

Next-Generation Sequencing

Thomas Dugé de Bernonville

Université de Tours, EA2106 Biomolécules et Biotechnologies Végétales

November 5, 2018


[HOME](#)
[ABOUT](#)
[SUBPROJECTS](#)
[PROTOCOL & STANDARDS](#)
[STATUS](#)

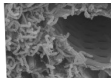
FISHT1K: Transcriptomes of 1,000 Fishes

Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000

Deep sequencing of 10,000 human genomes

Amalio Telenti^{a,b,1}, Levi C. T. Pierce^{a,c,1}, William H. Biggs^{a,1}, Julia di Iulio^{a,b}, Emily H. M. Wong^a, Martin M. Fabani^a, Ewen F. Kirkness^a, Ahmed Moustafa^a, Naisha Shah^a, Chao Xie^d, Suzanne C. Brewerton^d, Nadeem Bulsara^a, Chad Garner^a, Gary Metzker^a, Efren Sandoval^a, Brad A. Perkins^a, Franz J. Och^{a,c}, Yaron Turpaz^{a,d}, and J. Craig Venter^{a,b,2}

^aHuman Longevity Inc., San Diego, CA 92121; ^bJ. Craig Venter Institute, La Jolla, CA 92037; ^cHuman Longevity Inc., Mountain View, CA 94041; and ^dHuman Longevity Singapore Pte. Ltd., Singapore 138542



Hungate 1000

The Hungate 1000 project aims to produce a reference set of rumen microbial genome sequences by sequencing the genomes of available cultivated rumen bacteria and methanogenic archaea, together with representative cultures of bic fungi and ciliate protozoa.



10KP

Beta

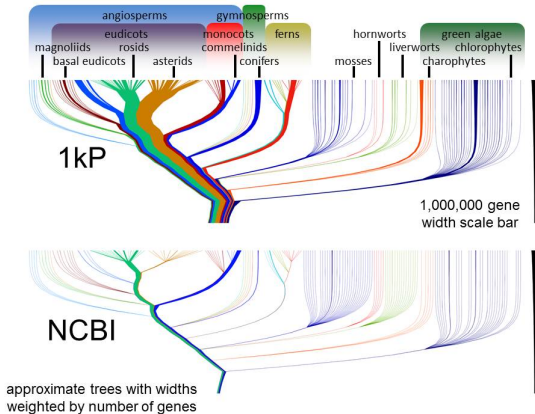
[HOME](#)
[STRATEGY](#)
[TIMELINE](#)
[SPECIES DATABASE](#)
[PRIORITY LIS](#)

 10KP: 10,000 Plant Genomes Project

Pourquoi séquencer des acides nucléiques?

- ▶ Comprendre le déterminisme génétique d'un phénotype
 - ▶ Outils de diagnostic/prédiction
 - ▶ Outils biotechnologiques
- ▶ Retracer des histoires évolutives
 - ▶ Génomique comparative
 - ▶ Phylogénie moléculaire (phylogénomique)

Pourquoi séquencer des acides nucléiques?



Méthodes de séquençage

- ▶ Sanger
- ▶ NGS (**N**ext **G**eneration **S**equencing) ou High Throughput Sequencing



Méthodes de séquençage

- ▶ **Sanger**
- ▶ NGS (**N**ext **G**eneration **S**equencing) ou High Throughput Sequencing

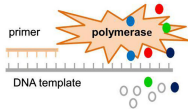


Séquençage première génération

Sanger - principe

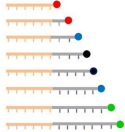
reaction mixture:

