Bridging the gaps in bioinformatics/Raw data QC

Introduction to sequencing

February 2025, Søren Hallstrøm, Statens Serum Institute, Denmark

Outline

This session consists of the following elements

1. Introduction to Sequencing
2. Brief overview of the evolution of sequencing technology
3. Sequencing in National surveillance of bacterial pathogens

Objectives

Specific objectives of this session:

1. What is sequencing
2. The technological advancements
3. When is sequencing an advantage

The Lecturer

Søren Hallstrøm, Ph.D. in molecular microbiology

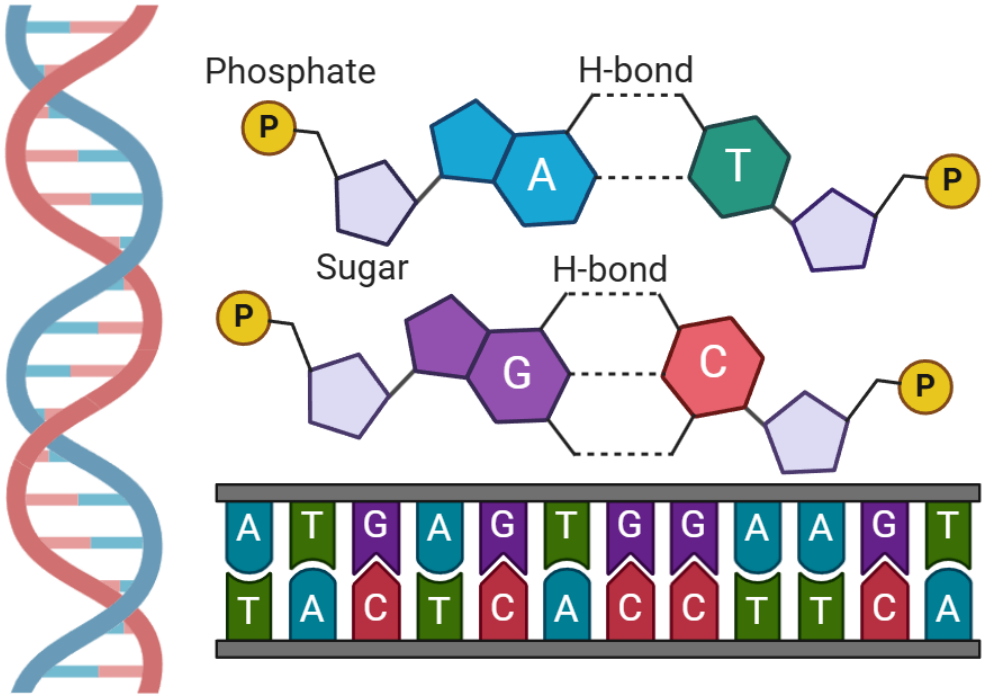
Academic staff
the Sequencing Core Facility -
Department of Sequencing & Bioinformatics,
Statens Serum Institute, Denmark

- Method development, mainly bacterial Whole Genome Sequencing (WGS)
- Illumina and Oxford Nanopore Technology (ONT or nanopore)
- Quality assurance and maintenance (semi-)automated NGS workflows

The Definition

DNA sequencing is the determination of a precise order of the nucleotides –

Adenine, Guanine, Cytosine and Thymine in a given DNA fragment.



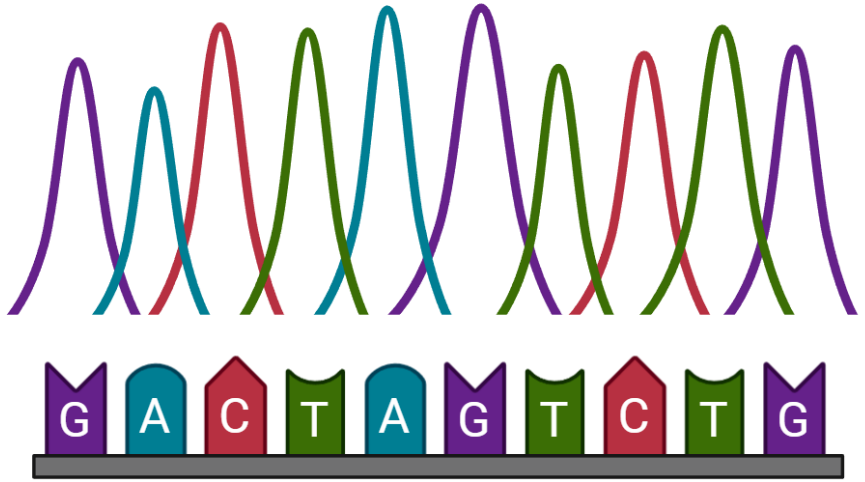
DALL-E 2: Double stranded DNA

The Basics

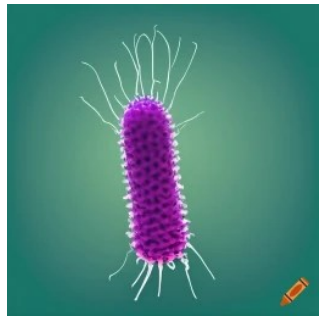
Genetic information is stored in DNA sequences

This information can be extracted by determining the correct order of the nucleotides in a given DNA sequence.

The precise order of the nucleotides in a given DNA fragment determines the structure and function of gene products.

