing cost is constantly dropping
- Recently more competitors (again) (MGI, Nanopore)
- 100\$ human genome announced in 2022



<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

# Nanopore sequencing



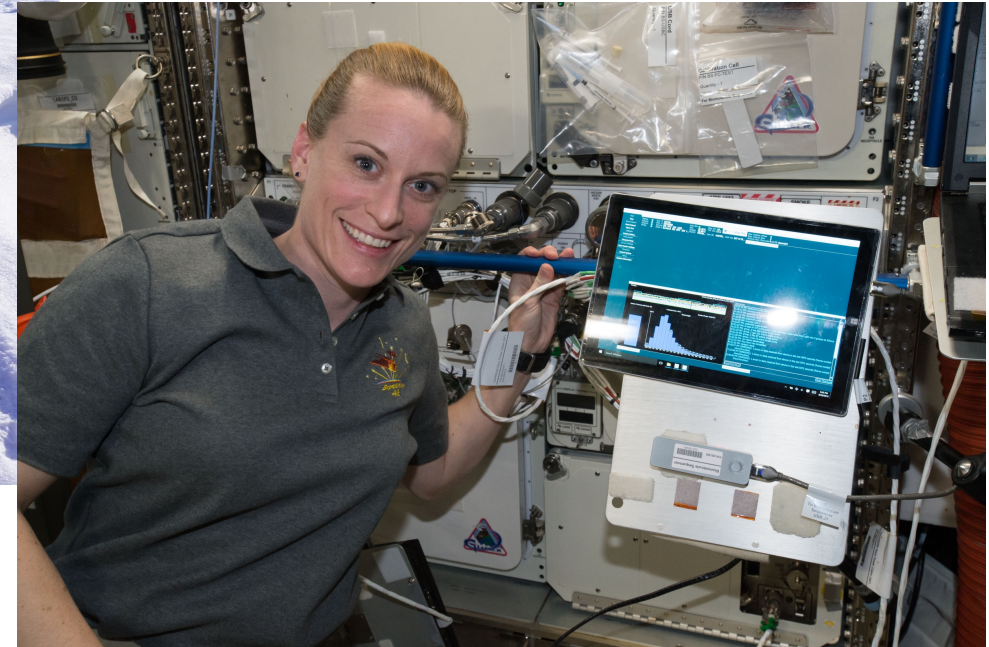
# Nanopore sequencing – sequencing everywhere