- primer
- DNA template
- DNA polymerase
- dNTPs
- fluorescently labeled ddNTPs

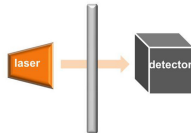


ddATP
ddTTP
ddCTP
ddGTP
dNTPs

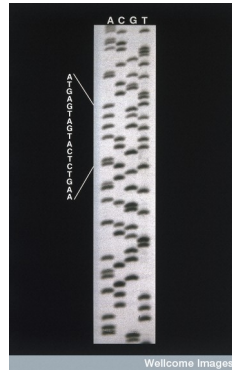
chain extension & chain termination



gel electrophoresis & fluorescence detection



computational sequence analysis

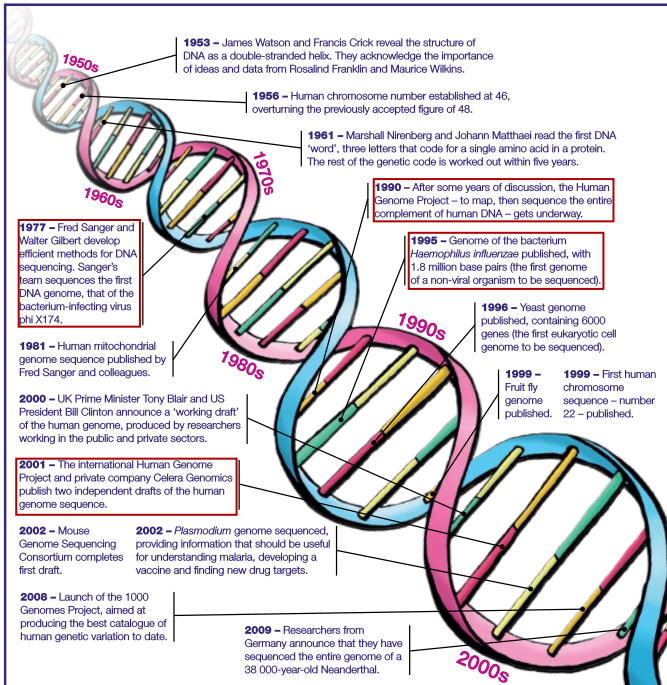


Séquençage première génération

Séquençage des premiers génomes

Sanger - limitations

- ▶ Faible débit
- ▶ Coût



Méthodes de séquençage

- ▶ Sanger
- ▶ **NGS (Next Generation Sequencing) ou High Throughput Sequencing**



(a)					
No.	Plant	Tissue	Roche GS-FLX Titanium		
			SRA accession	Number of raw reads	Number of cleaned reads
1	<i>Amsonia hubrichtii</i>	Leaf	SRS160815	715,432	627,544
2	<i>Arnicia montana</i>	Leaf	SRS212543	739,124	660,971
3	<i>Centella asiatica</i>	Roots	SRS212540	712,683	626,382
4	<i>Chelidonium majus</i>	Stem	SRS150402	504,849	408,742

(b)					
No.	Plant	Tissue	Illumina genome analyzer		
			SRA accession	Number of raw reads	
1	<i>Amsonia hubrichtii</i>	Leaf	SRS260004	75,544,560	
2	<i>Arnicia montana</i>	Leaf	SRS260095	67,378,080	
3	<i>Centella asiatica</i>	Roots	SRS260019	77,853,840	
4	<i>Chelidonium majus</i>	Stem	SRS259976	62,362,800	



RESULTS

Assembly of the *G. hirsutum* genome

The allotetraploid genome of Upland cotton *G. hirsutum* has been estimated, using various methods, as 2.25–2.43 Gb^{13,14}. We generated a total of 445.7 Gb, or 181-fold haploid genome coverage, of raw paired-end Illumina reads by sequencing whole genome shotgun (WGS) libraries of homozygous cv. 'TM-1' with fragment lengths ranging from 250 bp to 40 kb (Supplementary Table 1). Owing to the existence of abundant repetitive sequences and homeologous chromosomes (Supplementary Table 2), we were unable to assemble this allotetraploid genome satisfactorily using only the WGS data. Supplemental use of a bacterial artificial chromosome (BAC-to-BAC) sequencing strategy substantially improved the assembly. A total of 100,187 BACs, that corresponded to about fivefold genome coverage, were