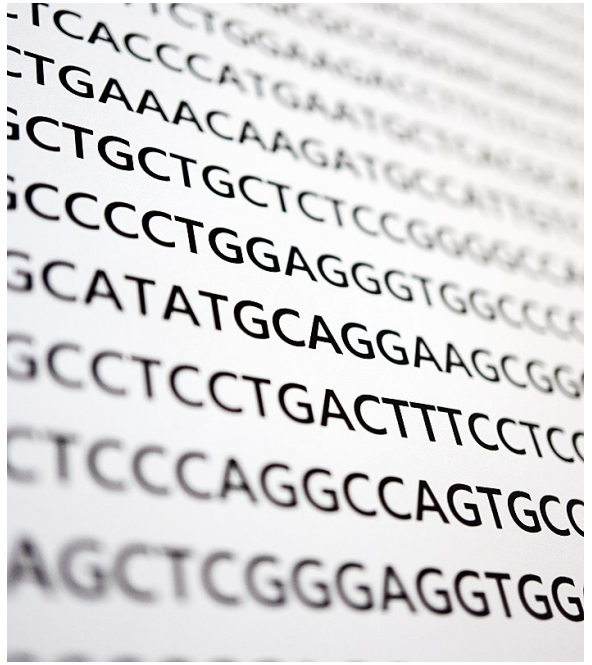


Sequencing fundamentals



Nucleotide Extraction

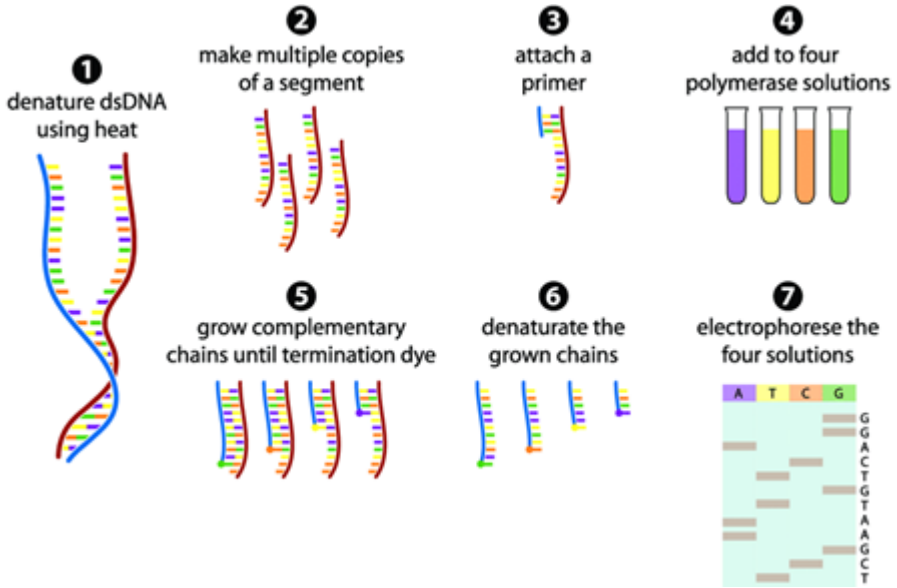
Sequencing Library preparation



A brief history of sequencing technologies

1st generation Single molecule Sanger	2nd generation Short read Illumina	3rd generation Long Read Nanopore
Chain termination Single gene fragments	Sequencing by Synthesis Parallel sequencing Complex DNA Libraries	Long read sequencing Real time basecalling
First sequence of a DNA genome: bacteriophage φX174, in 1977	Illumina (Solexa) On board amplification of sequencing library (Bridge amplificaion	Oxford Nanopore Technologies Low price -Low quality Q10-20 data
Q20 - Q30 data	Q30 – Q40 data	Pacific Biosciences PacBIO High price -> High quality Q30-40 data

1st generation Sanger sequencing



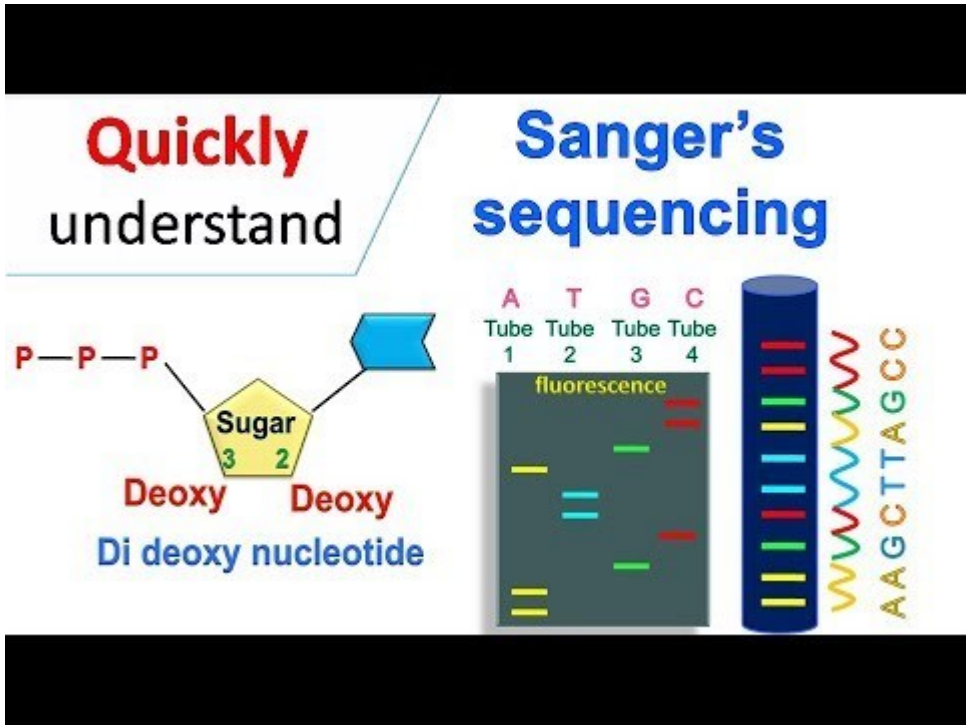
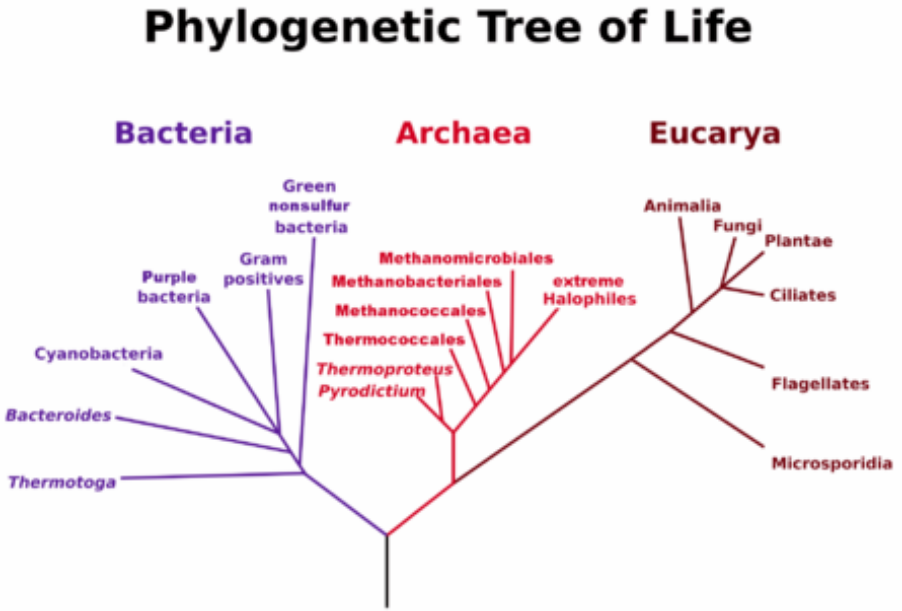
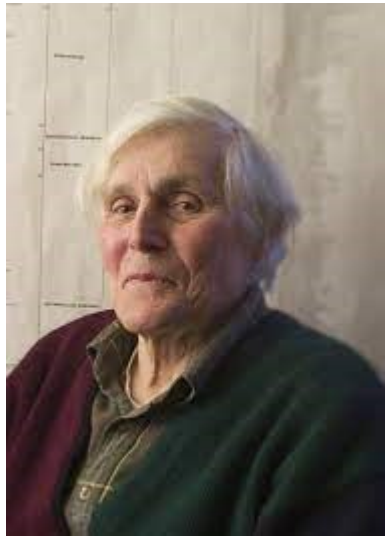