RAIN FORREST BRAZIL



CANADIAN HIGH ARCTIC



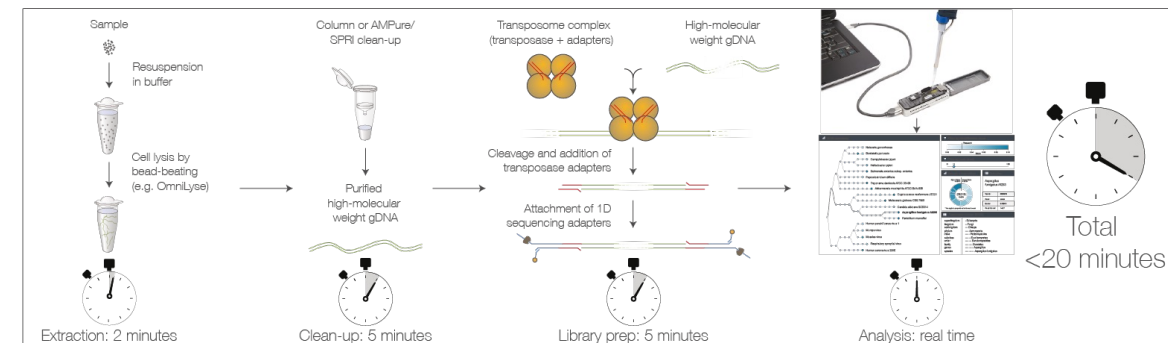
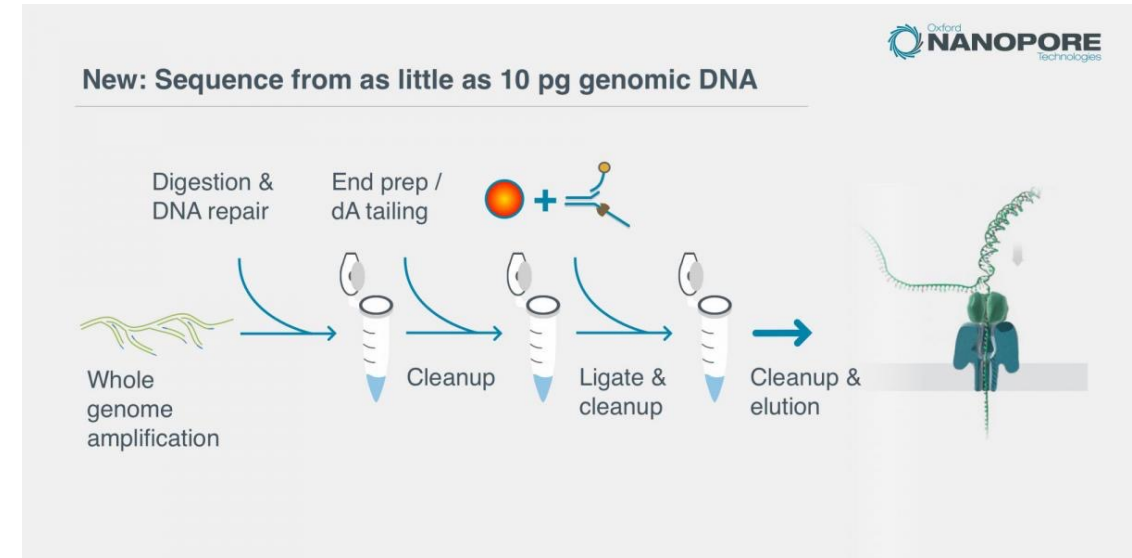
INTERNATIONAL SPACE STATION (ISS)

# Nanopore DNA extraction

- Long read sequencing requires good (long) input DNA
  - Check DNA size distribution on a gel
- Sequencing length mainly limited by library size
- Many extraction methods developed for PCR or Illumina sequencing

