History of sequencing, overview & comparison of NGS technologies

Bacterial Genomics: Sequencing and Bioinformatics Training

15-July-2025

MicroHub Lab Kericho

Facilitator: Collins Kigen

Overview of Sequencing

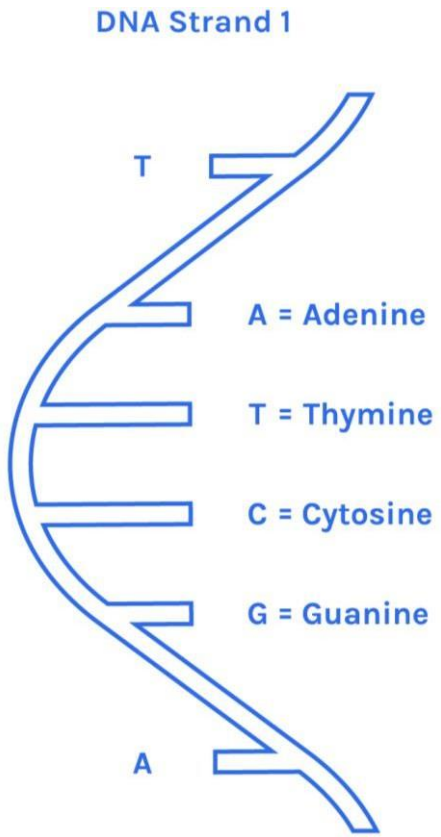
Definition of sequencing

What is DNA sequencing?

- Determining the order of the 4 nucleotides that make up a DNA strand or oligonucleotide

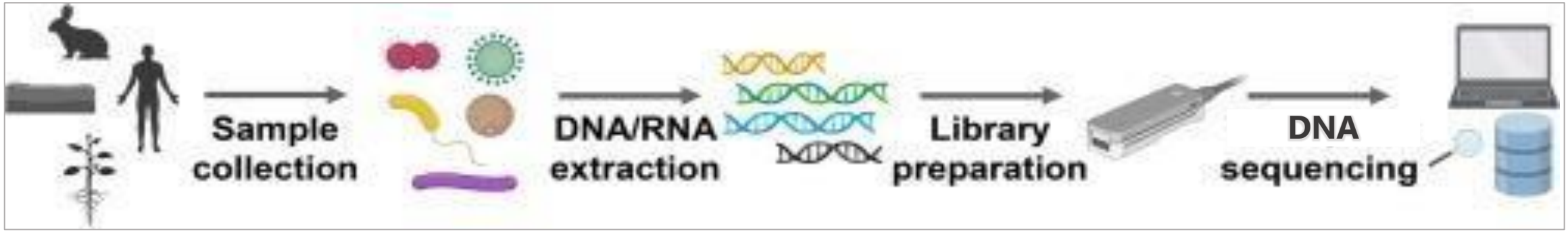
What is Next Generation Sequencing - NGS?

- Technology used for sequencing many DNA strands **at the same** time instead of **one at a time** as with traditional Sanger sequencing by capillary electrophoresis.
- Also called "**massively-parallel sequencing**".
- Enabled by use of unique **barcodes** or **indexes** to label DNA libraries of individual samples.



https://medium.com/prime-movers-lab/how-dna-sequencing-started-the-microbiome-revolution-and-what-it-means-for-the-future-e660006e0d98