NONHUMAN GENOMICS

Long-read sequence assembly of the gorilla genome

David Gordon,* John Huddleston,* Mark J. P. Chaisson,* Christopher M. Hill,* Zev N. Kronenberg,* Katherine M. Munson, Maika Malig, Archana Raja, Ian Fiddes, LaDeana W. Hillier, Christopher Dunn, Carl Baker, Joel Armstrong, Mark Diekhans, Benedict Paten, Jay Shendure, Richard K. Wilson, David Haussler, Chen-Shan Chin, Evan E. Eichler†

INTRODUCTION: The accurate sequence and assembly of genomes is critical to our understanding of evolution and genetic variation. Despite advances in short-read sequencing technology that have decreased cost and increased throughput, whole-genome assembly of mammalian genomes remains problematic because of the presence of repetitive DNA.

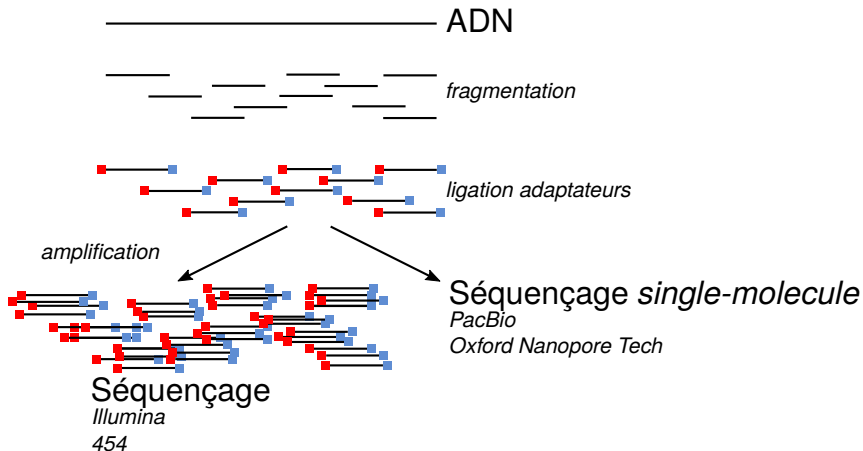
RATIONALE: The goal of this study was to sequence and assemble the genome of the western lowland gorilla by using primarily

single-molecule, real-time (SMRT) sequencing technology and a novel assembly algorithm that takes advantage of long (>10 kbp) sequence reads. We specifically compare the properties of this assembly to gorilla genome assemblies that were generated by using more routine short sequence read approaches in order to determine the value and biological impact of a long-read genome assembly.

RESULTS: We generated 74.8-fold SMRT whole-genome shotgun sequence from peripheral



Préparation des banques



Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

NGS 3e génération

Nanopores

SMRT

Comparaisons et applications

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

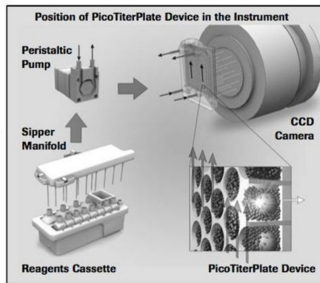
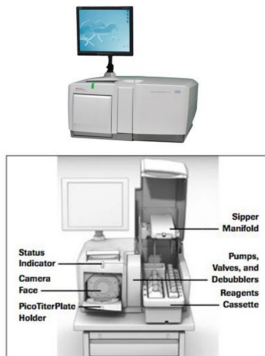
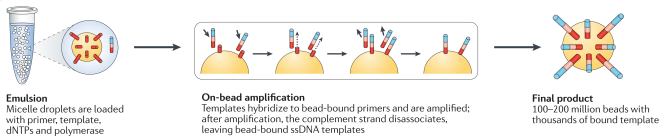
Séquençage par ligation