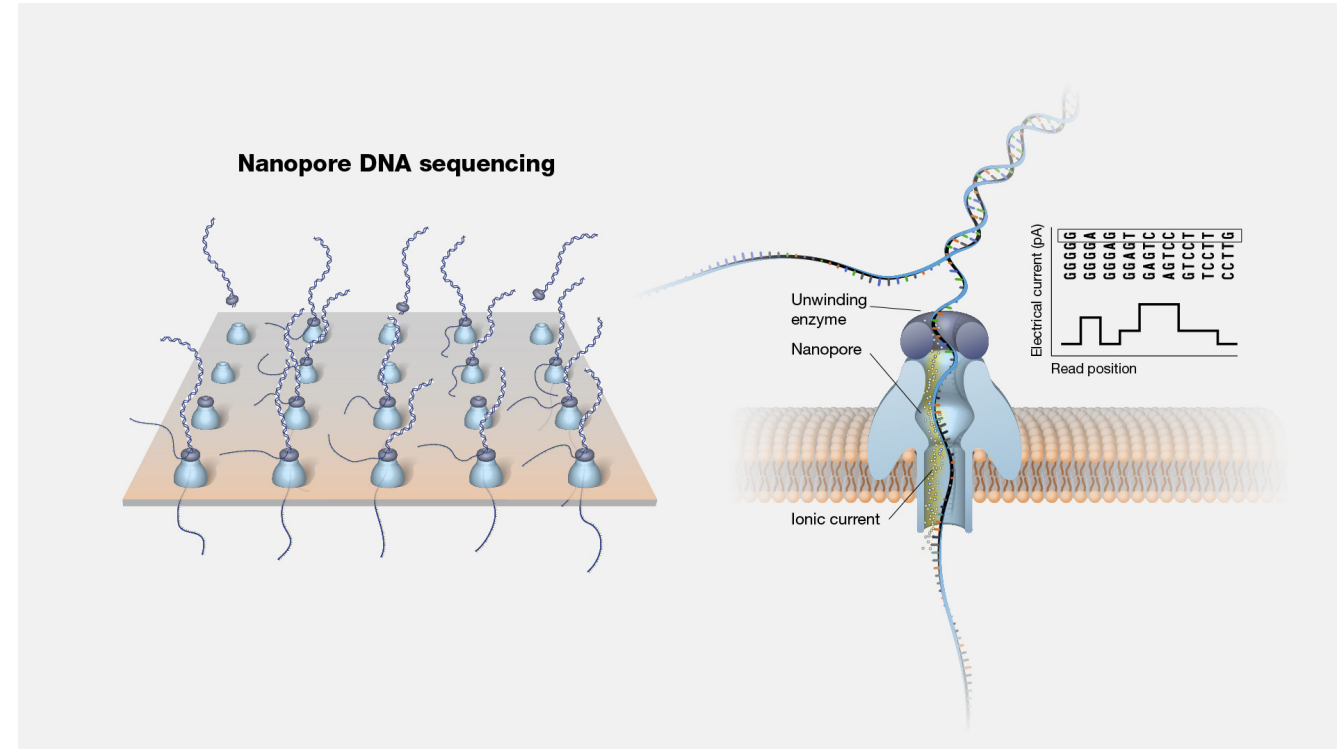
# Nanopore library prep

- Add adapter and motor protein to DNA fragments
- Currently 2 strategies
  - Ligation
    - Longer libraries
    - 4-5 h protocol
    - Good for de-novo sequencing
  - Tagmentation
    - Used in this course
    - Shorter libraries
    - Fast – 20 minutes
    - Good to sequence plasmids

