Bridging the gaps in bioinformatics/Raw data QC

Introduction to sequencing

February 2024, Søren Hallstrøm, Statens Serum Institut, Denmark

Outline

This session consists of the following elements

1. Introduction to Sequencing
2. Brief summary of the evolution of sequencing technology
3. Basics of sequencing in clinical microbiology

Objectives

Specific objectives of this session:

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

The Lecturer – Background

Søren Hallstrøm

Ph.D. in molecular microbiology

Extensive experience with development of custom workflows

- Amplicon sequencing (16S, functional gene targets)
- Whole Genome Sequencing (WGS)

The Lecturer - Presently

Academic staff at the Sequencing Core Facility -
NGS laboratory at Statens Serum Institute, Denmark
Department for infectiuos desease surveillance

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

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.

Finding the precise order of the nucleotides in a DNA fragment is very important to know about the structure and function of the genes.

Sequencing fundamentals

Nucleotide Extraction

DNA

Sequencing Library preparation

A brief history of sequencing technologies

1st generation Sanger Sequencing

Chain termination
Single gene fragments

First sequence of a DNA genome: bacteriophage φX174, in 1977

Q20 - Q30 data

2nd generation Illumina

Sequencing by Synthesis
Parallel sequencing
Complex DNA Libraries

Illumina (Solexa)
On board amplification of sequencing library (Bridge amplificaion

