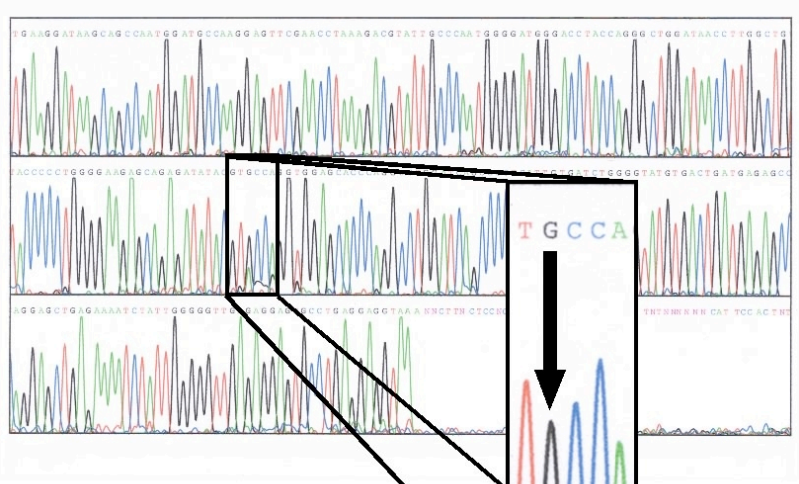
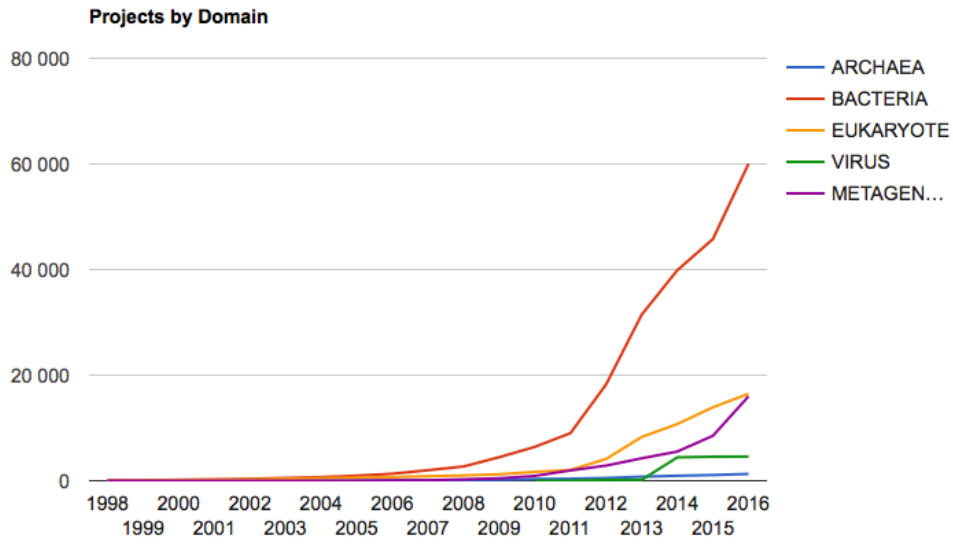
GENOME SEQUENCING

Introduction

Sequencing the genomes

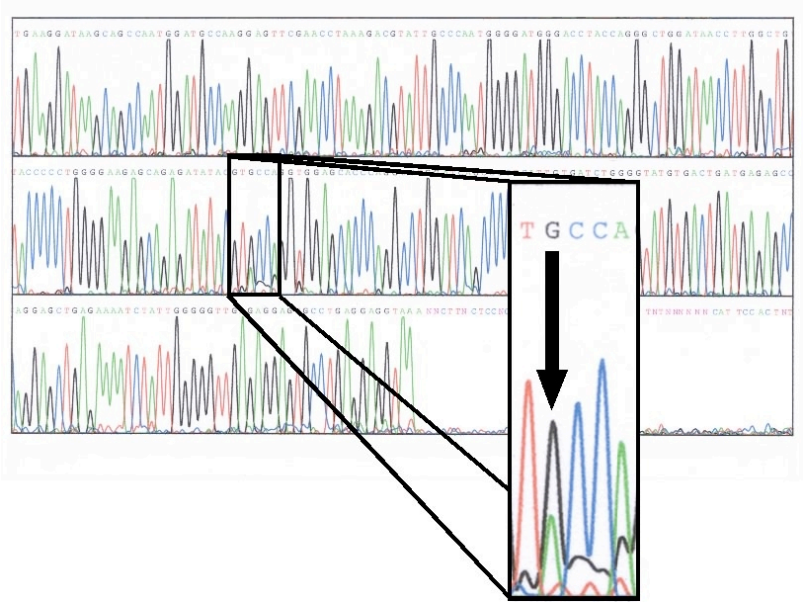




http://www.yourgenome.org/facts/timeline-organisms-that-have-had-their-genomes-sequenced

Introduction

Sequencing the genomes



<u>Why sequencing these genomes?</u>

Complete inventory of the genes of various organisms

Information about the sequence of every polypeptides

Basis for new technologies

Global data at the level of the whole cell

New data for evolutionnary and species classification analysis

Medical applications

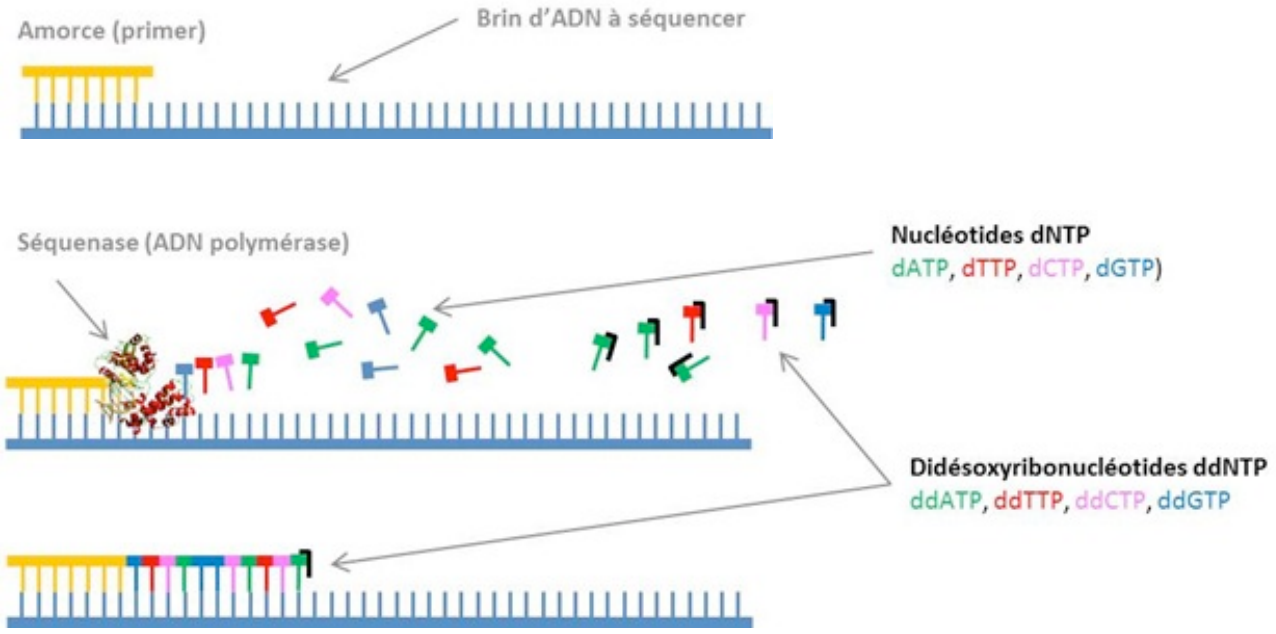
I. Genome sequencing

The Sanger Method (1977)