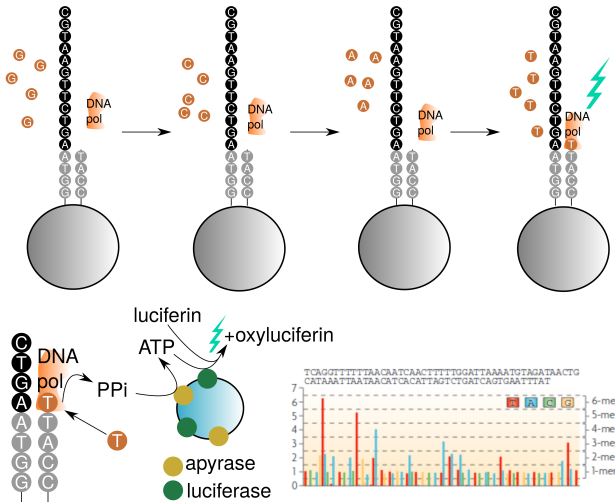
NGS 3e génération

Nanopores

SMRT

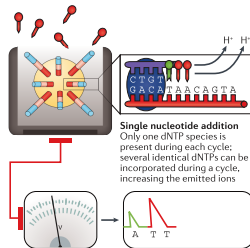
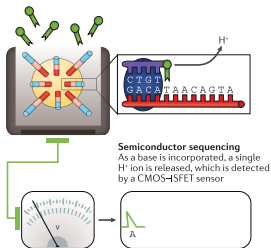
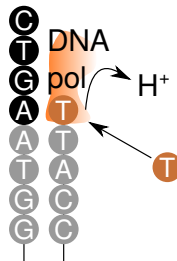
Comparaisons et applications





- ▶ jusqu'à 1000 pb
- ▶ jusqu'à 1 million de lectures
- ▶ moins de 1 jour
- ▶ 1Gb: \$9500