Q30 – Q40 data

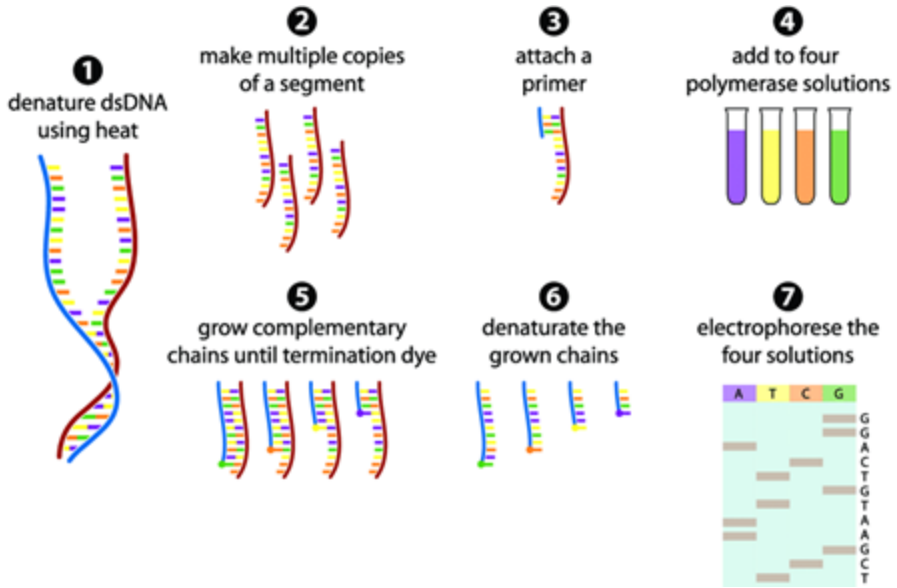
3rd generation Nanopore

Long read sequencing
Real time basecalling

Oxford Nanopore Technologies
Low price -Low quality
Q10-20 data

Pacific Biosciences
PacBIO
High price -> High quality
Q30-40 data

1st generation Sanger sequencing



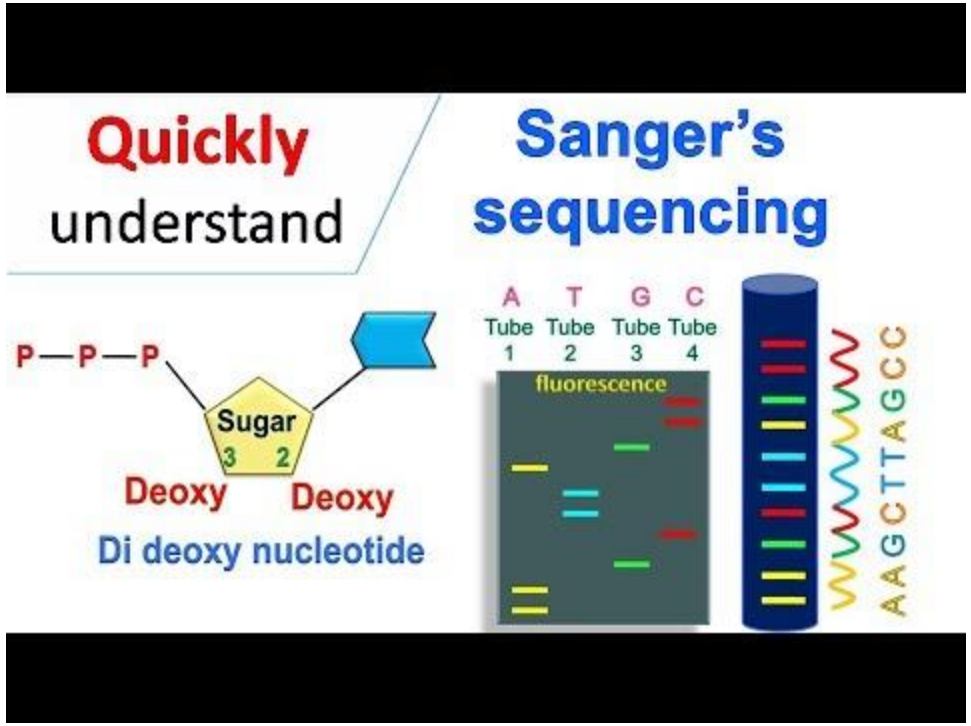
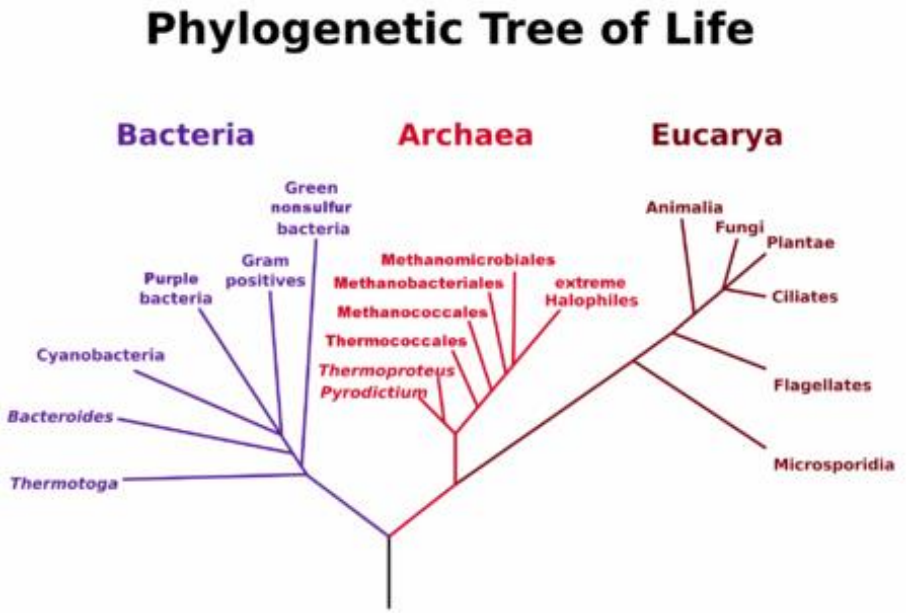


Image credit: Canva
https://frontlinegenomics.com/a-history-of-sequencing/#:~:text=The%20first%20major%20breakthrough%20in,him%20his%20second%20Nobel%20Prize.
https://www.youtube.com/watch?v=KTstRrDTmWI