Denaturation
+priming

Polymerization

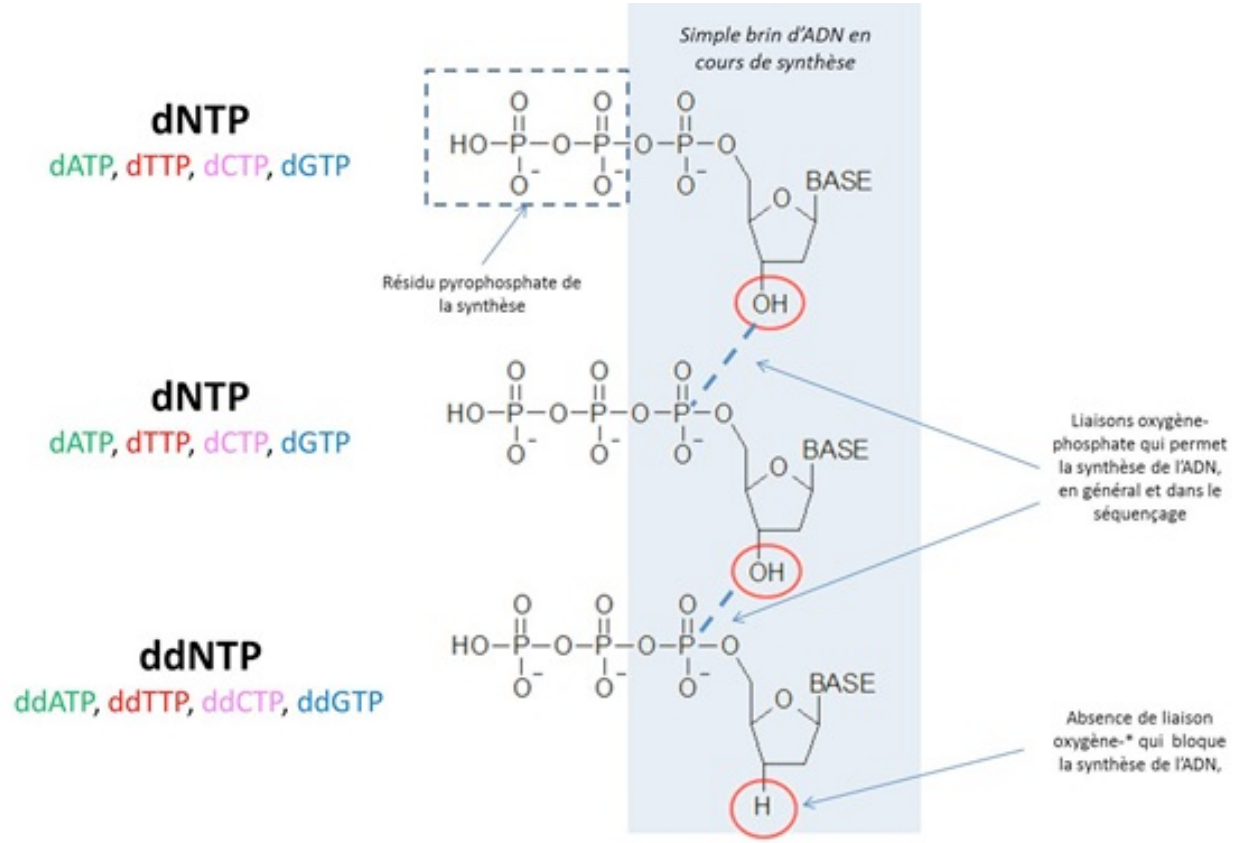
Arrest



I. Genome sequencing

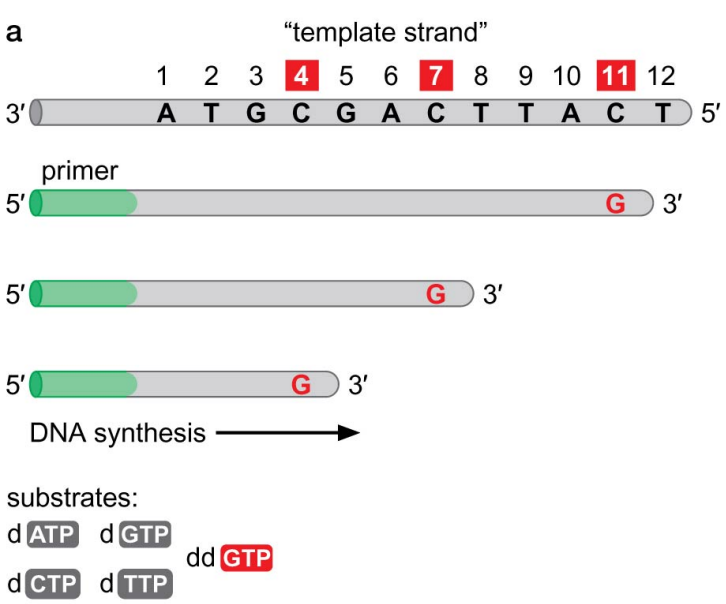
The Sanger Method (1977)

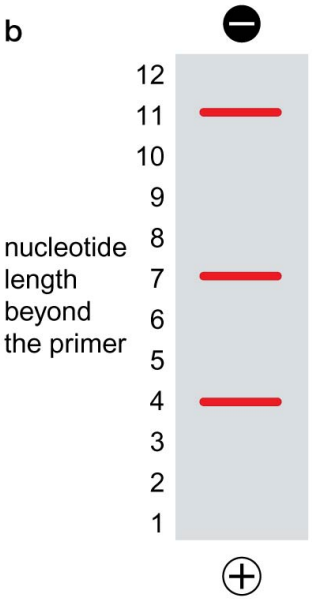
End of polymerization

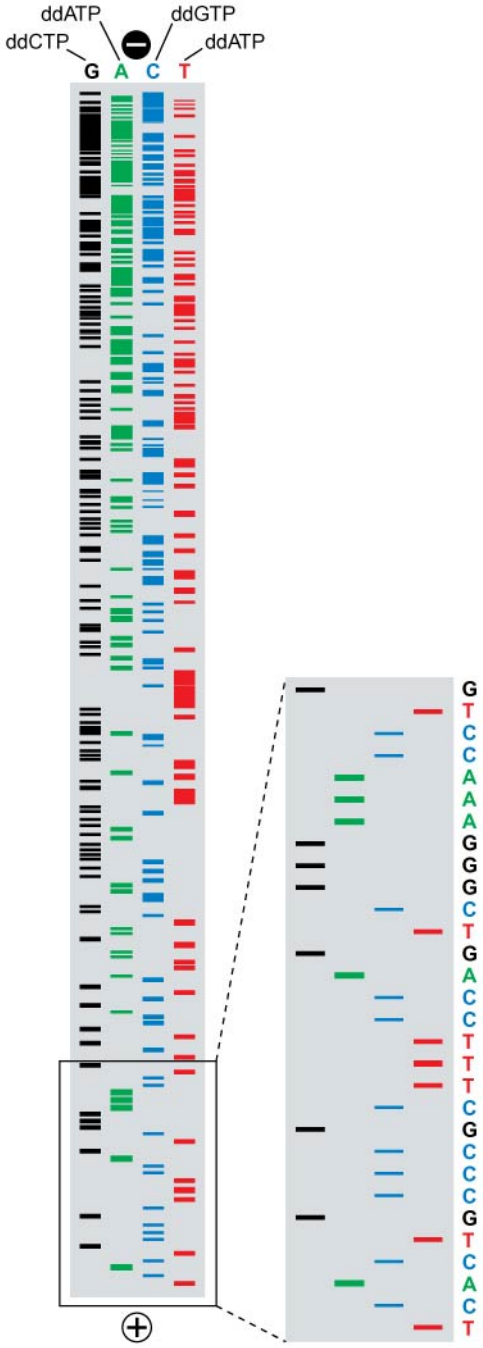


I. Genome sequencing

The Sanger Method (1977)







I. Genome sequencing

The caveats of the Sanger Method

Technically cumbersome and time consuming (limited samples processing)