IonTorrent



IonTorrent

- ▶ 200 ou 400 pb SE
- ▶ <100 000 lectures
- ▶ quelques heures
- ▶ 1Gb: \$80

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

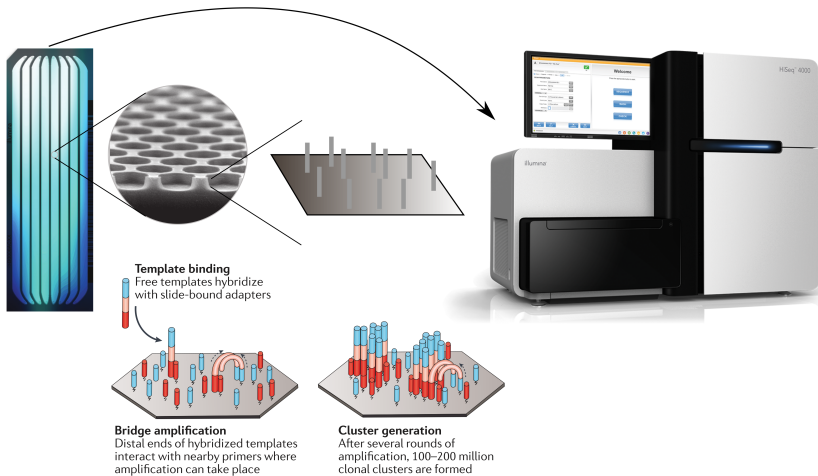
Séquençage par ligation

NGS 3e génération

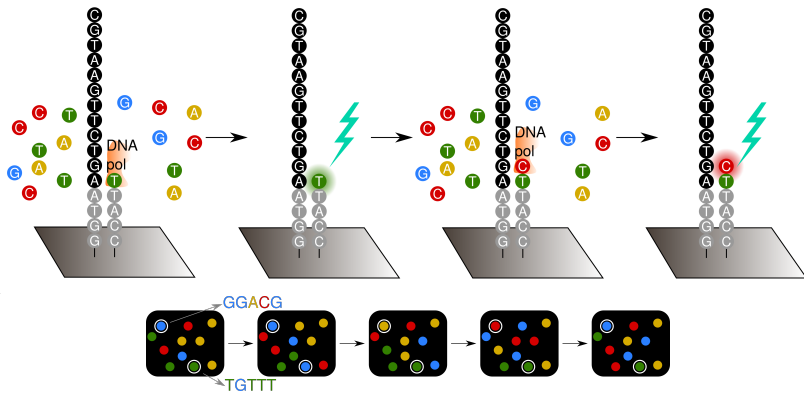
Nanopores

SMRT

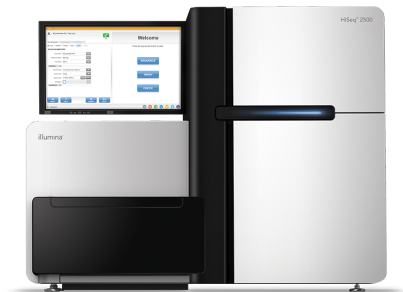
Comparaisons et applications



Illumina



HiSeq



- ▶ jusqu'à 150 pb
- ▶ jusqu'à 4 milliards de lectures
- ▶ de 1 à 6 jours
- ▶ 1Gb: \$30

HiSeq X



The HiSeq X Ten System is the world's first to break the \$1000 genome barrier for human whole-genome sequencing (WGS). The system consists of a set of 10 HiSeq X ultra-high-throughput instruments that deliver over 18,000 human genomes per year. Each HiSeq X System can sequence a human genome at 30x coverage or greater for significantly less than \$1000 in consumables.

MiSeq



- ▶ jusqu'à 300 pb
- ▶ jusqu'à 50 millions de lectures
- ▶ de 1 à 4 jours
- ▶ 1Gb: \$110

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

NGS 3e génération

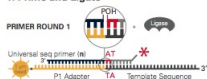
Nanopores

SMRT

Comparaisons et applications

Applied Biosystems - SOLiD

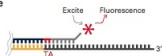
1. Prime and Ligate



4. Cleave off Fluor



2. Image

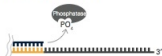


5. Repeat steps 1-4 to Extend Sequence

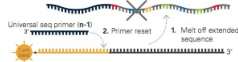
Ligation cycle 1 2 3 4 5 6 7... (n cycles)



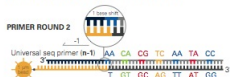
3. Cap Unextended Strands



6. Primer Reset



7. Repeat steps 1-5 with new primer



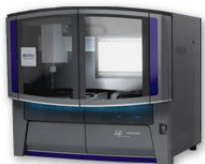
8. Repeat Reset with , n-2, n-3, n-4 primers

		Read Position																																			
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
Primer Round	1	Universal seq primer (n)																																			
	2	Universal seq primer (n-1)																																			
	3	Universal seq primer (n-2)																																			
	4	Universal seq primer (n-3)																																			
	5	Universal seq primer (n-4)																																			

• Indicates positions of interrogation

Ligation Cycle

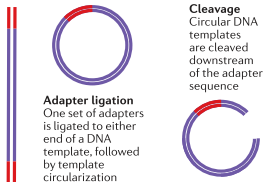
Applied Biosystems - SOLiD



- ▶ 50 ou 75 pb
- ▶ jusqu'à 700 millions de lectures
- ▶ 6 jours
- ▶ 1Gb: \$70
- ▶ **en cours d'abandon**

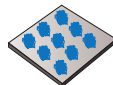
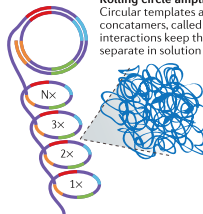
Beijing Sequencing

In-solution DNA nanoball generation (Complete Genomics (BGI))



Rolling circle amplification

Circular templates are amplified to generate long concatamers, called DNA nanoballs; intermolecular interactions keep the nanoballs cohesive and separate in solution