Sanger sequencing – The tree of life and the third domain



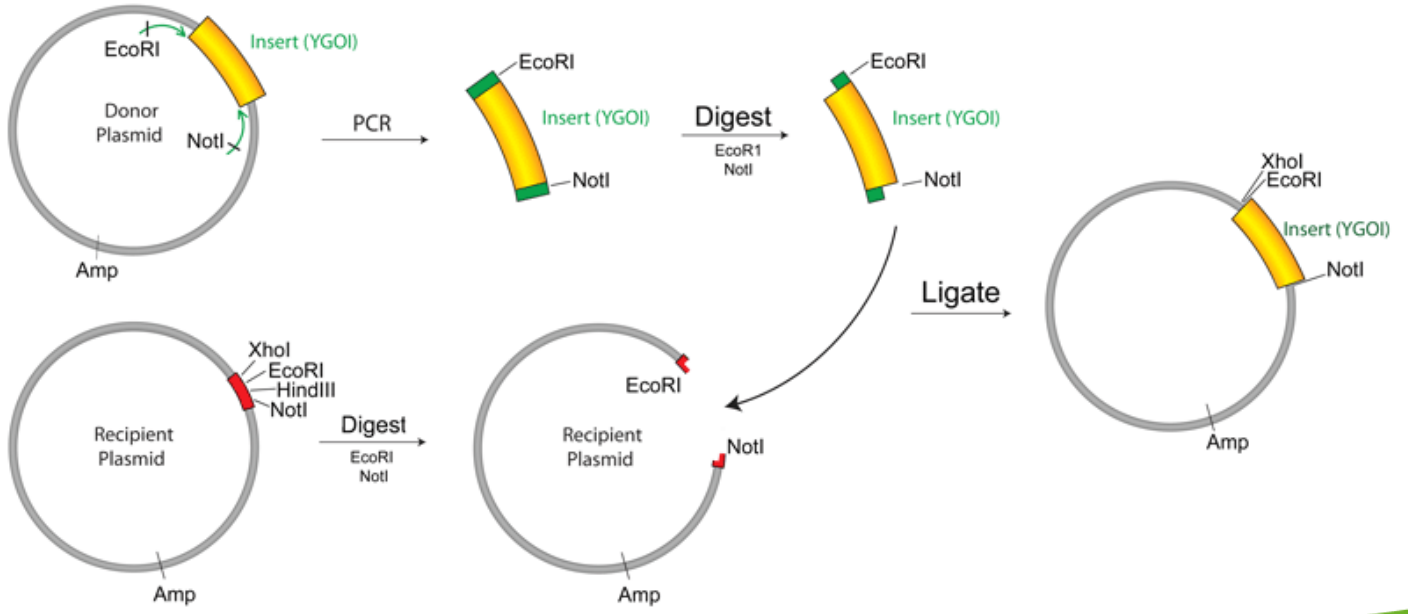
Carl Woese, 1928 - 2012

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

Sanger Sequencing in use today

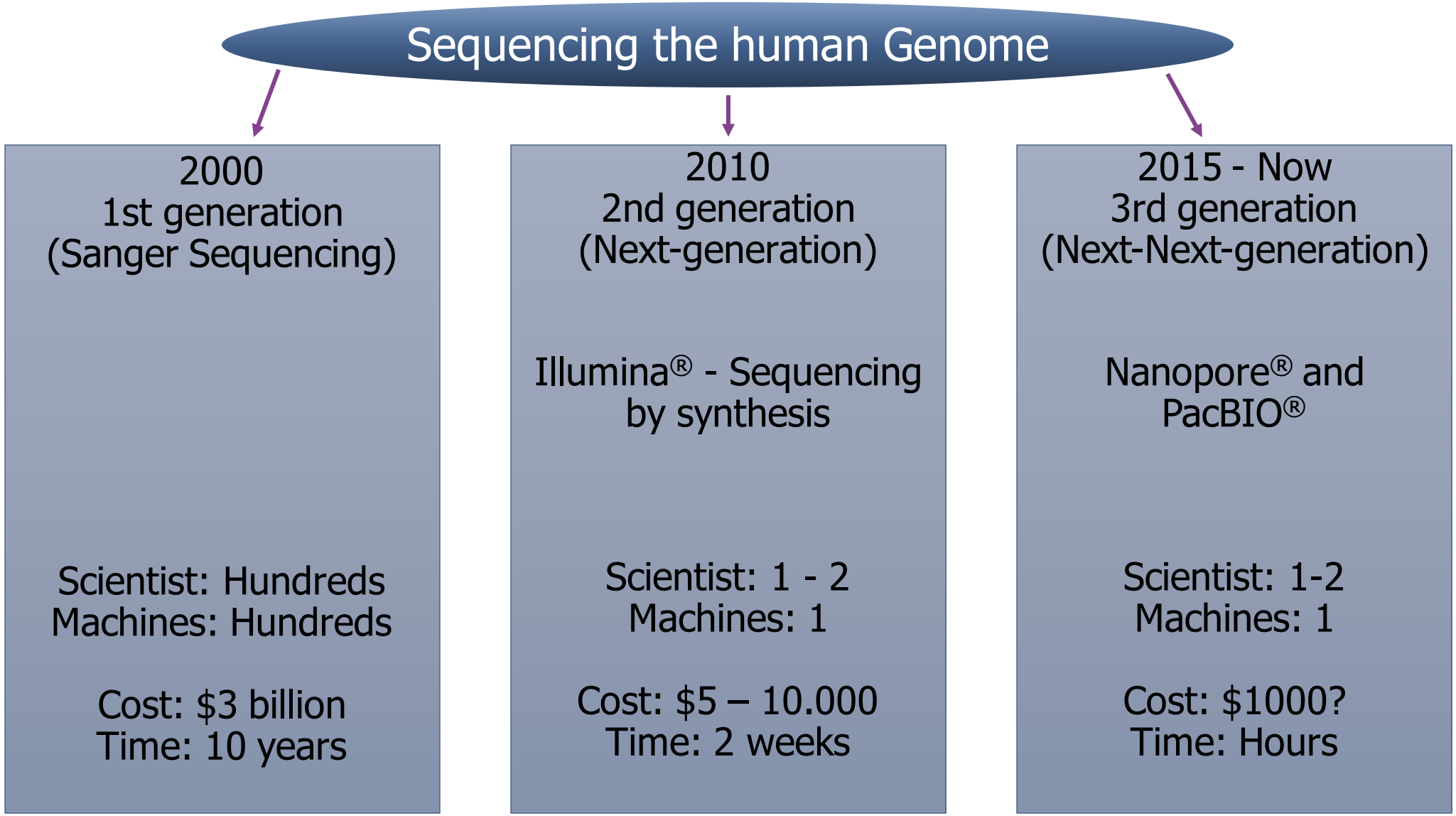
Still the method of choice for confirming correct insertion of gene fragements into plasmids for cloning