Image credit: Canva
https://frontlinegenomics.com/a-history-of-sequencing/#:~:text=The%20first%20major%20breakthrough%20in,him%20his%20second%20Nobel%20Prize.
https://www.youtube.com/watch?v=KTstRrDTmWI

Sanger sequencing – The tree of life and the third domain





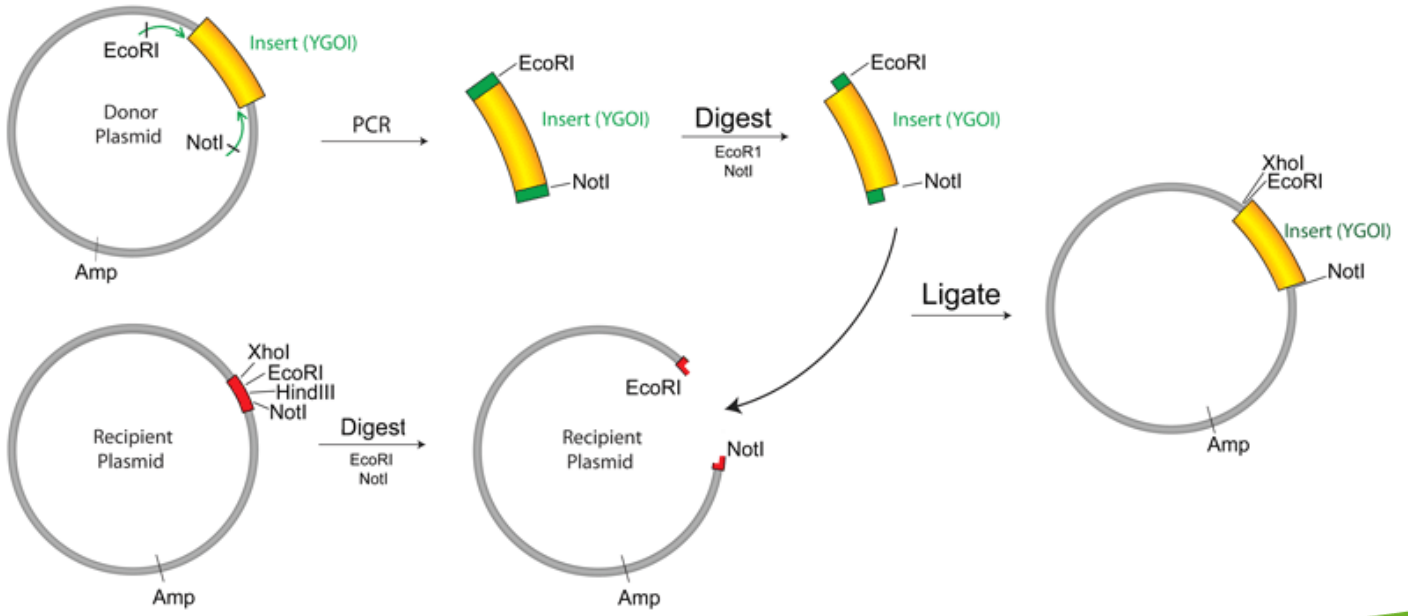
Carl Woese, 1928 - 2012

https://distributedmuseum.illinois.edu/exhibit/carl-woese/
https://en.wikipedia.org/wiki/Carl_Woese