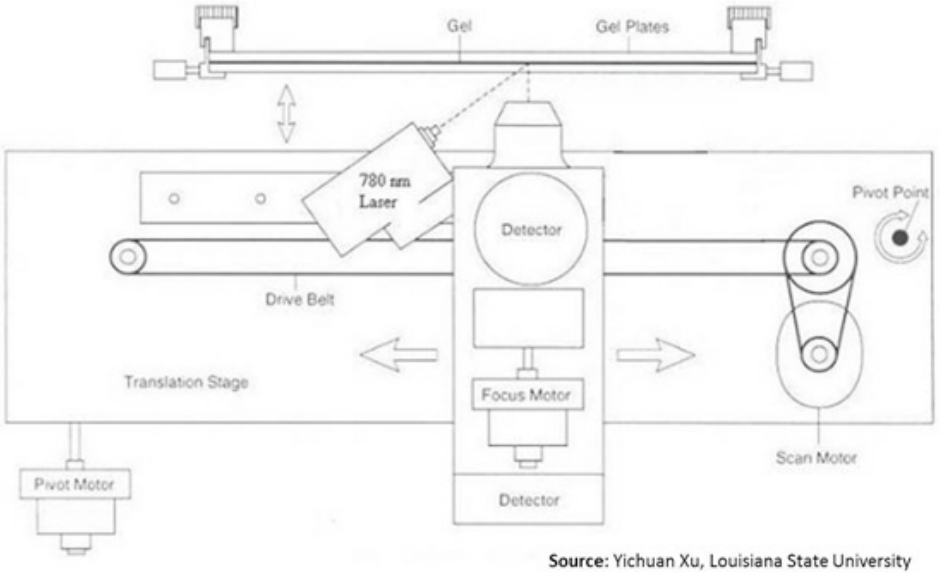
High amounts of starting material

➔Not well-suited for high-throughput analyses (genome sequencing)

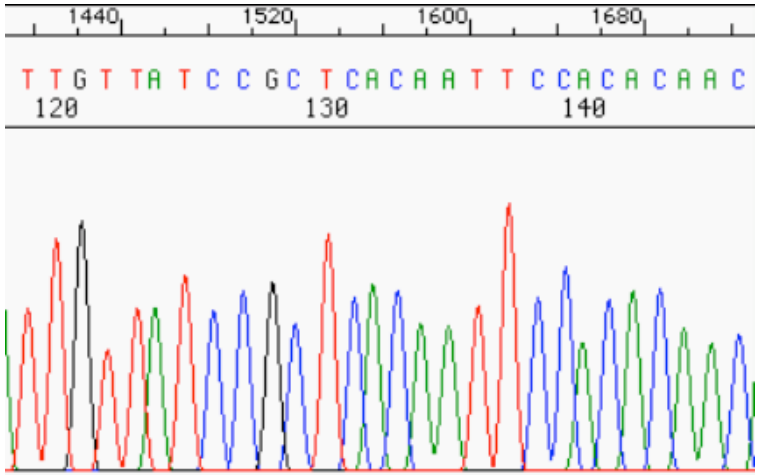
I. Genome sequencing

The Sanger Method (1977)

1st generation sequencing : Improvement of sequence analysis (column electrophoresis)



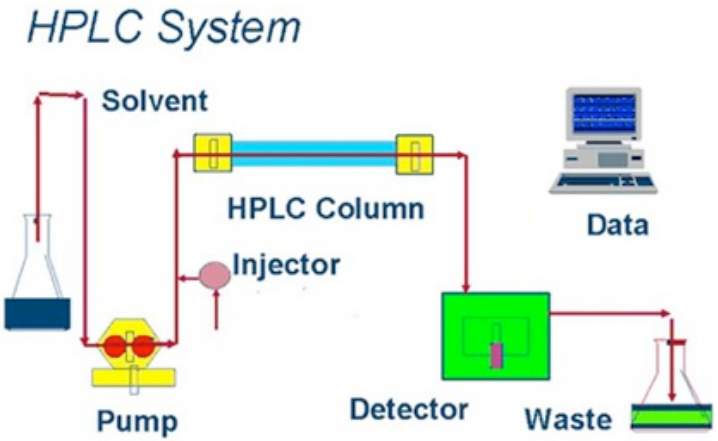
Source: Yichuan Xu, Louisiana State University



I. Genome sequencing

The Sanger Method (1977)

1st generation sequencing : Improvement of sequence analysis (capillary electrophoresis)



Source: College of Arts and Sciences, New Mexico State University

Agilent 1290 Infinity Quaternary LC System

PerkinElmer Flexar FX-15 UHPLC

Nexera UHPLC

Dionex UltiMate 3000 Standard LC Systems

I. Genome sequencing

1st generation sequencing : Improvement of the methodology

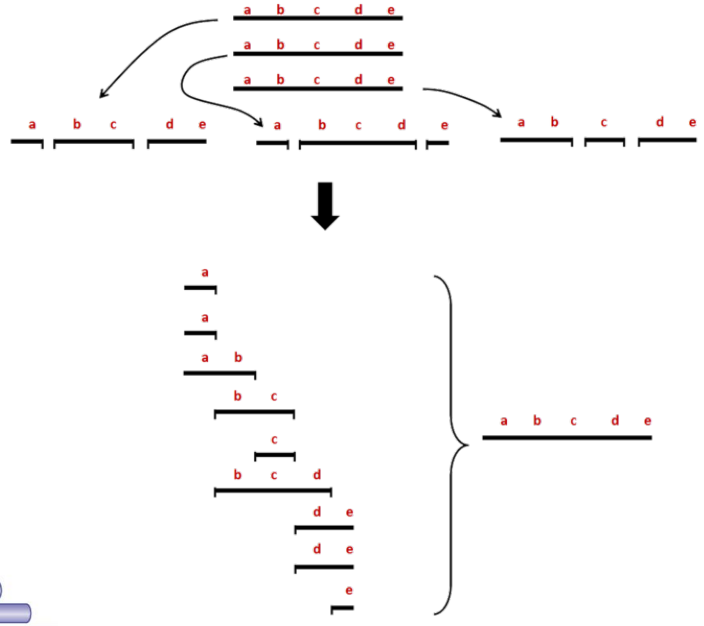
Parallel random sequencing : after fragmentation (shotgun sequencing) and cloning

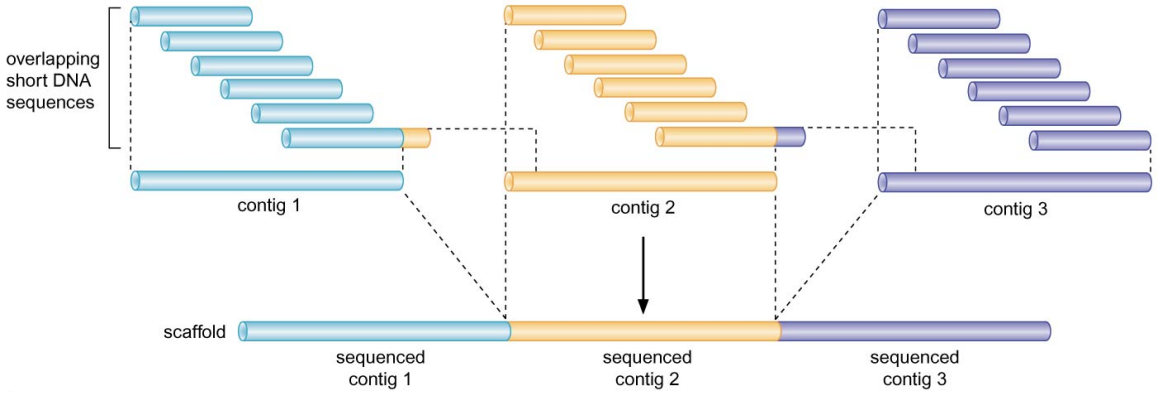
- genomic DNA is fragmented
- cloned to a plasmid vector -> transform