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



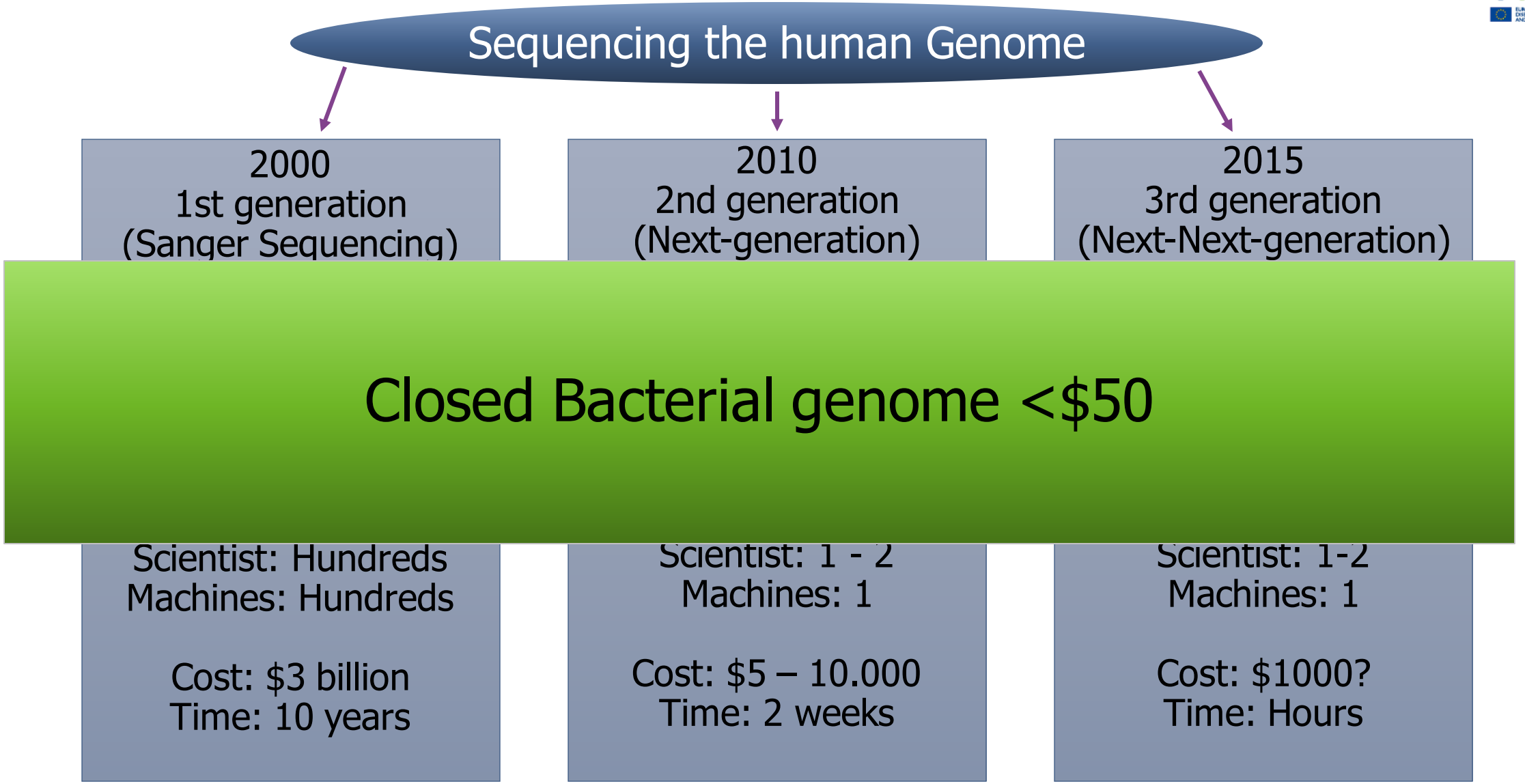
https://www.addgene.org/protocols/pcr-cloning/