General steps of sequencing



Qubit

NanoDrop

DNA quantification

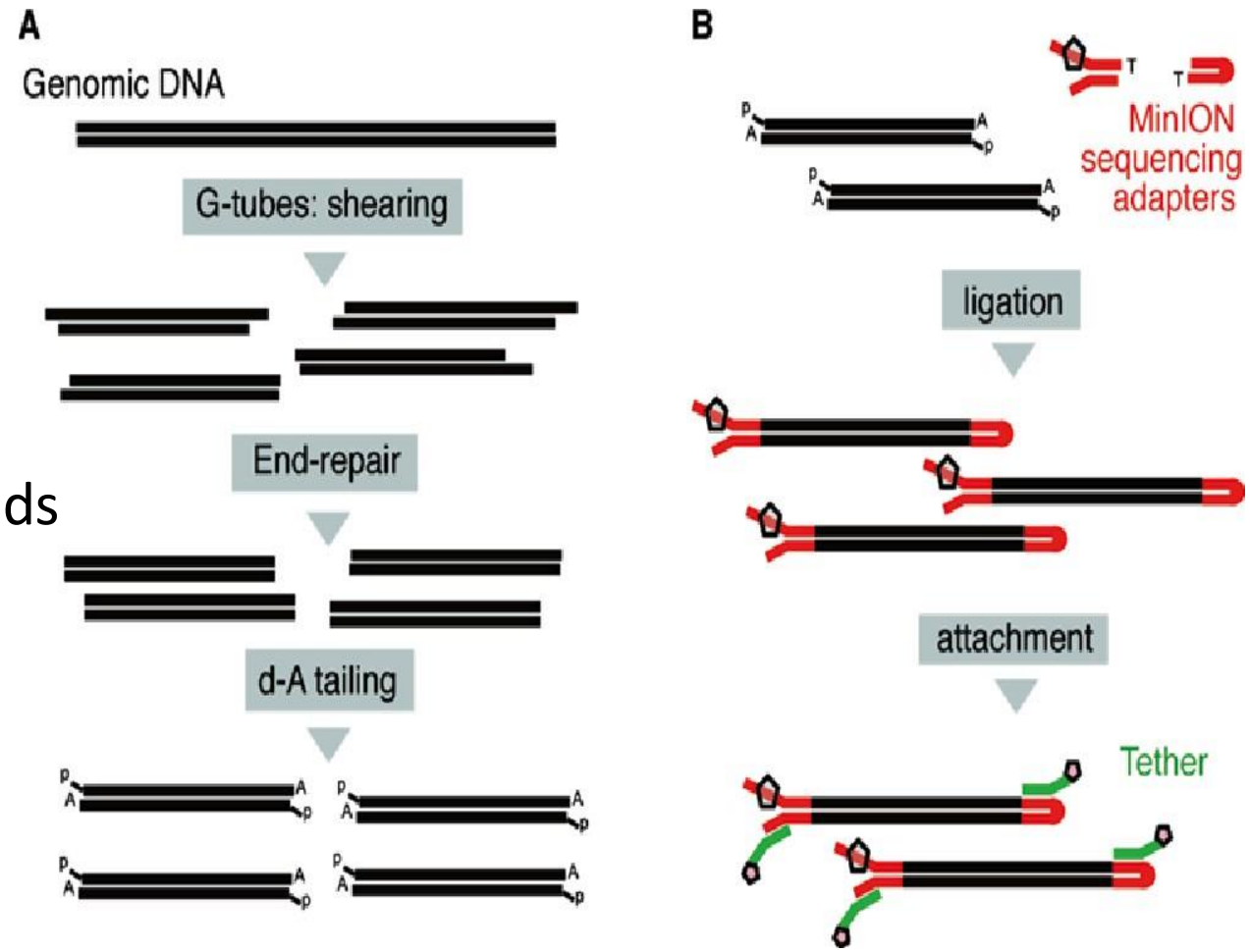
Data acquisition & analysis

https://medium.com/prime-movers-lab/how-dna-sequencing-started-the-microbiome-revolution-and-what-it-means-for-the-future-e660006e0d98

General steps of sequencing

Library preparation

1. DNA fragmentation (A)
   a. Covaris tubes
   b. Insulin Needles
2. End-repair (A)
   a. Converting stick ends to blunt ends
   b. Adding short oligo A tail
3. Barcoding (B)
4. Adapter ligation (B)



History of Sequencing

History of sequencing

1869	First DNA isolation by Friedrich Mietscher
1953	Discovery of DNA structure by Watson, Crick and Franklin
1965	Robert Holley sequenced the first tRNA
1972	Sequenced the 1st complete protein coding gene – coat protein of bacteriophage MS2
1977	Fredrick Sanger discovered the first sequencing method – Chain Termination method
	Sanger sequenced the first complete genome - bacteriophage PhiX174

History of sequencing

- First Generation sequencing methods

Year of Discovery		Technology	Maximum read length
1977	Fredrick Sanger	Chain Termination Method	~50bp
1977	Maxam & Gilbert	Chemical modification of DNA	~50bp
1984	Fritz Pohl	Direct blotting electrophoresis system GATC1500	N/A
1987	Leroy Hood & Michael Hunkapiller	ABI 370 – Applied Biosystems - First automated Sanger sequencer	20 – 30 bp