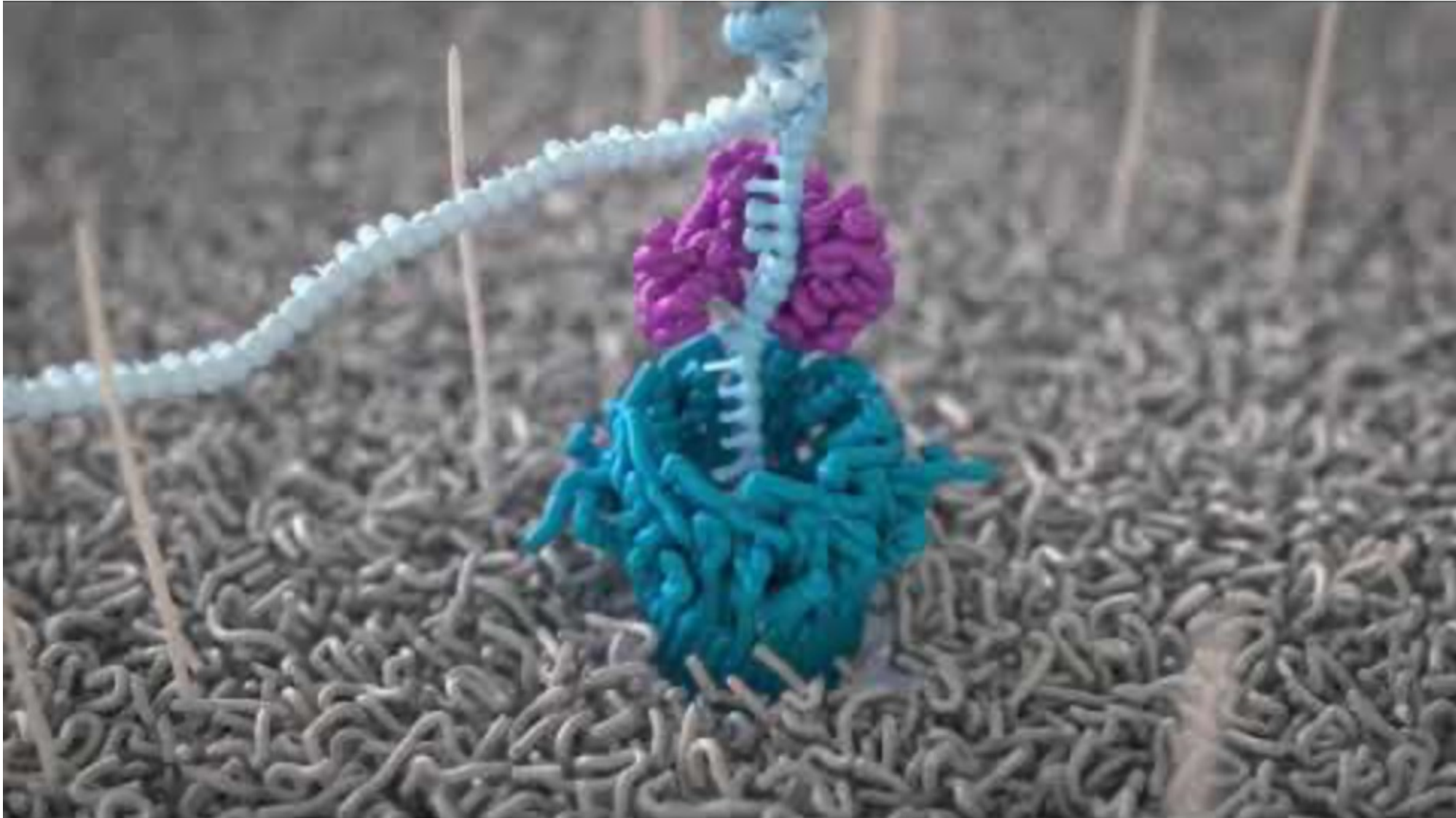


# Nanopore sequencing

- Use pores to “read” the DNA
- Pores embedded in membrane
- Apply tension to membrane to create a constant flow of ions through the pore which can be measured by electrodes
- Motor protein moves DNA through the pores
- While DNA passes through the pore, the flow of ions is changed depending on the base that is currently in most narrow site of the pore
- Even modifications of bases (e.g. methylation) can be measured



# Nanopore sequencing



# Nanopore sequencing - limitations

Data quality!

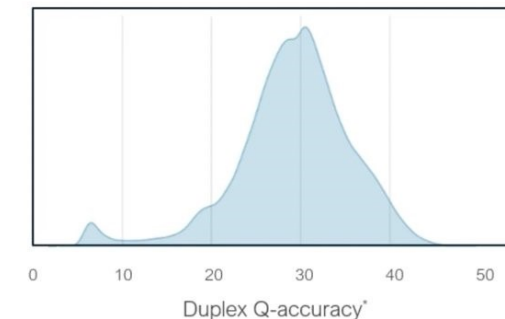
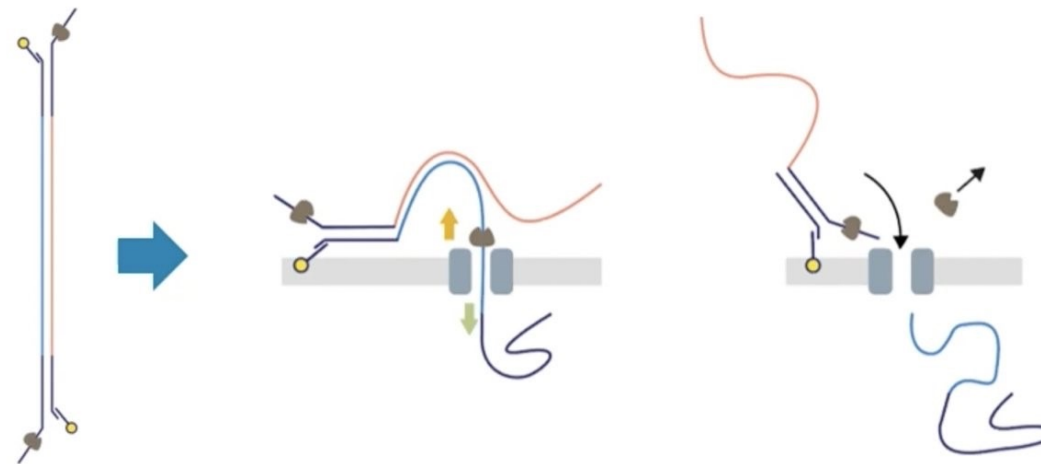
Originally 5% error rate (1 in 20 bases wrongly called)

New flowcells released: R10.4.1

based on base caller (fast / super accuracy / duplex)