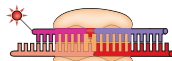
Hybridization
DNA nanoballs are immobilized on a patterned flow cell

2. Loading DNBS

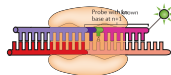
3. Sequencing



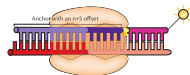
Beijing Sequencing



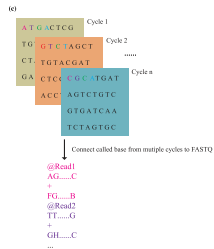
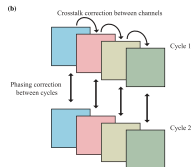
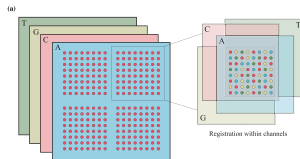
Single-base-encoded probes
A probe with a single known base and degenerate bases hybridizes to a template and is imaged



Paired-end sequencing
Sequencing is performed for both the left and right sides of the adapter



Offset anchors
Subsequent rounds of hybridization and ligation use offset anchors to sequence more-distant bases



Beijing Sequencing

- ▶ jusqu'à 100 pb
- ▶ nombre de lectures?
- ▶ 24h
- ▶ Coût non défini

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

NGS 3e génération

Nanopores

SMRT

Comparaisons et applications

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

NGS 3e génération

Nanopores

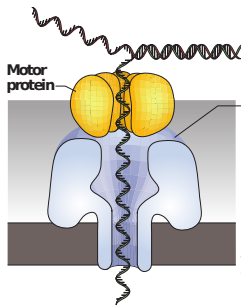
SMRT

Comparaisons et applications

ONT

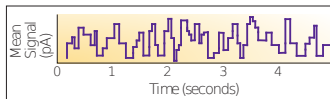


Leader-Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing



Alpha-hemolysin
A large biological pore capable of sensing DNA

Current
Passes through the pore and is modulated as DNA passes through



ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

ONT

- ▶ jusqu'à 200 Kpb
- ▶ jusqu'à 100 000 lectures
- ▶ moins de 2 jours
- ▶ 1Gb: \$750

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

NGS 3e génération

Nanopores

SMRT

Comparaisons et applications



- ▶ Lectures longues: jusqu'à 20 Kbp

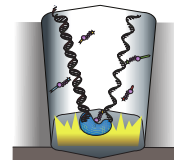
SMRTbell template
Two hairpin adapters allow continuous circular sequencing



ZMW wells
Sites where sequencing takes place



Labelled nucleotides
All four dNTPs are labelled and available for incorporation



Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light

PacBio output
A camera records the changing colours from all ZMWs; each colour change corresponds to one base