Sanger Sequencing in use today

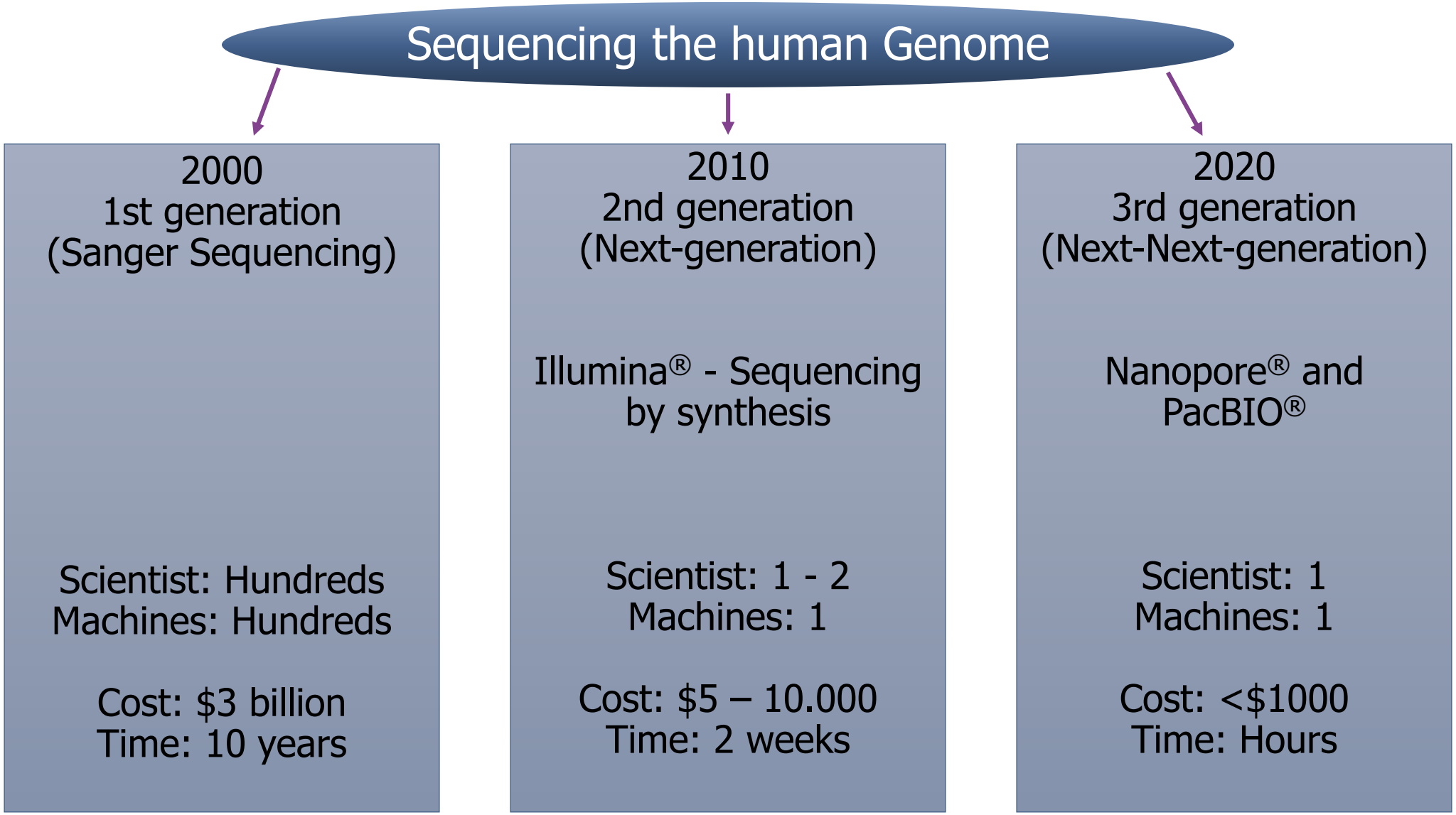
Still the method of choice for confirming correct insertion of gene fragements into plasmids

- Requires a primer site upstream of the insert site to act as starting point for the PCR reaction
- Read length ~1 kb