History of sequencing

- Second generation sequencing

Year of Discovery	Developers/ manufacturers	Technology	Maximum read length
1996	Mostafa Ronaghi, Mathias Uhlen & Pål Nyŕen	Pyrosequencing - sequencing by synthesis	1000bp
2005	Jonathan Rothberg & colleagues	Roche 454 Sequencing System - Automated pyrosequencing	1000bp
2006	Applied Biosystems	Life Technologies SOLiD system – sequencing by ligation	60bp
2007	Illumina Inc.	Illumina sequencing – sequencing by synthesis	150 – 500bp
2010	Ion Torrent Systems Inc	Ion Torrent - pH-mediated sequencer	100bp

History of sequencing

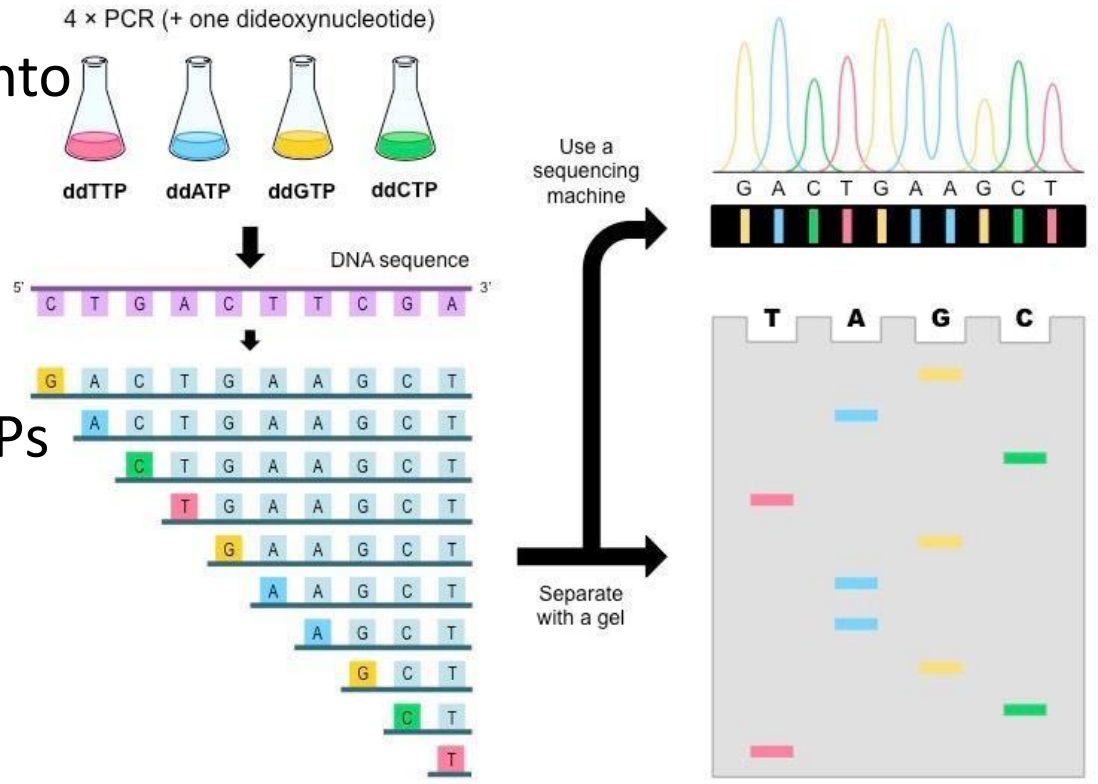
- Third Generation Sequencing

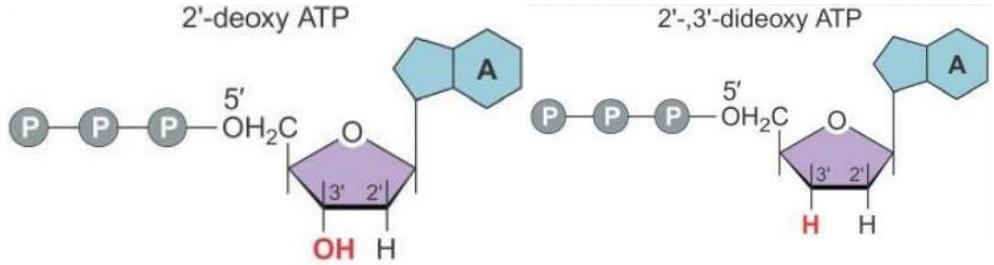
Year of Discovery	Manufacturer	Technology	Average read length
1996	Oxford Nanopore Tech Limited	Nanopore sequencing	Long 10 – 100kb Ultralong - 2Mb
2005	Pacific Biosciences SMRT	PacBio sequencing	13 - 20kb

Principles of various Sequencing Technologies

Sanger sequencing - principle

- The double-stranded DNA (dsDNA) is denatured into two single-stranded DNA (ssDNA).
- A primer that corresponds to one end of the sequence is attached.
- Four polymerase solutions with four types of dNTPs and only one type of ddNTP are added.
- The DNA synthesis reaction initiates and the chain extends until a termination nucleotide is randomly incorporated.
- The resulting DNA fragments are denatured into ssDNA. The denatured fragments are separated by gel electrophoresis and the sequence is determined.