E. coli

- a single bacterial colony is picked ->

plasmid DNA isolated

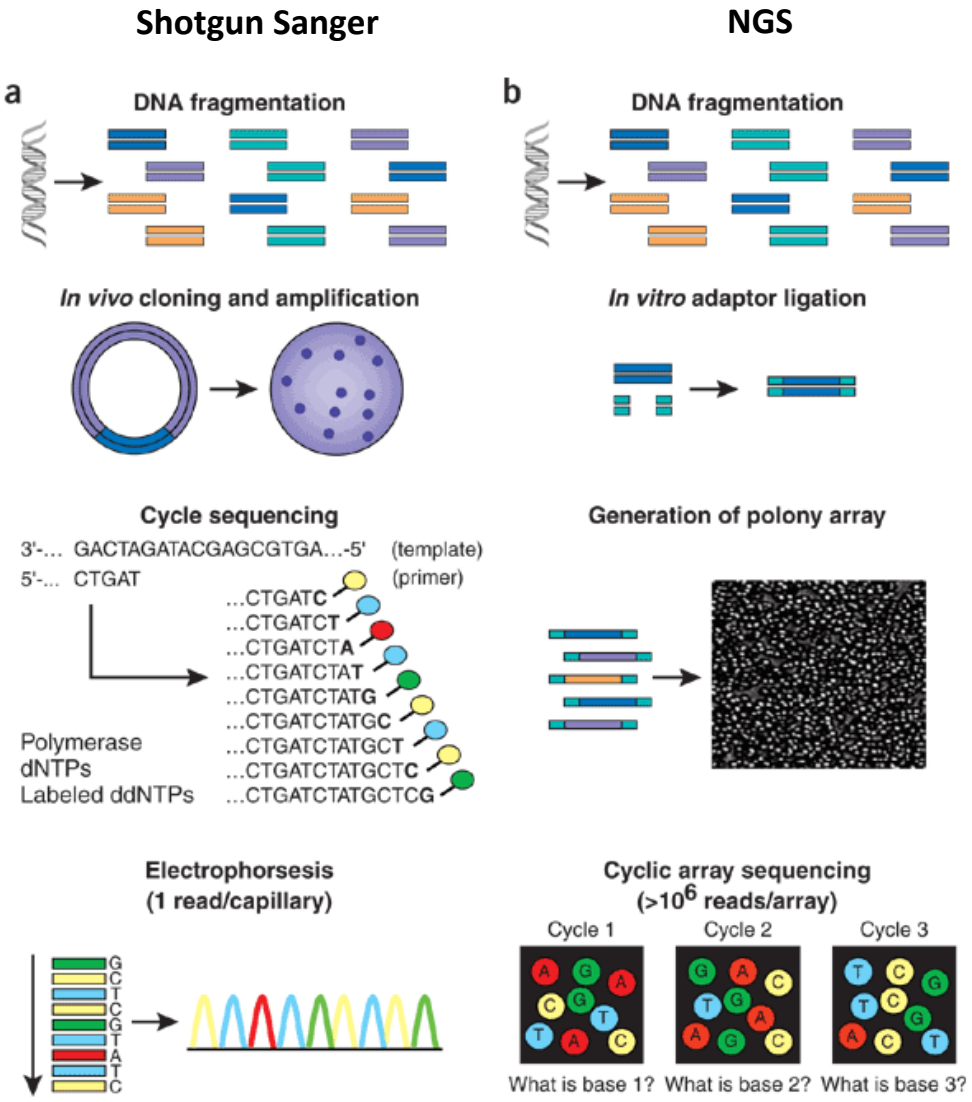




I. Genome sequencing

2nd generation and next generation (NGS) sequencing : Areas of improvement

- Template preparation
  amplification by PCR
- Sequencing and imaging
- Data analysis

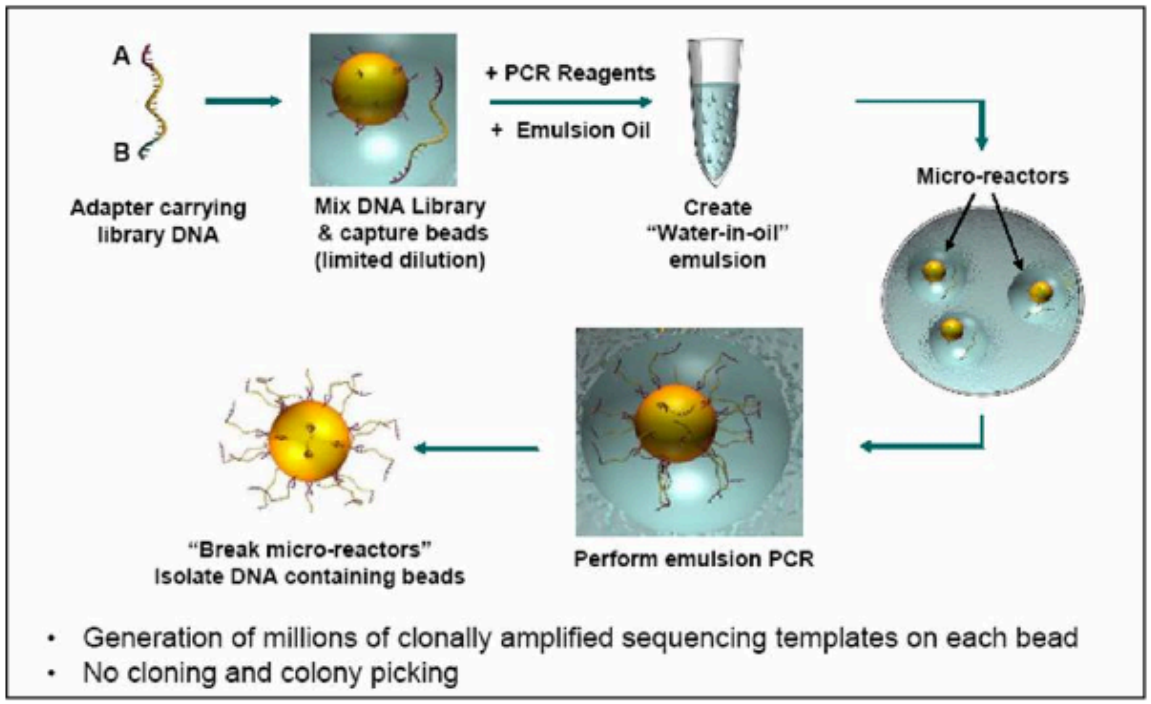


I. Genome sequencing

2nd generation sequencing/NGS :

Improvement of template preparation : amplification by PCR

Emulsion PCR (emPCR)



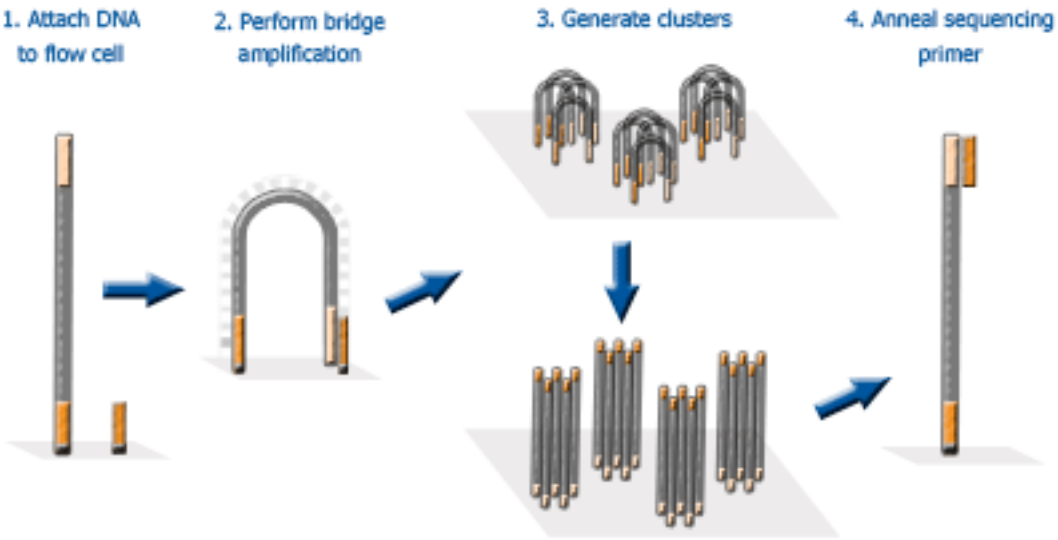
Used by 454 (life science), Polonator, Ion torrent and Solid