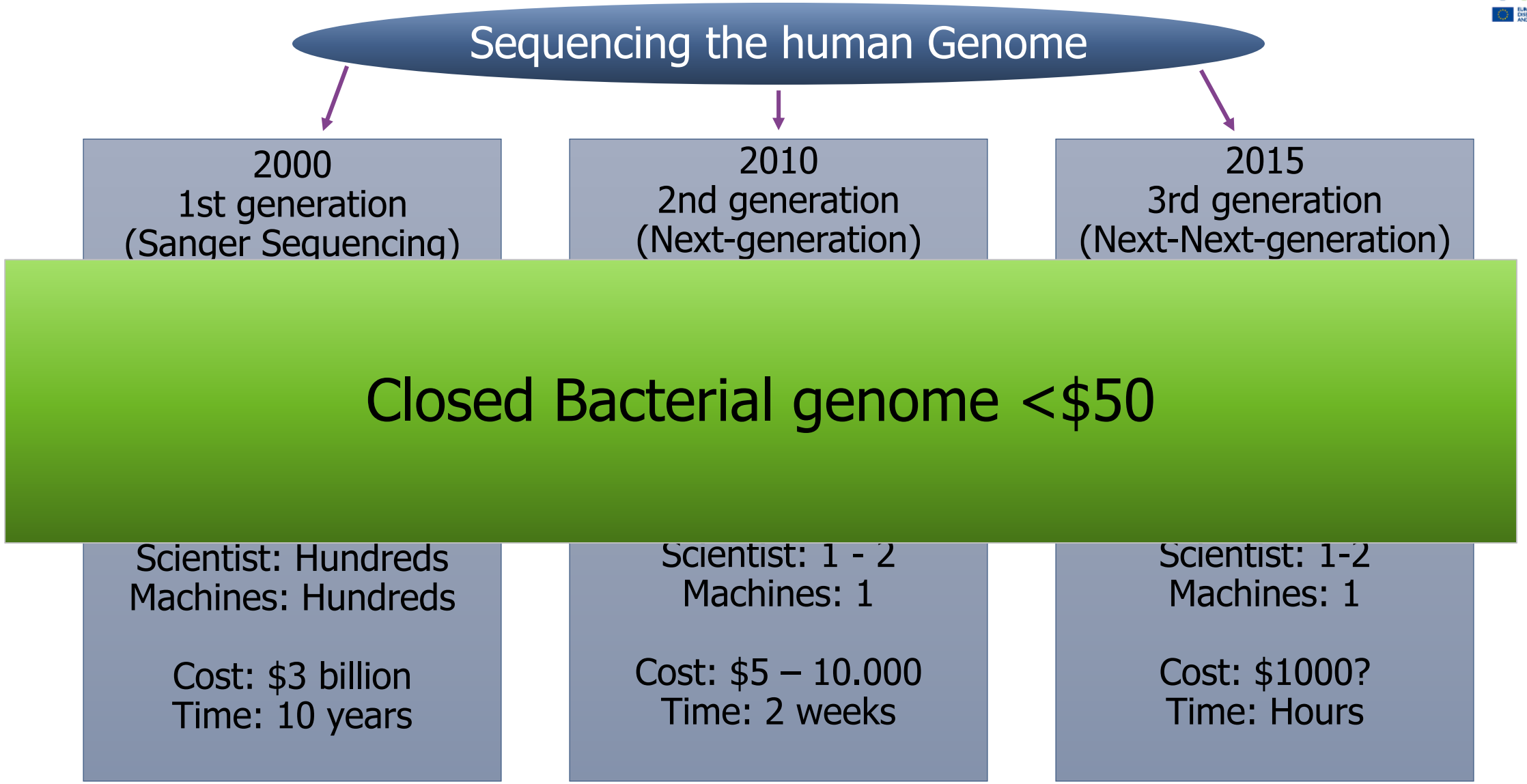
https://www.addgene.org/protocols/pcr-cloning/

The rapid evolution of the Technology



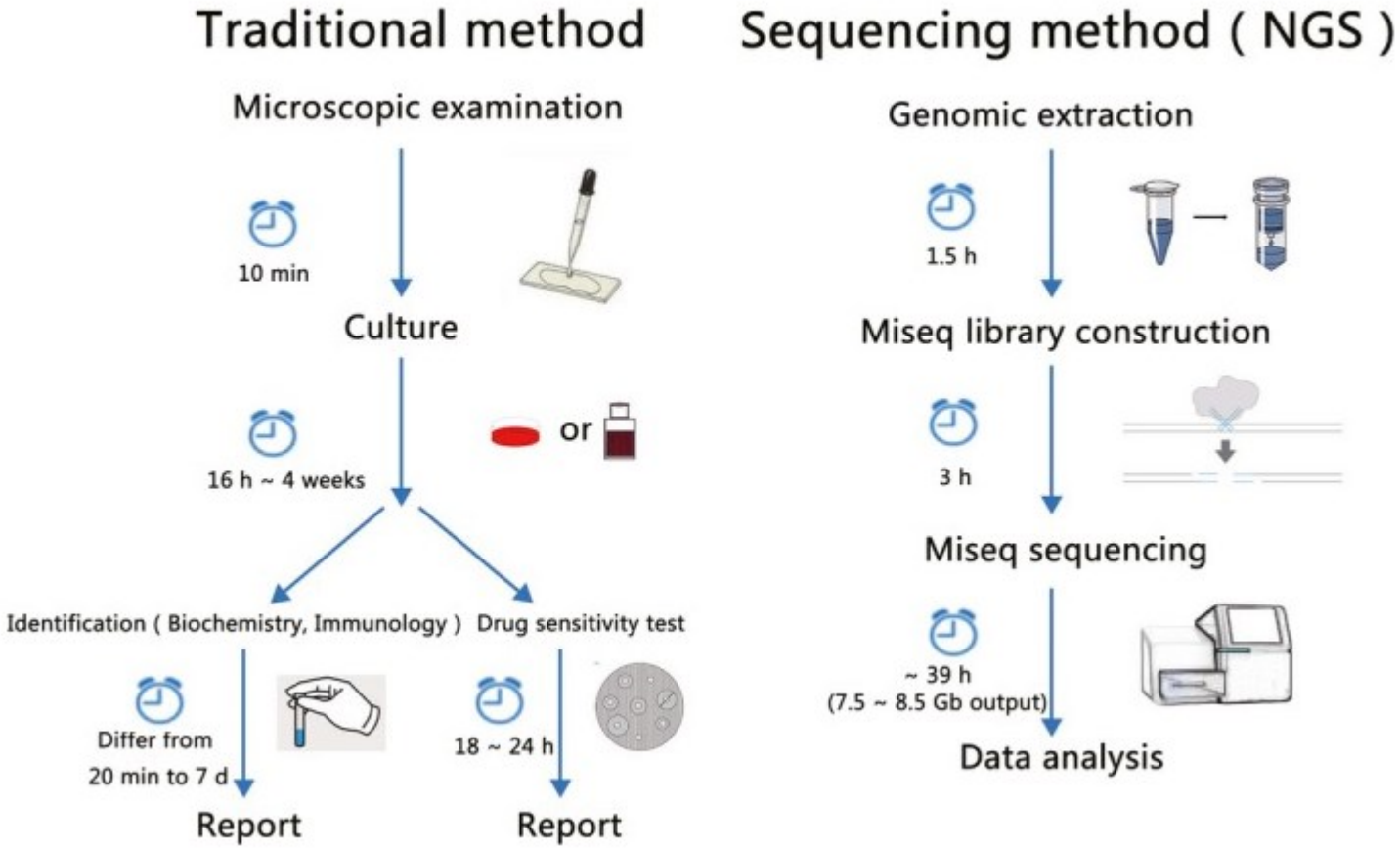
https://www.europeanpharmaceuticalreview.com/article/10409/dna-sequencing-technologies-and-emerging-applications-in-drug-discovery/