- ▶ jusqu'à 20 Kpb
- ▶ jusqu'à 55 000 lectures
- ▶ 4h
- ▶ 1Gb: \$1000

Plan

NGS 2e génération

Sequencing-by-synthesis: Single Nucleotide Addition

Sequencing-by-synthesis: Cyclic Reversible Termination

Séquençage par ligation

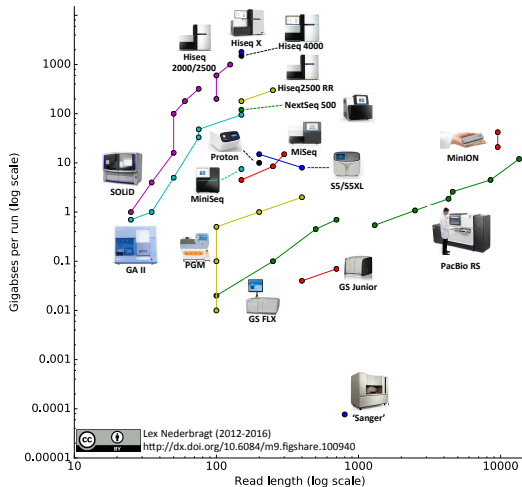
NGS 3e génération

Nanopores

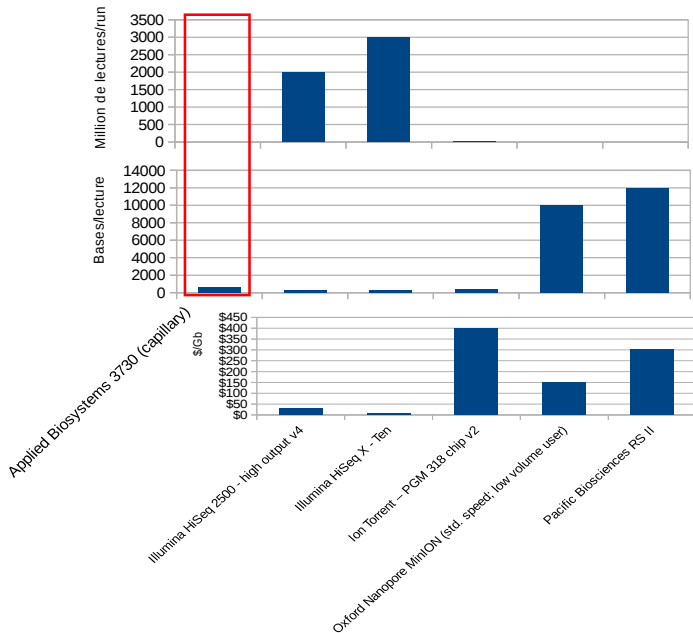
SMRT

Comparaisons et applications

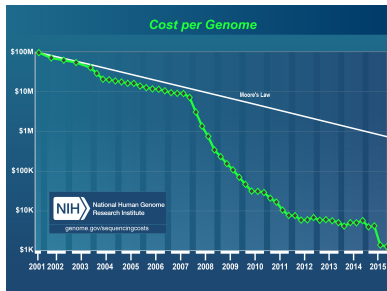
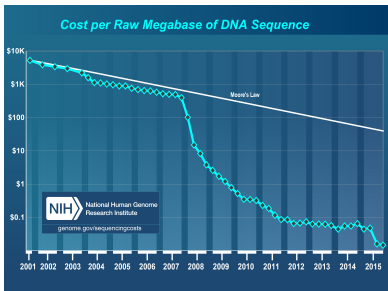
Débit et longueur des lectures



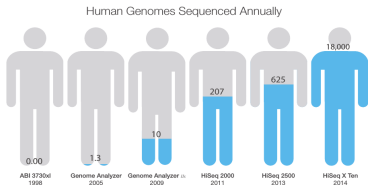
Débit et longueur des lectures



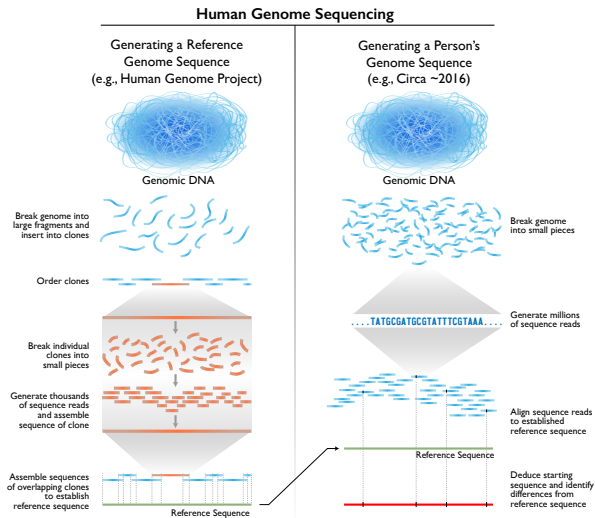
Les coûts de séquençage ont très largement chuté



- ▶ Annonce d'un génome humain pour 1000\$
- ▶ Augmentation du nombre de projets de séquençage de manière générale



Stratégies pour le séquençage des génomes



Applications