The rapid evolution of the Technology



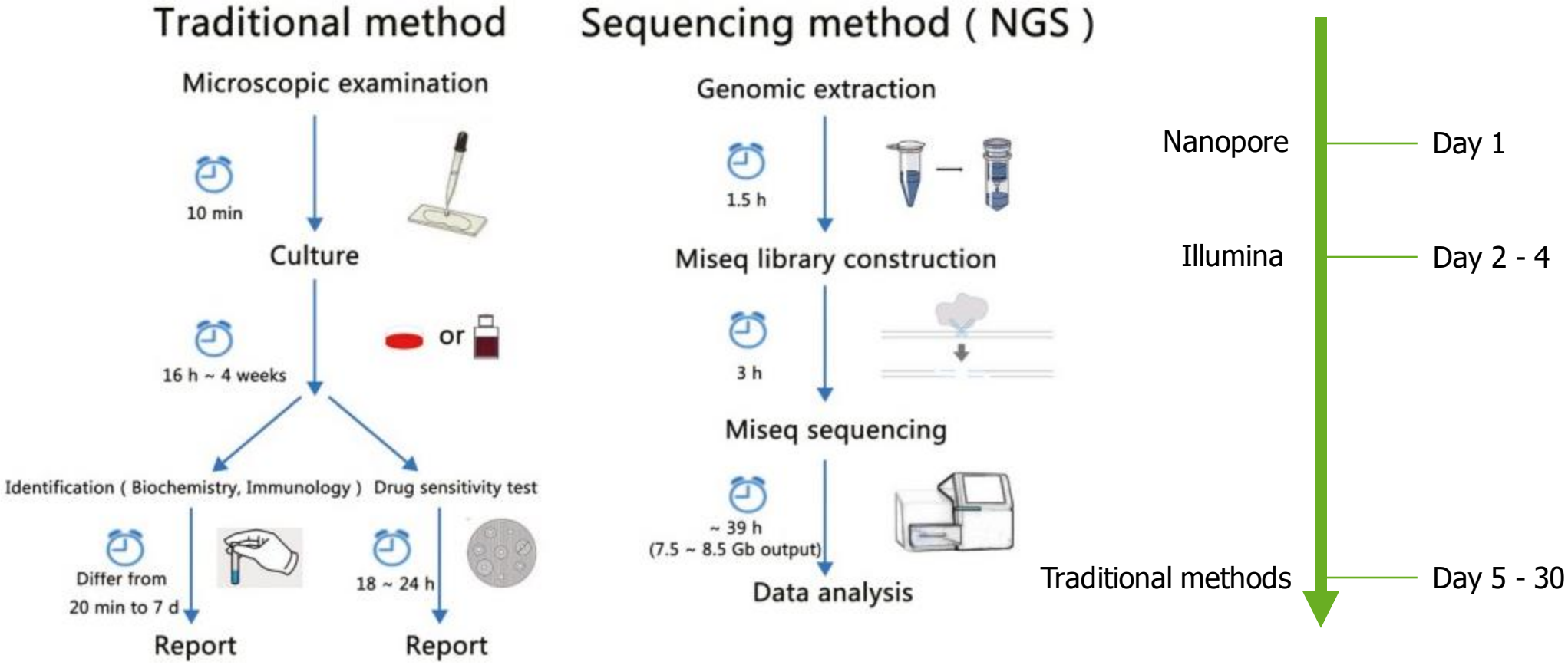
https://www.europeanpharmaceuticalreview.com/article/10409/dna-sequencing-technologies-and-emerging-applications-in-drug-discovery/