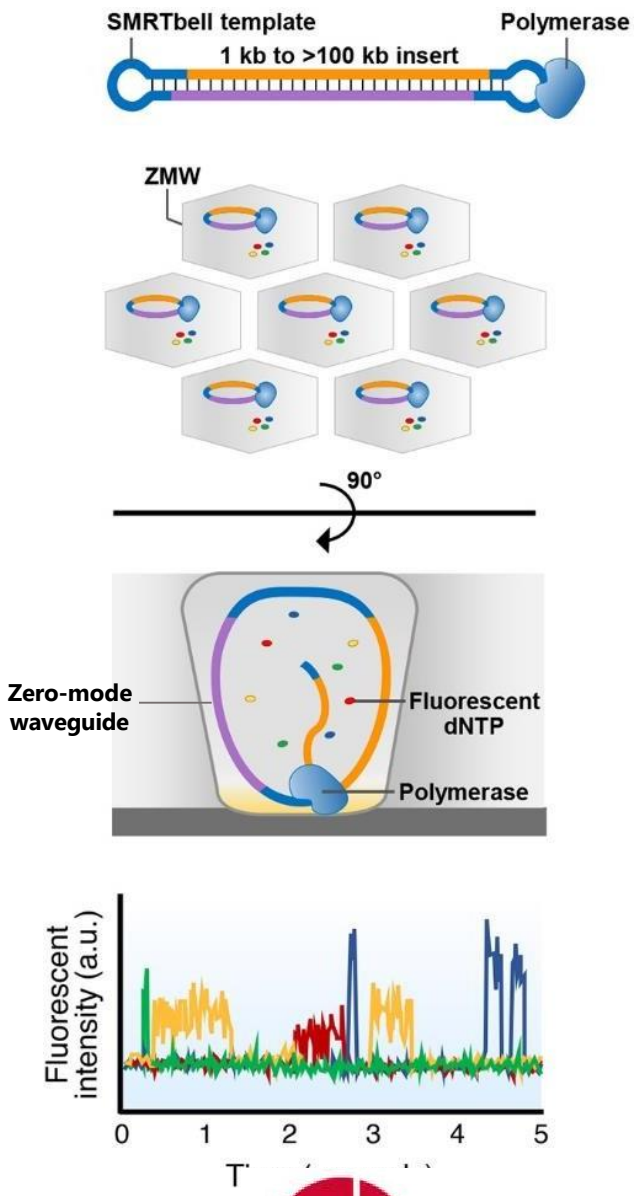


https://www.slideserve.com/dahlia-hendrix/molecular-biology-techniques-a-primer-powerpoint-ppt-presentation