I. Genome sequencing

2nd generation sequencing :

Improvement of template preparation : amplification by PCR

Bridge PCR (illumina plateform)



I. Genome sequencing

2nd generation sequencing :

Improvement of template preparation : Single Molecule Sequencing (SMRT)

- Requires less strating material
- Provides longer contigs

■ Immobilized on the solid surface by

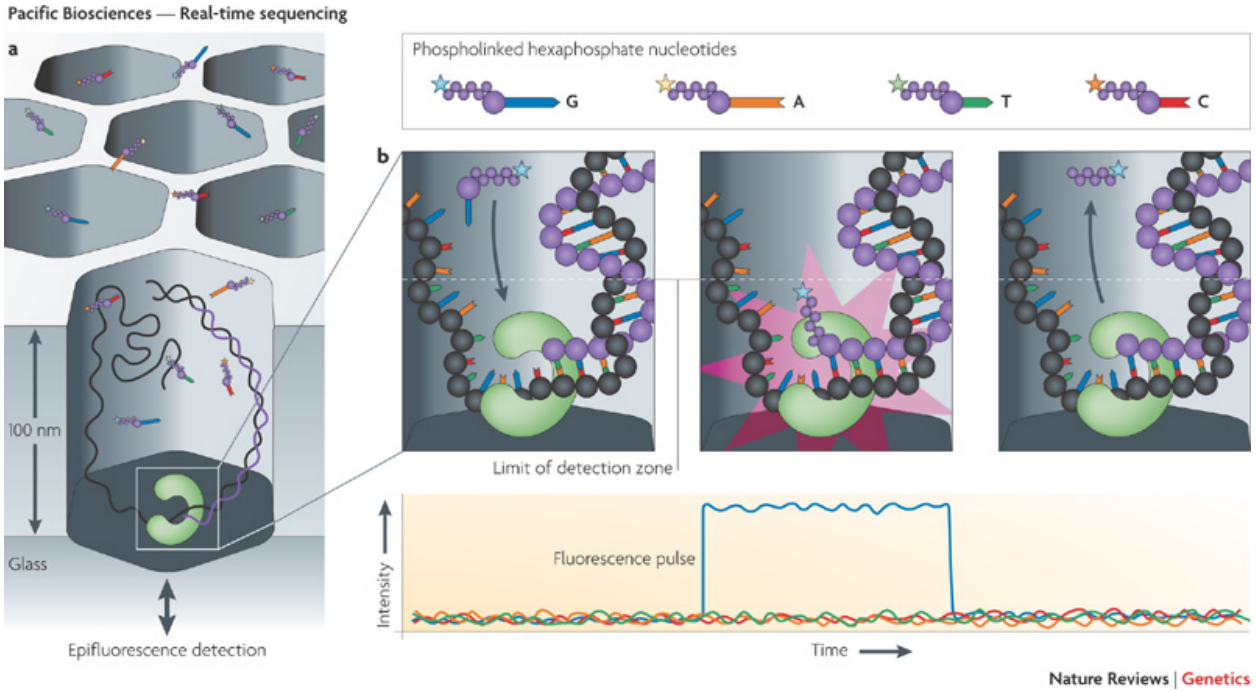
Primers: Helicos BioSciences
Template: Helicos BioSciences
Polymerase: Pacific Biosciences, Life/Visigen, LI-COR Biosciences

I. Genome sequencing

2nd generation sequencing : real-time sequencing

SMRT Sequencing Technology : Single-Molecule Real-Time Sequencing

A true continous imaging of dye-labelled nucleotides incorporation

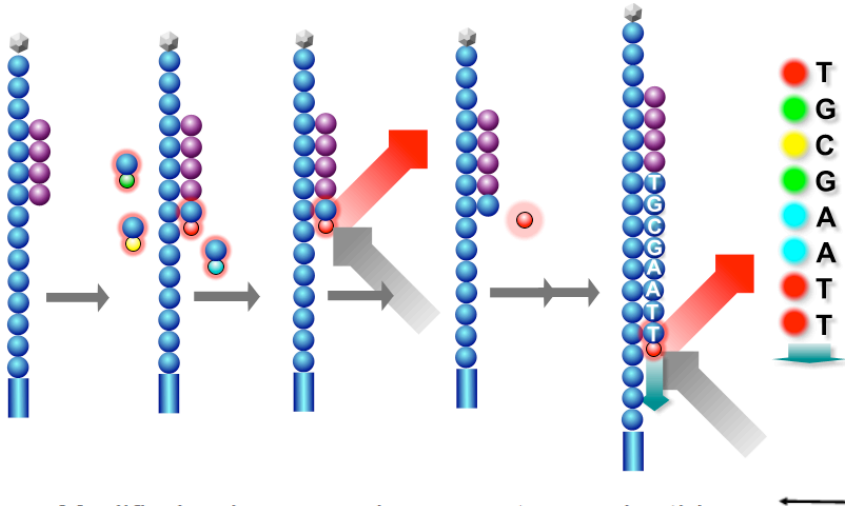


[Pacific Biosciences] : SMRT : Single Molecule Real-Time Sequencing

I. Genome sequencing

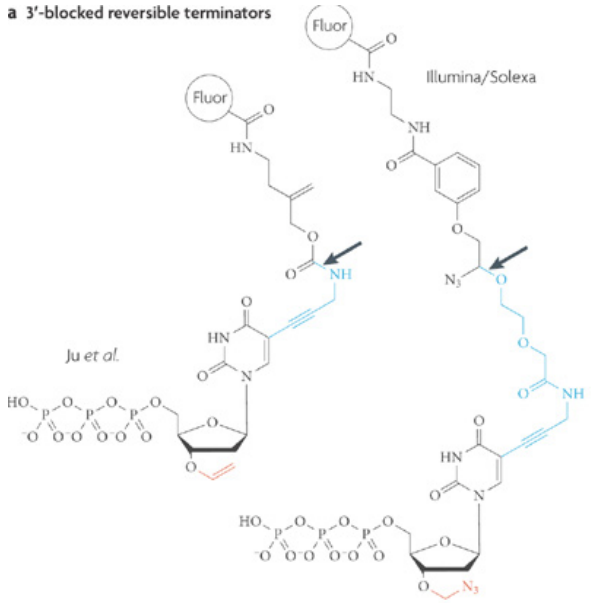
2nd generation sequencing : real-time sequencing

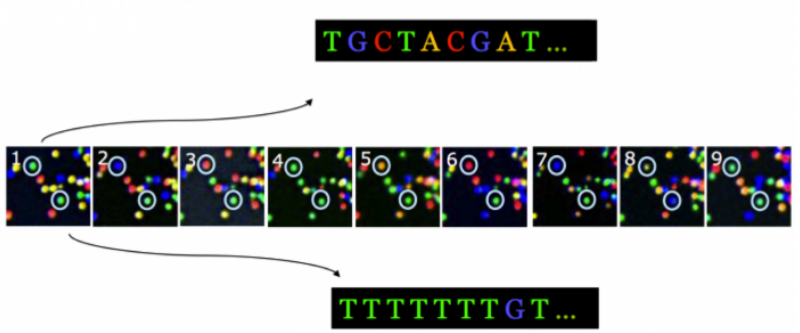
Sequencing **by reversible terminator (illumina/Solexa genome Analyzer)**



Modified polymerase incorporates nucleotides

- after each nucleotide incorporation process stops
- camera reads fluorophore signal (filter for each nucleotide type)
- terminator and labeling is removed and cycle starts again