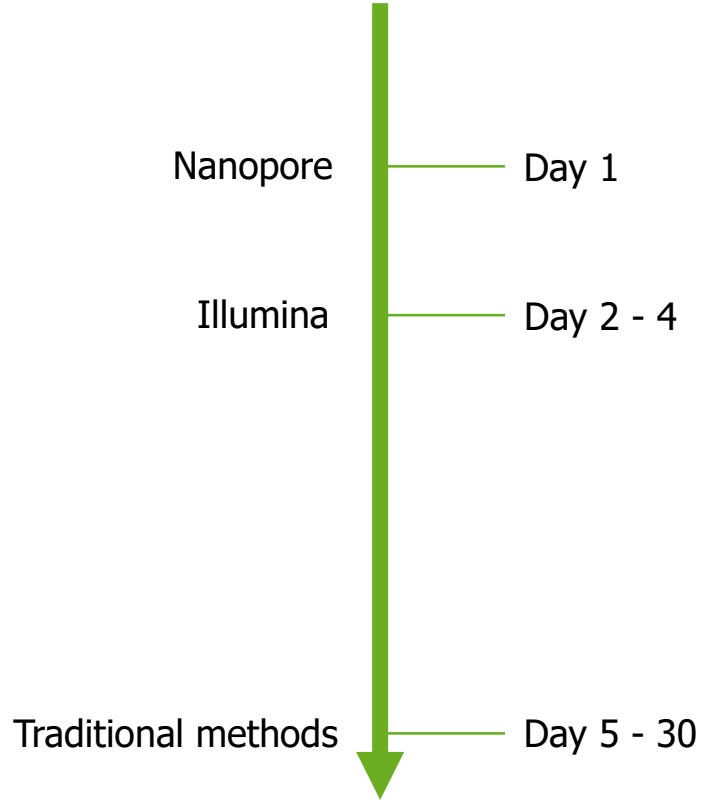
The rapid evolution of the Technology



https://www.europeanpharmaceuticalreview.com/article/10409/dna-sequencing-technologies-and-emerging-applications-in-drug-discovery/

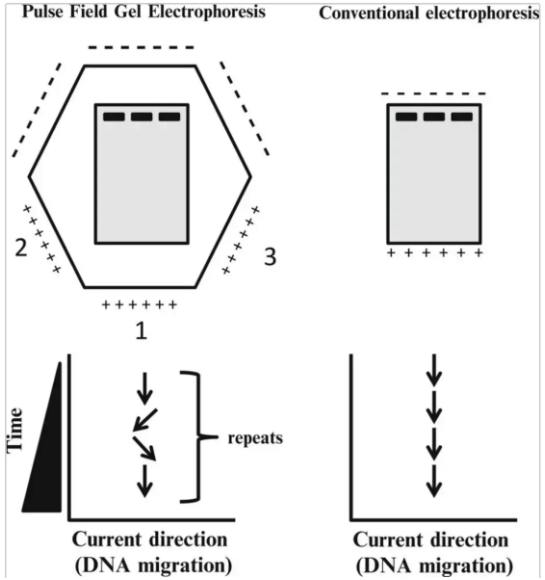
Why is sequencing an advantage - Phenotypic characterization