https://microbenotes.com/dna-sequencing-maxam-gilbert-and-sanger-dideoxy-method/

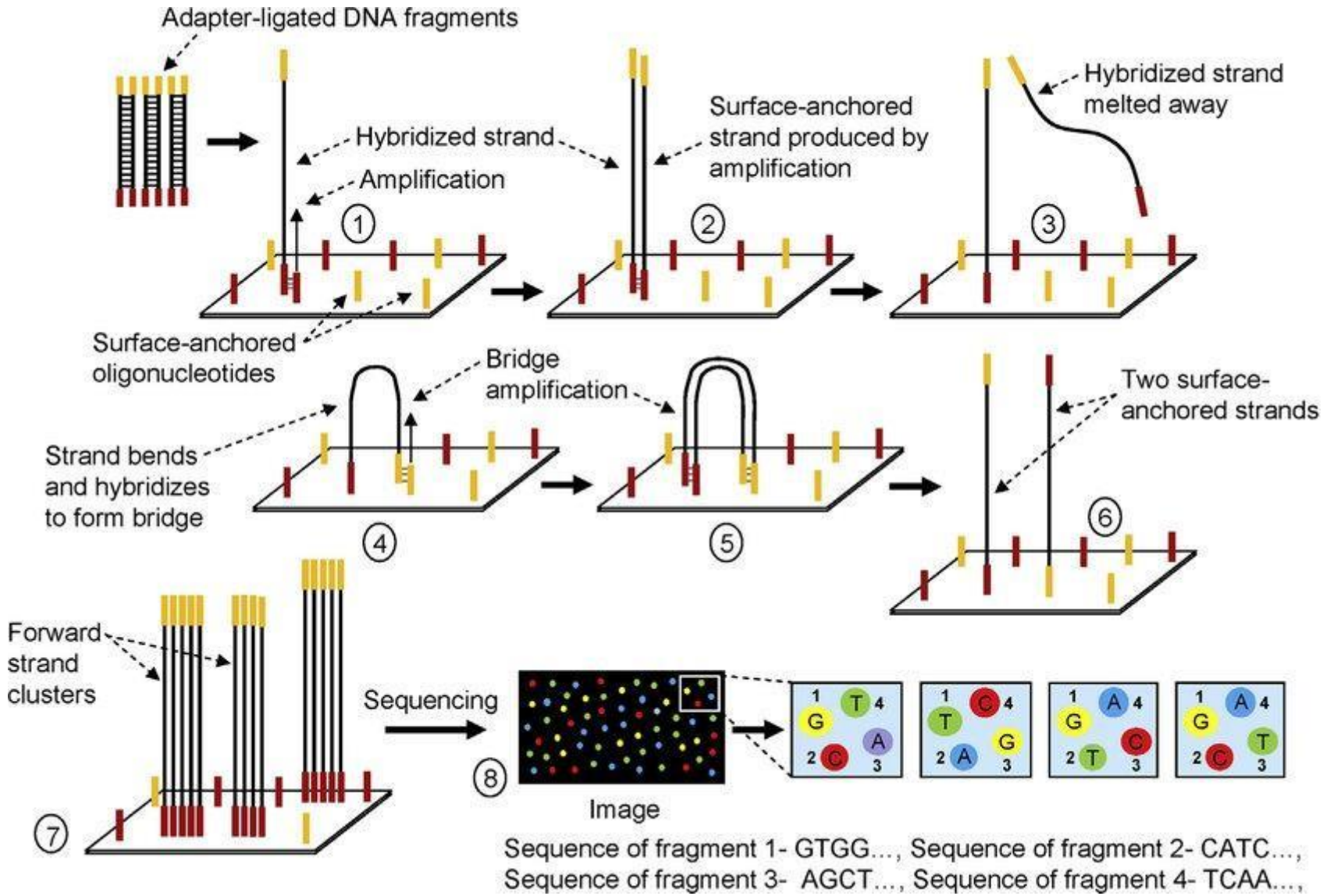
PacBio sequencing SMRT- principle

- SMRT – single molecule real time
- ZMW – zero mode waveguide
- SMRT bell generated during DNA lib prep
- Its bound by DNA polymerase & loaded into a SMRT cell with >8 million ZMW chambers
- Elongation occurs through addition of one flourescent labelled dNTP to SMRTbell template at a time
- A light pulse excites the fluorophore; emission detected by camera and converted to corresponding base
- Fluorophore cleaved and released into sequencing buffer to complete one cycle