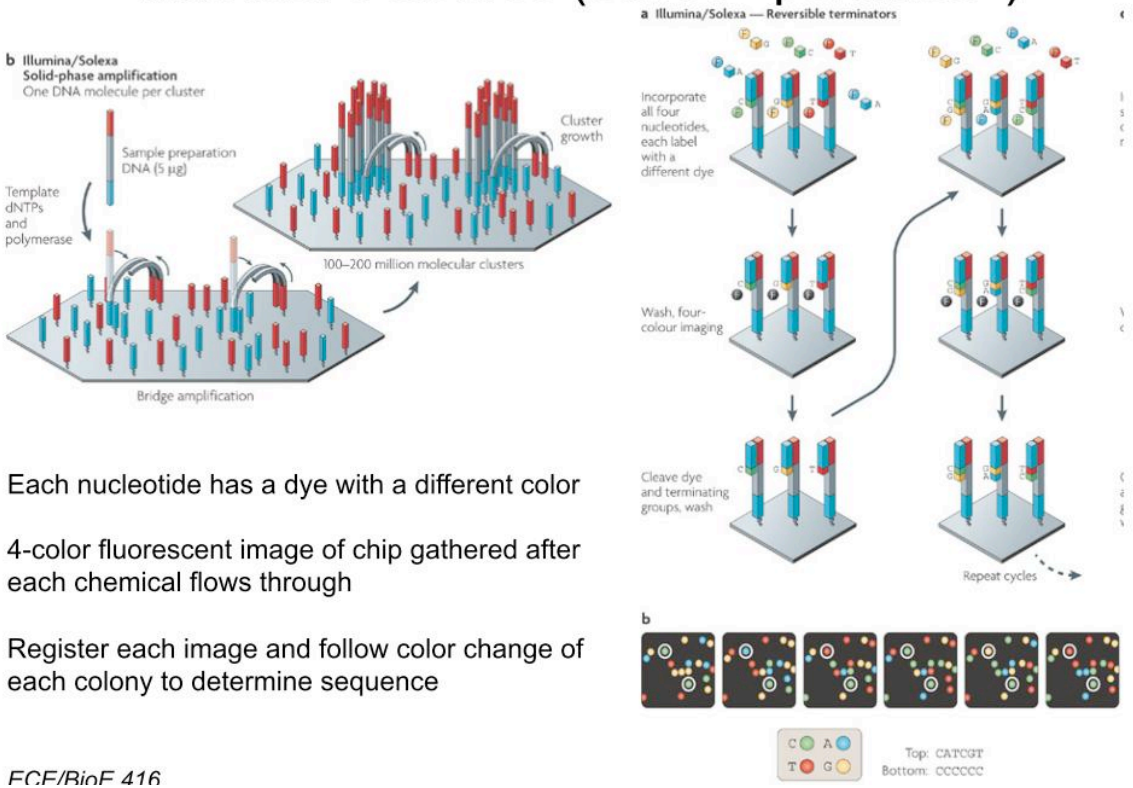
I. Genome sequencing

2nd generation sequencing : real-time sequencing

Sequencing **by reversible terminator**

Used in combination with bridge PCR in the illumina/Solexa genome Analyzer

Illumina Colonies (called “polonies”)



Each nucleotide has a dye with a different color

4-color fluorescent image of chip gathered after each chemical flows through

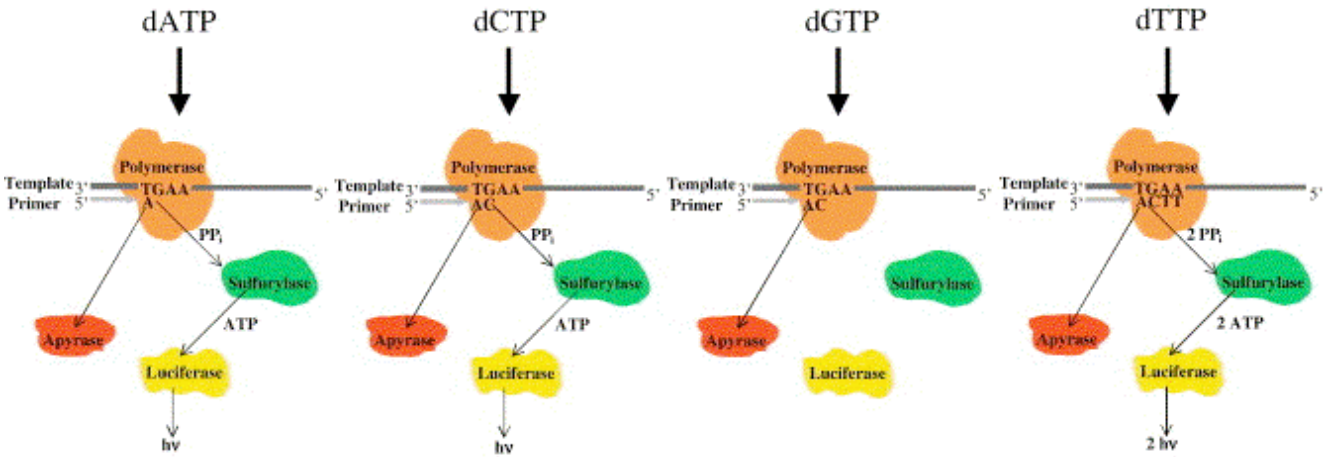
Register each image and follow color change of each colony to determine sequence

https://www.youtube.com/watch?v=fCd6B5HRaZ8

I. Genome sequencing

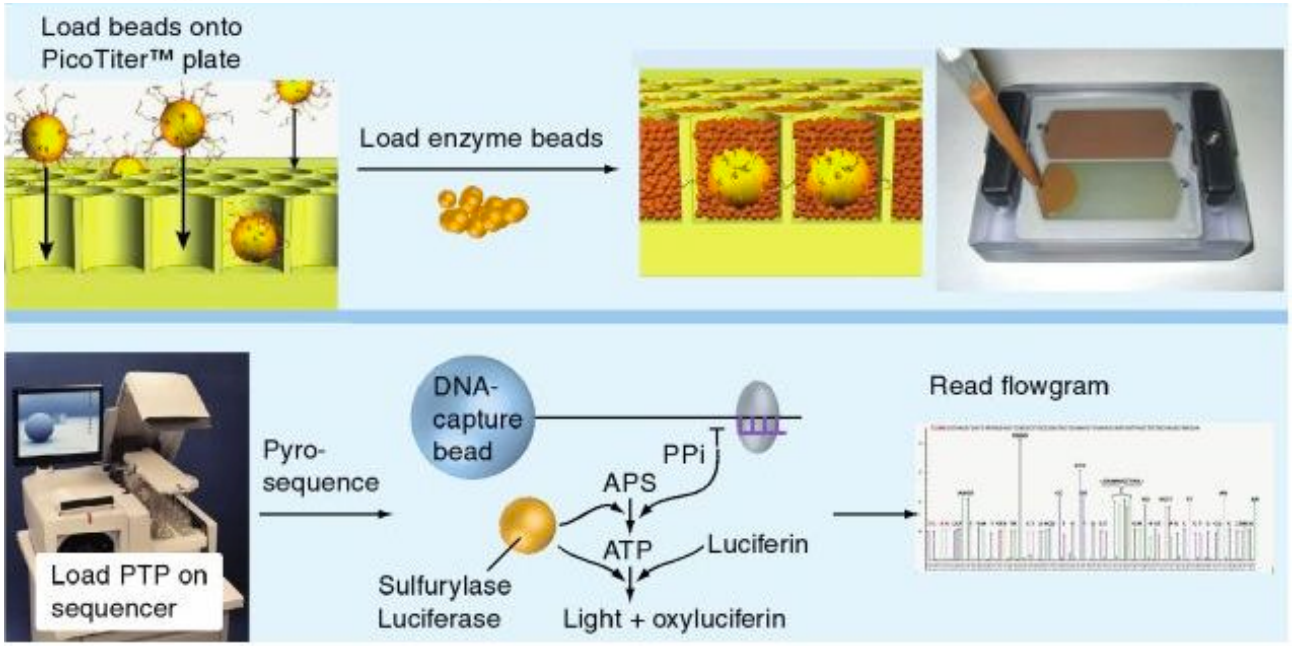
2nd generation sequencing : real-time sequencing

Pyrosequencing



Used in the 454 plateform (Life Sciences)