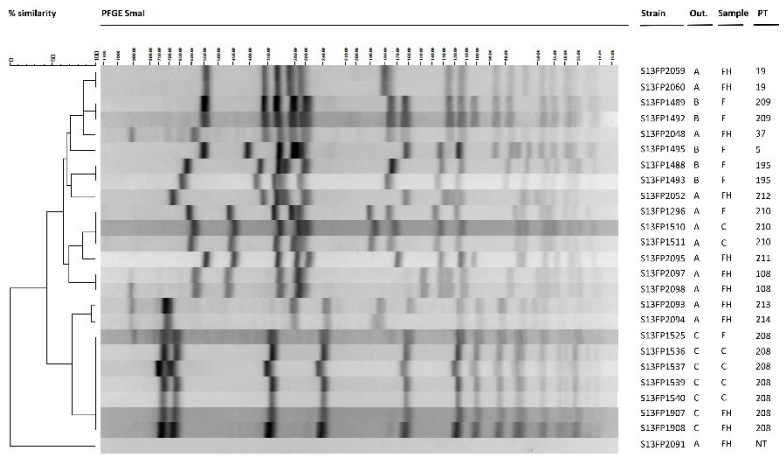


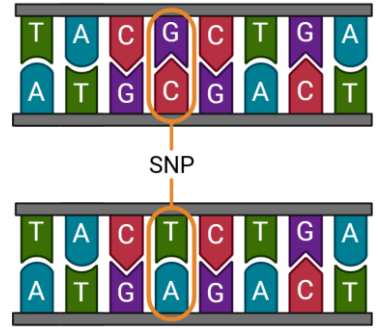


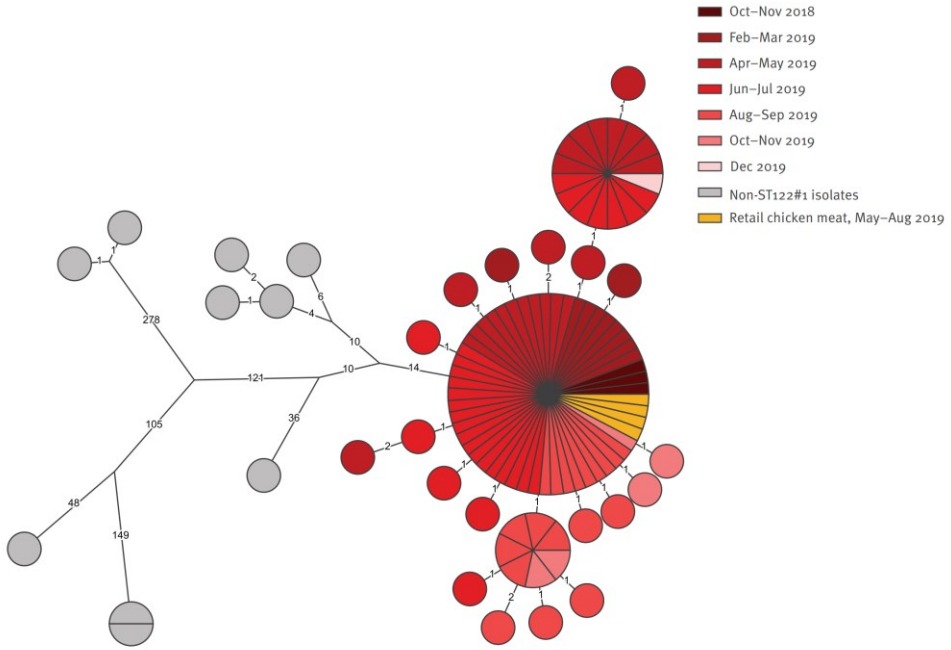
Yu *et al.*, 2019, J Cell Mol Med (CC licens)

Why is sequencing an advantage - Outbreak clusters of Foodborne pathogens







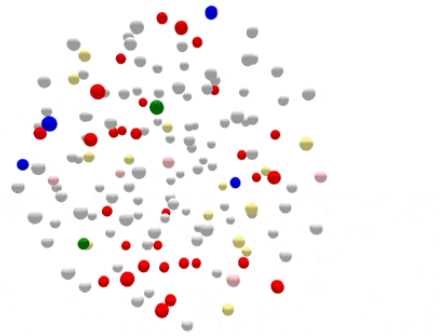


https://biologynotesonline.com/pulse-field-gel-electrophoresis-pfge-protocol/

Denayer, S. *et al.* (2017) *Toxins*
Joensen, KG, *et al* (2021) Euro Surveill

Why is sequencing an advantage - Outbreak clusters of Foodborne pathogens

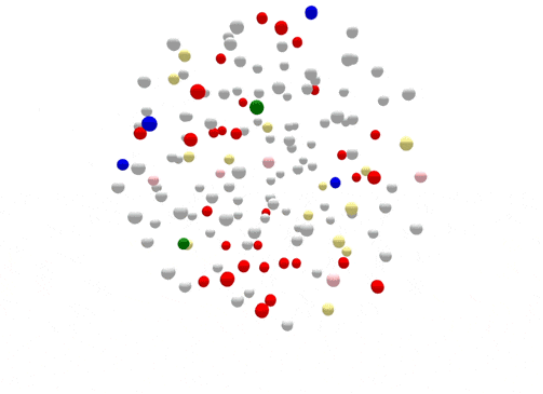
Unclustered



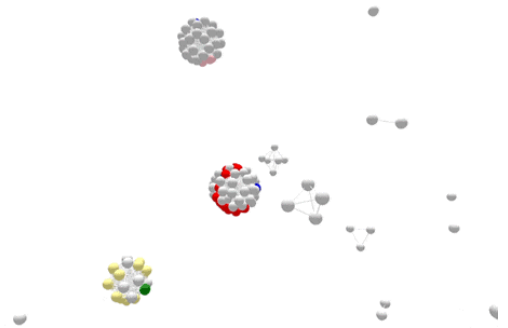
https://www.cdc.gov/advanced-molecular-detection/index.html

Why is sequencing an advantage - Outbreak clusters of Foodborne pathogens

Unclustered



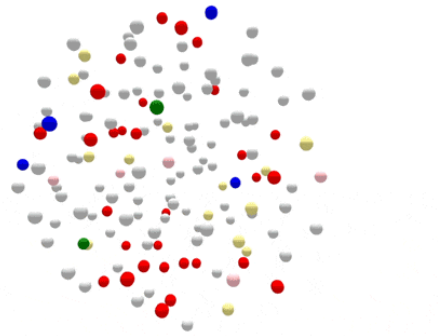
PFGE



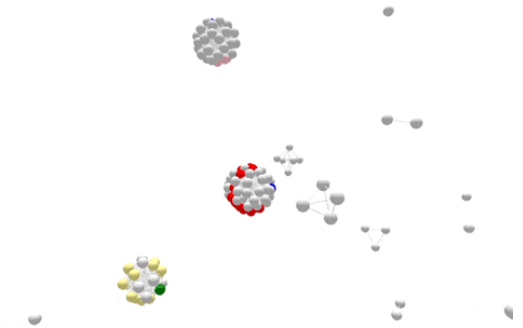
https://www.cdc.gov/advanced-molecular-detection/index.html

Why is sequencing an advantage - Outbreak clusters of Foodborne pathogens

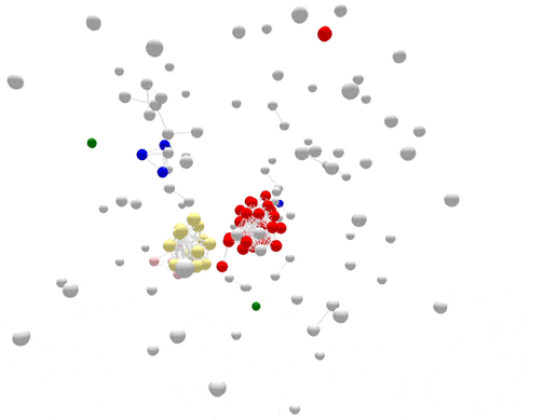
Unclustered



PFGE



WGS



https://www.cdc.gov/advanced-molecular-detection/index.html

When is sequencing an advantage

High resolution sequence typing (cgMLST)