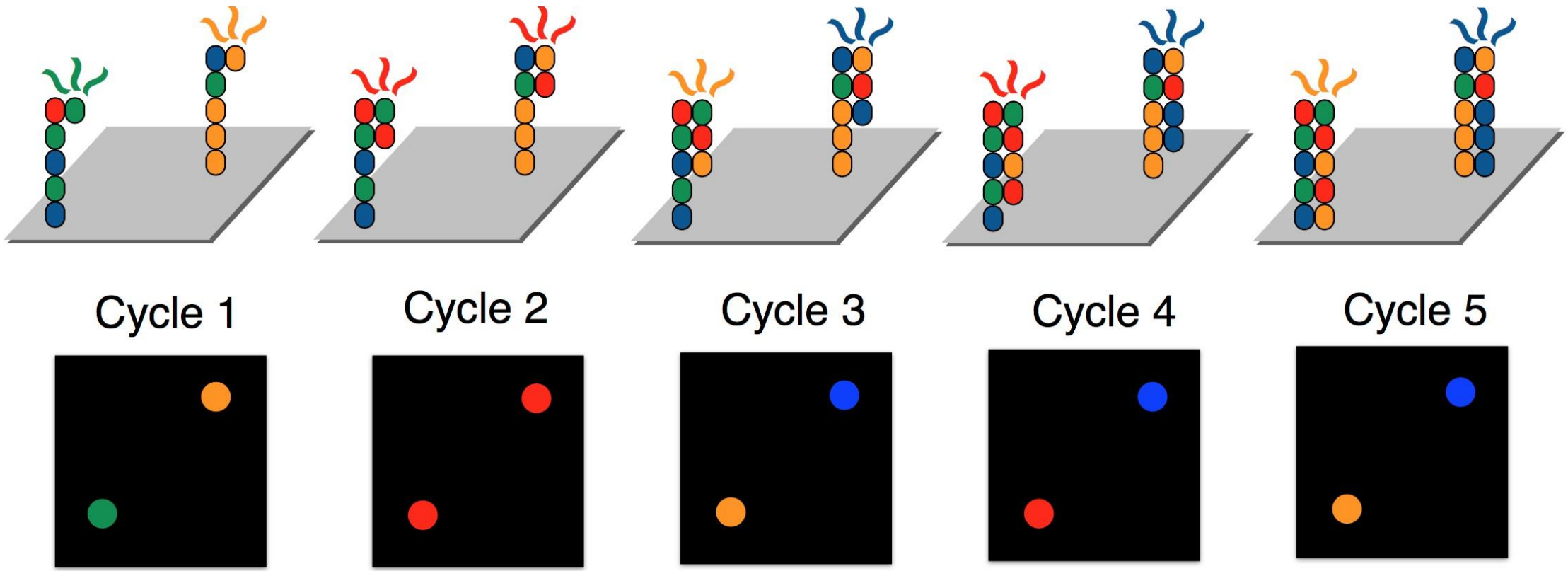


PACBIO®

Illumina sequencing - principle



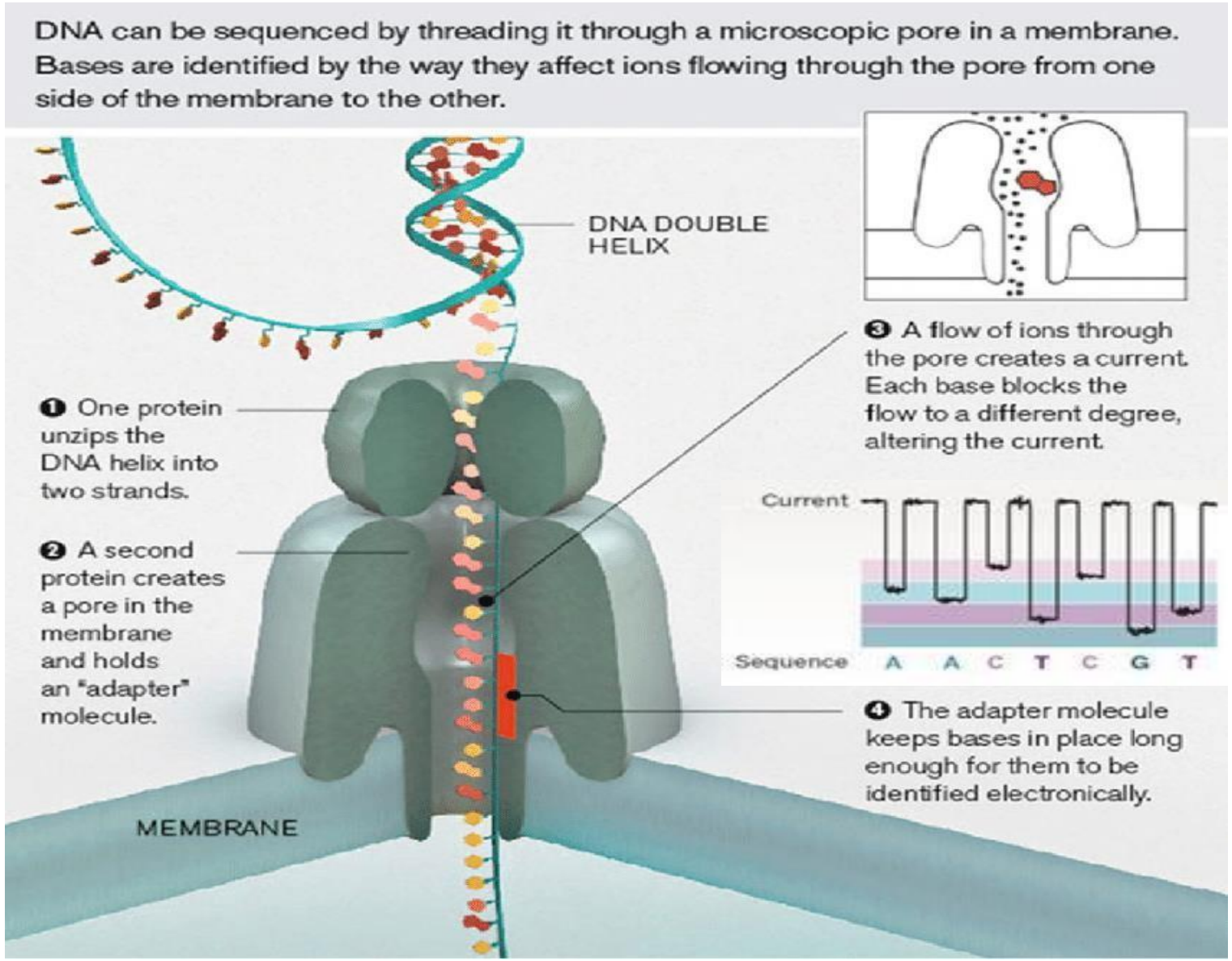
Illumina sequencing - principle



Illumina sequencing - devices

Sequencing System	iSeq™	MiniSeq™	MiSeq®	NextSeq®	HiSeq®	HiSeq® X	NovaSeq®
					4000	Five/Ten	6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1.5 Tb	1.8 Tb	1 Tb - 6 Tb[1]