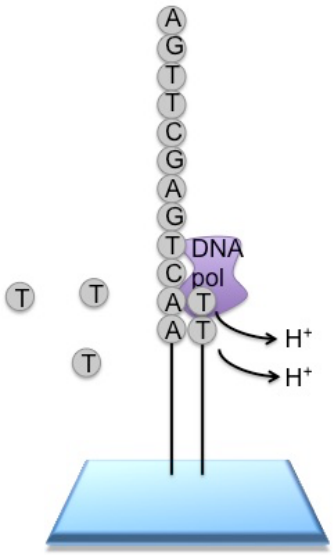
emPCR+Pyrosequencing

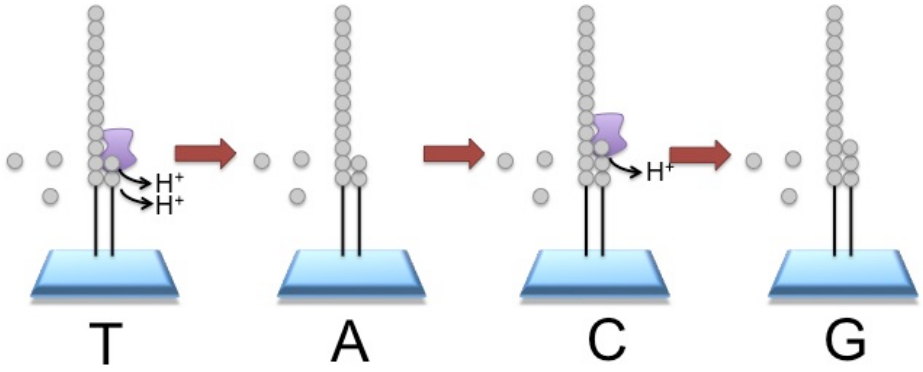


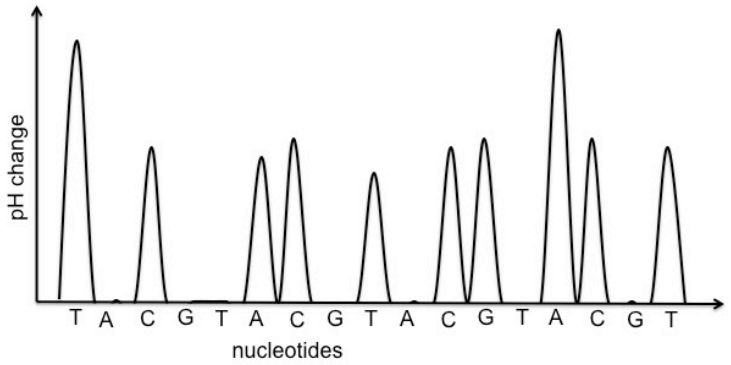
I. Genome sequencing

2nd generation sequencing : real-time sequencing

Ion proton sequencing (Ion torrent)





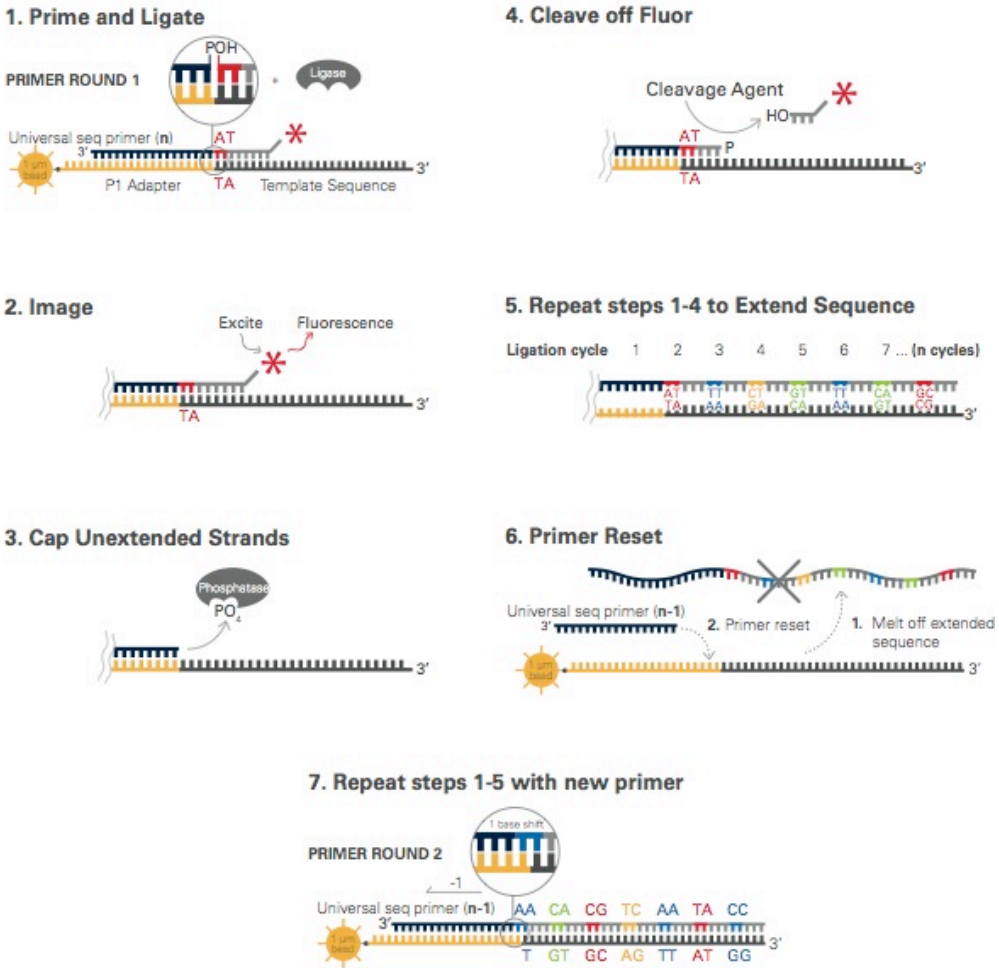


I. Genome sequencing

2nd generation sequencing : real-time sequencing

Ligase-based sequencing (ABI's Solid technology)

- Difference – DNA ligase
- Hybridization of a fluorescently labelled probe
- SOLiD cycle of 1,2-probe hybridization

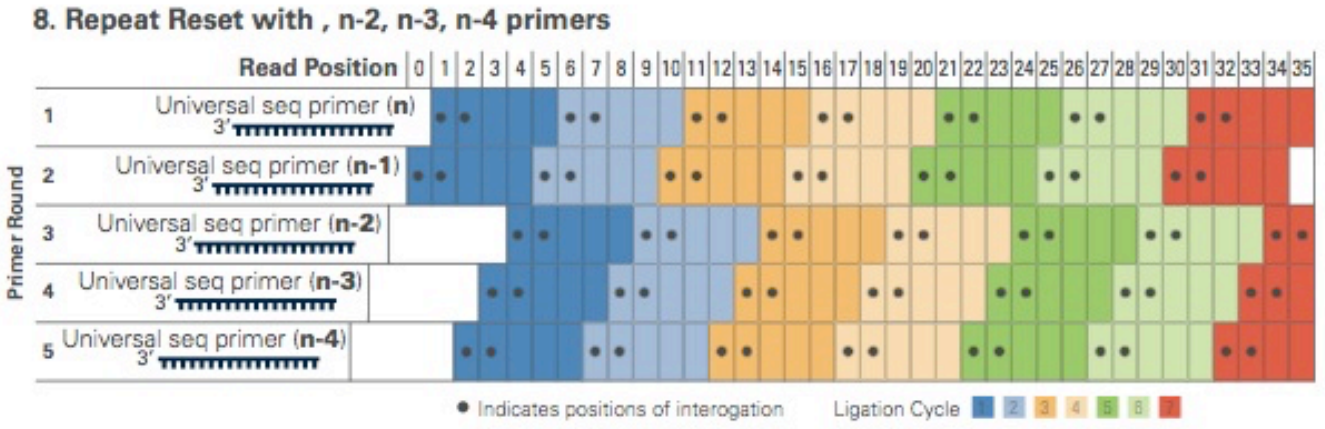


I. Genome sequencing

2nd generation sequencing : real-time sequencing

Ligase-based sequencing

SOLiD DNA Sequencing Technology (Applied Biosystems)