Less automation

Less standardization



<https://nanoporetech.com/resource-centre/clive-brown-ncm-update-2021>

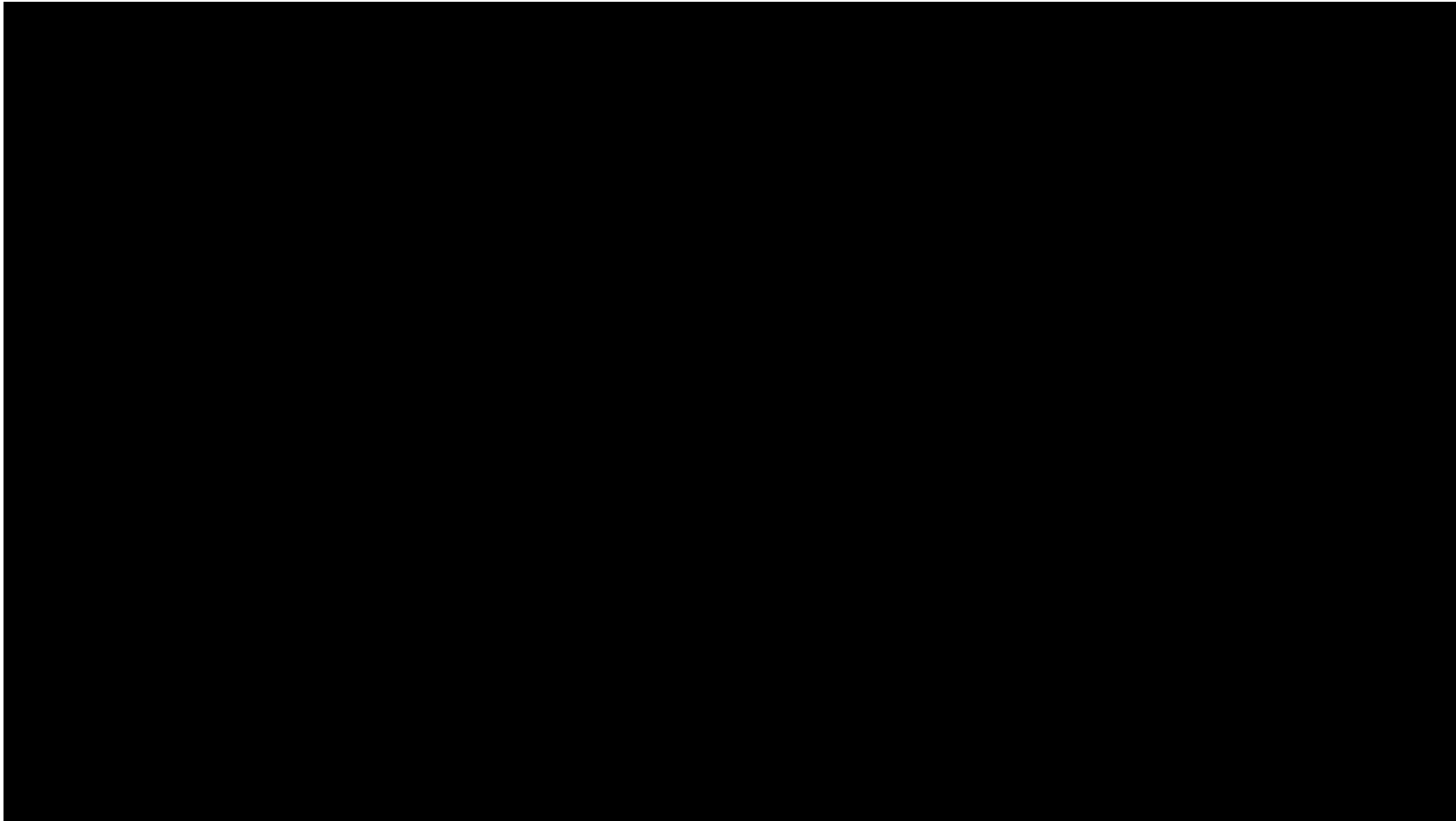
# Nanopore sequencing – adaptive sequencing

Select what you want to sequence

- Positive selection
- Negative selection

Up to 13.87-fold enrichment

Only works on fragments  
> 400bp



<https://doi.org/10.1186/s13059-021-02582-x>