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 - example from a clinical setting



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

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 resistancetracking

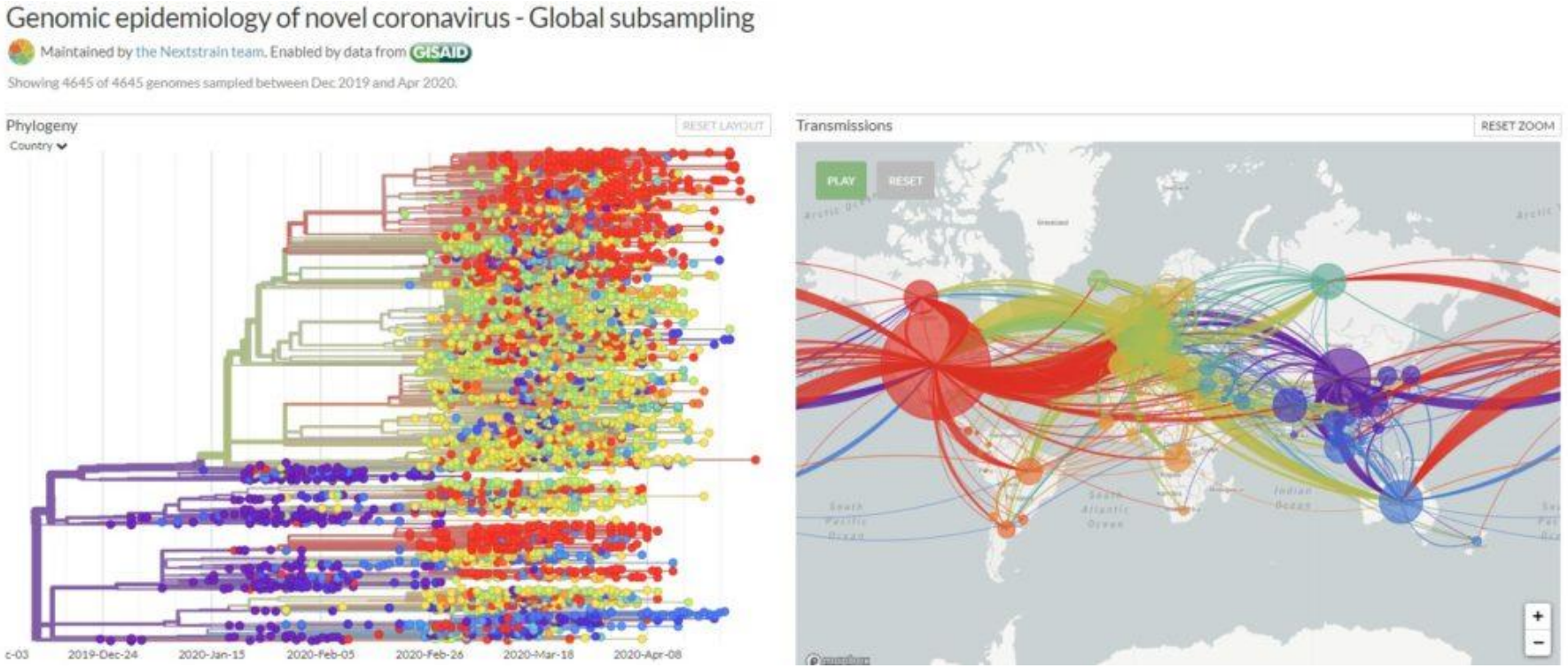
Genomic epidemiology

- 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(?)

Which sequencing platforms do you have available?

Key take home

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

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

- The technology
- The sample
- The aim

Acknowledgements

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