Instrument price	$19.9K	$49.5K	$99K	$275K	$900K	$6M[2]/$10M[2]	$985K
Installed base[3]	NA	~600	~6,000	~2,400	~2,300[4]		~285

Oxford Nanopore sequencing - principle

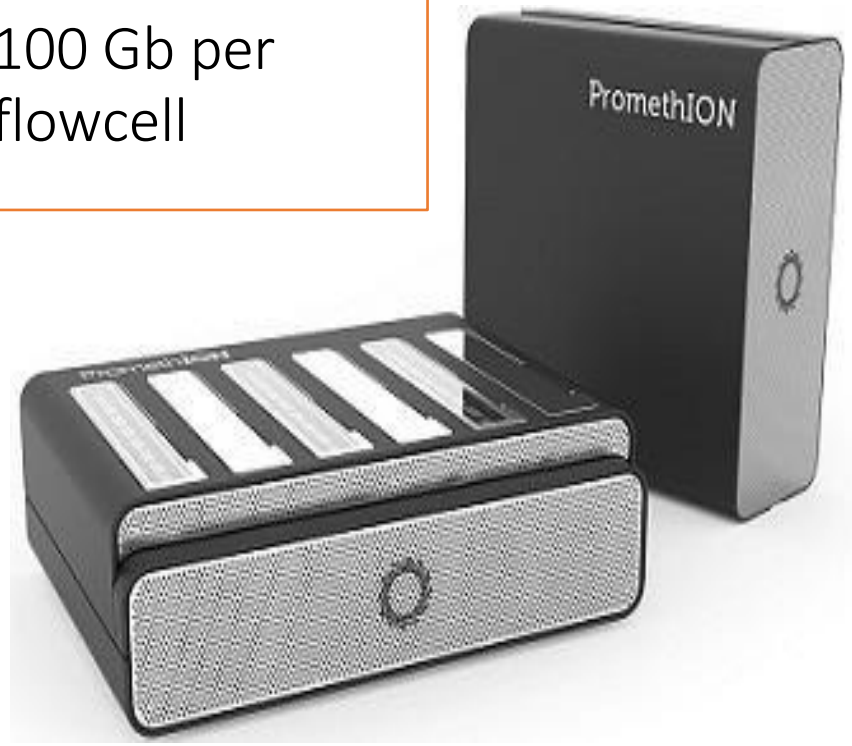


Nanopore sequencing - devices

1. MinION
   - 1 flowcell:
   - 2048 pores
   - 512 pores are used at a time
   - ~ 400 bases/sec are read
   - 10 -25 Gb data