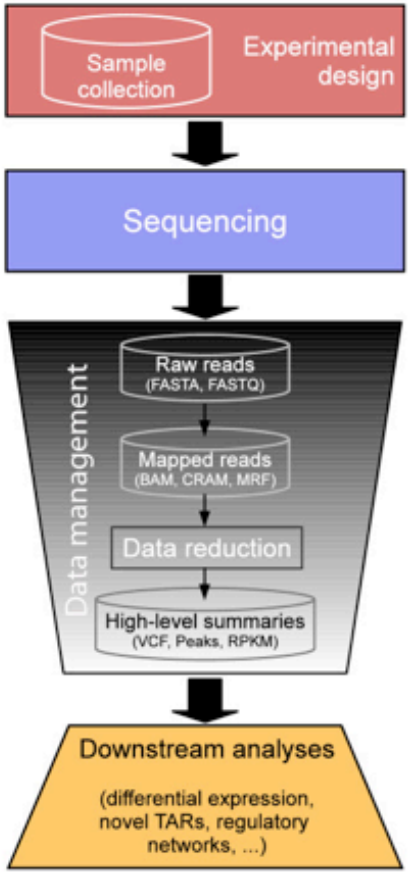


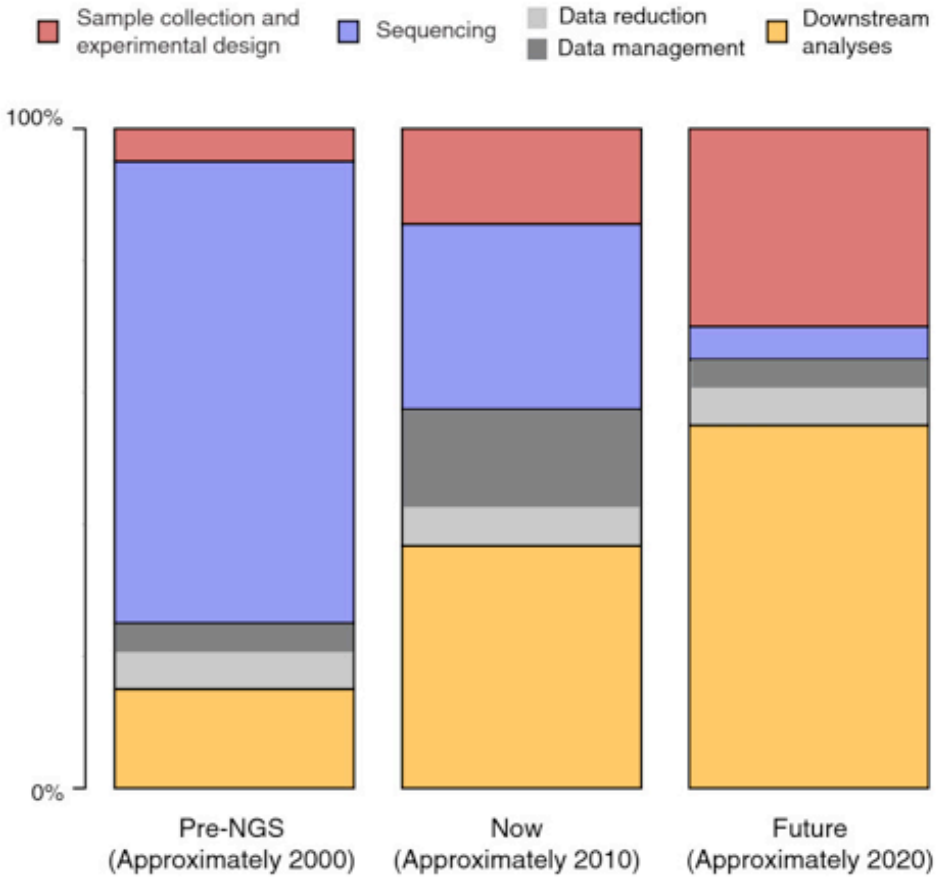
I. Genome sequencing

	Throughput	Length	Quality	Costs
Sanger	6 Mb/day	800nt	10^{-4} - 10^{-5}	500$/Mb
454	750Mb/day	400nt	10^{-3} - 10^{-4}	~20$/Mb
Ion Torrent	1600Mb/day	200nt	10^{-2} - 10^{-3}	~10$/Mb
Illumina	100000Mb/day	125nt	10^{-2} - 10^{-3}	~0.40$/Mb
SOLiD 4	100000Mb/day	125nt	10^{-2} - 10^{-3}	~0.40$/Mb
Helicos	5000Mb/day	32nt	10^{-2}	~0.40$/Mb

Today's technologies generate huge amount of short, high quality sequencing data

II. Data analysis



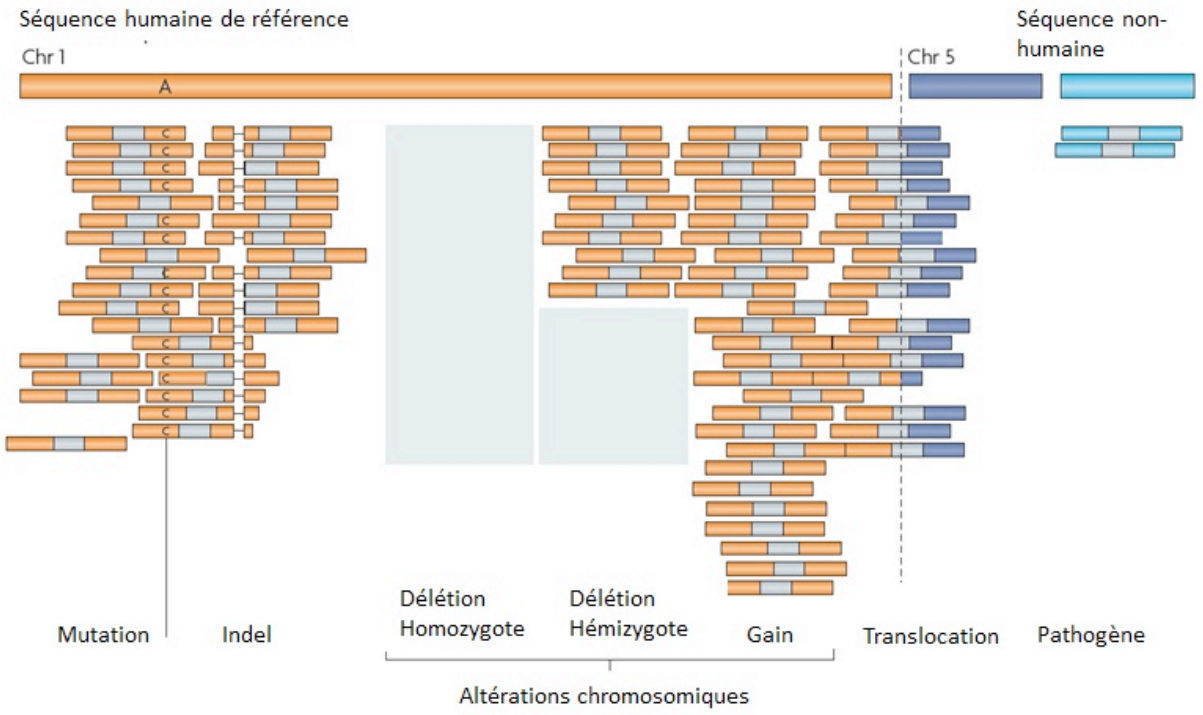


II. Data analysis

- Bioinformatics tools for:
  - Alignment
  - Base calling/polymorphism detection
  - *De novo* assembly
  - Genome browsing or annotation
- Challenging problems:
  - *De novo* assembly of short reads -> mate-paired libraries required
  - Reads in repetitive regions

III. NGS Applications

- Genome resequencing: polymorphism and mutation discovery in humans (1000 Genomes Project)



III. NGS Applications

- Transcriptome sequencing:

  Gene expression

  Alternative splicing

  Transcript annotation