- core genome multi locus sequence typing cgMLST

Outbreak detection

- SNP = Single Nucletide Polymorphism

Resistance specific genotype

- Plasmid medited resistance tracking

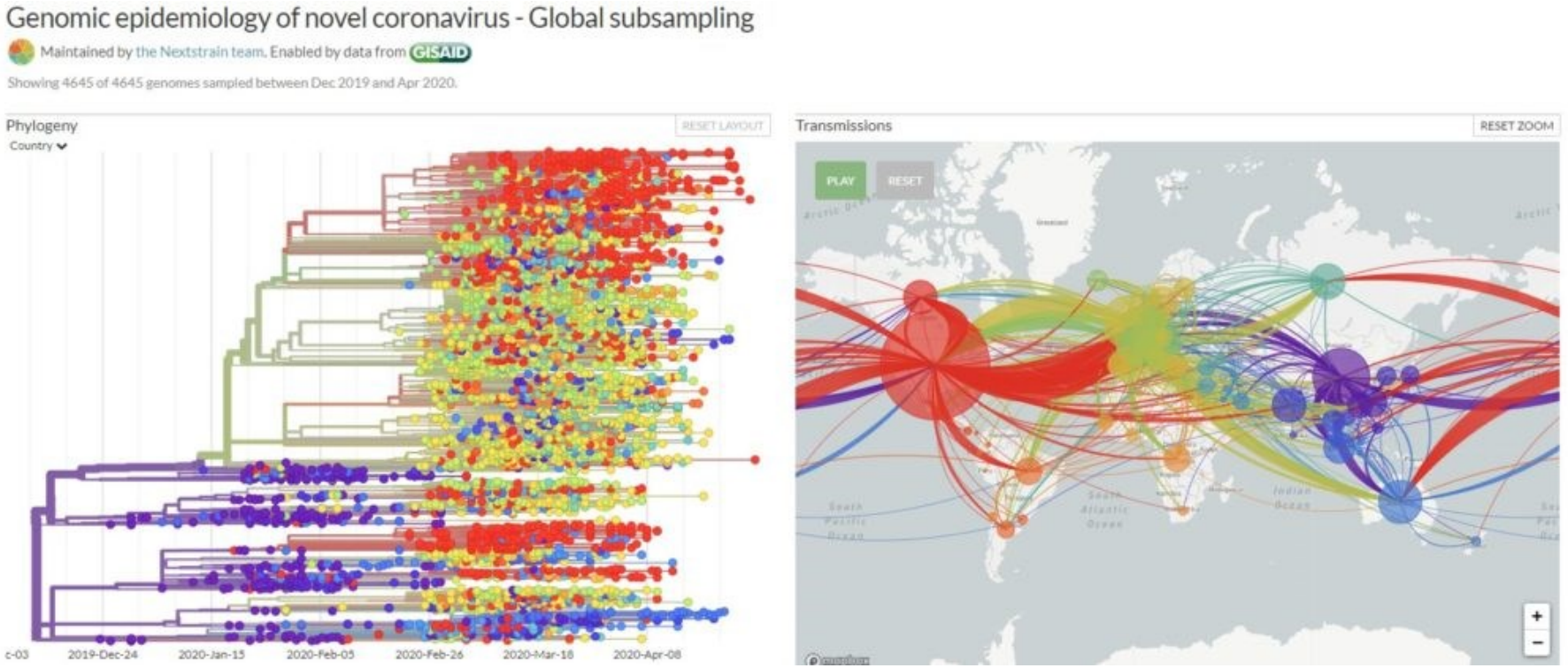
Genomic epidemiology

- Evolution and spread of clones

When is sequencing an advantage
The genomic evolution SARS-CoV-2

Genomic epidemiology

- Evolution and spread of clones



https://nextstrain.org/ncov/gisaid/global/6m

Selection of the most suited sequencing assay

Sanger Sequencing

- Short single gene fragments
- Genetic constructs (e.g. gene insertions into cloning vectors)
- Research

Massive parallel sequencing (Illumina)

- Short reads – High quality
- Genetic epidemiolgy
- SNP variant detection
- Research and Clinic

Long Read sequencing

- Closed genomes and plasmids
- Research… but moving towards clinical applications

Round off discussion

Which sequencing platforms do you have available?

Key take home

DNA sequencing is fundamental to modern surveillance of infectious disease, outbreak cluster detection, and genomic epidemiology

To select the right technology for the right task one need to think about

- The sample
- The aim
- ►The technology

Further reading

Yu X, Jiang W, Shi Y, Ye H, Lin J. Applications of sequencing technology in clinical microbial infection. J Cell Mol Med. 2019 Nov;23(11):7143-7150. doi: 10.1111/jcmm.14624. Epub 2019 Sep 2. PMID: 31475453; PMCID: PMC6815769.

Peters TM. Pulsed-field gel electrophoresis for molecular epidemiology of food pathogens. Methods Mol Biol. 2009;551:59-70. doi: 10.1007/978-1-60327-999-4_6. PMID: 19521867.

Joensen KG, Schjørring S, Gantzhorn MR, Vester CT, Nielsen HL, Engberg JH, Holt HM, Ethelberg S, Müller L, Sandø G, Nielsen EM. Whole genome sequencing data used for surveillance of Campylobacter infections: detection of a large continuous outbreak, Denmark, 2019. Euro Surveill. 2021 Jun;26(22):2001396. doi: 10.2807/1560-7917.ES.2021.26.22.2001396. PMID: 34085631; PMCID: PMC8176674.

Acknowledgements

The creation of this training material was commissioned by ECDC to Statens Serum Institute (SSI) with the direct involvement of Søren Hallstrøm