2. GridION
   - 5 flowcells
   - 2048 x 5 pores
   - 100 Gb data

3. PromethION
   - 48 flowcells
   - 1200 pores per cell
   - 100 Gb per flowcell



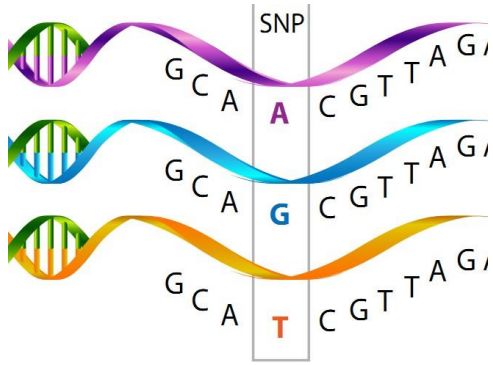
Comparison of Sequencing Technologies

Long versus Short sequencing technologies

Feature	Sanger	Oxford Nanopore	Illumina	PacBio HiFi
Read length	Short reads: 200 - 600bp	Long reads 10 – 100kbp Ultra-long ~3Mb	Short reads 150 – 250bp	Long reads 10 – 20kbp
Read type	Paired end reads	Single reads	Paired-end reads: forward & reverse	Single reads
Accuracy	100% Gold Standard	R.9.4.1 89 – 93% R10.4.1 99%	99.9%	99.9%
Applications	Amplicon sequencing, Clone checking ChIP-Sequencing RNA Sequencing	Whole genome sequencing, Plasmid reccnstruction	Small genome sequencing, amplicon sequencing, clone checking & ChIP-Seq, RNA Seq	Whole genome sequencing

Advantages of Illumina sequencing

- High accuracy of ~99.9% - SNP analysis, allele identification
- High sequence coverage – good sequence depth >100
- High throughput data - sequence billions of template strands simultaneously
- Parallel sequencing – use of unique indexes to label diff samples



Disadvantages of Illumina sequencing

- Initially expensive to install
- Non-contiguous or incomplete assemblies
- Poor resolution of regions rich in repeat sequences & AT rich genomes

illumina®

Advantages of Nanopore sequencing

- Ultra-long read lengths – longest reads up to date is >4 Mb in length; Resolve plasmid sequences & repetitive regions; completion of small genomes (viruses, bacterial)
- Real-time data analysis: no fixed run-time; stop run when data is sufficient; sequence new genome of unknown length
- Direct molecular analysis - Sequencing native DNA (and RNA) avoids amplification bias
- Portable – fits in adult palm – Flongle, Minion and Mrk1c

Disadvantages of Nanopore sequencing

- Sequencing error of ~5% - early chemistries R9.4.1, R10.3

  New chemistry R10.4.1 promises higher accuracy of 99.9%

Long reads versus short reads – Human Genome

Human Genome Project (1990-2003)

Generated first draft of human genome – 92% complete

Used Bacterial Artificial Chromosome cloning and Sanger short read sequencing

Telomere-to-Telomere (T2T) consortium 31st March 2022

Completed the remaining 8% (Nurk et al 2022) that were complex regions:

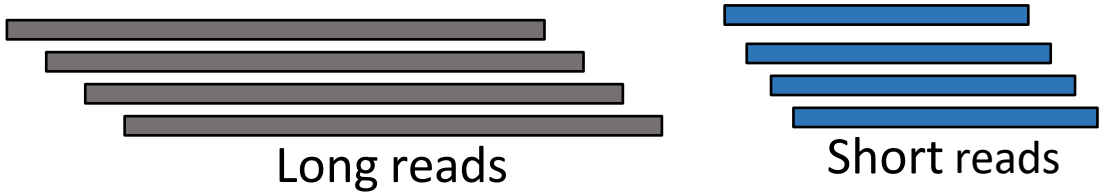
centromeric satellite arrays, subtelomeric regions, rDNA arrays, short arms of acrocentric chromosomes – rich in long repeat sequences

Used using PacBio HiFi and ONT:

- ONT produced ultra-long reads - >100kb
- PacBio HiFi produced high accurate reads - ~20kb

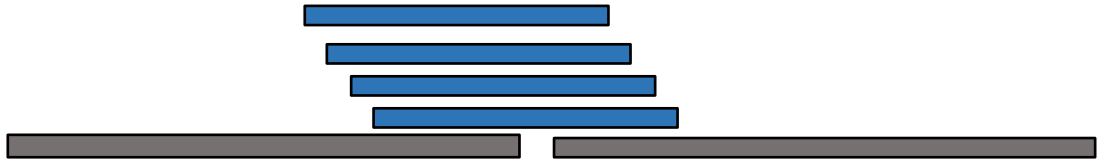
Hybrid genome assembling

Use of long and short reads to polish an assembly



Two strategies of generating a contiguous sequence:

1. Short reads to polish long read assembly



2. Use long reads to polish a short read assembly

Question & Answer session