

## Secuenciación de Genomas Bacterianos: Herramientas y Aplicaciones

**BU-ISCIII**

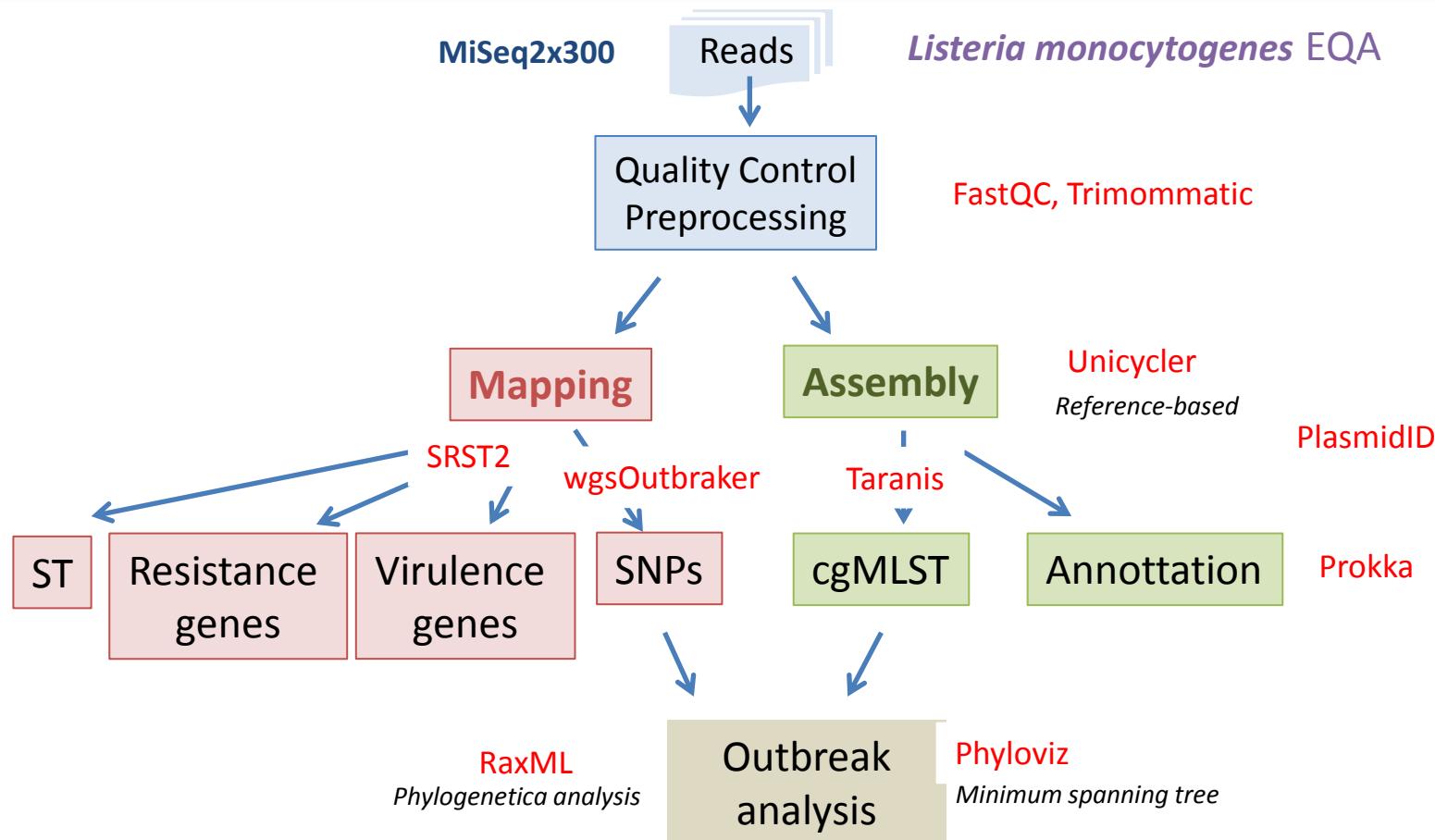
**Unidades Comunes Científico Técnicas - SGSAFI-ISCIII**

05-09 Noviembre 2018, 1<sup>a</sup> Edición  
Programa Formación Continua, ISCIII

# Learning aims and outcomes

- Understand some principles behind NGS and its applications to whole genome sequencing.
- Know the format files generated in NGS data analysis and the workflow analysis.
- Understand the uses of WGS in: specie, antimicrobial resistance genes and virulence factor genes identification, and for typing.
- Outbreak characterization based on SNPs or gene by gene approaches.

# Training workflow



# Teachers

- Sara Monzón Fernández, Biotecnóloga y Bioinformática (Analista de datos). Titulado Superior Especialista OPIS (nivel 24).
- Pedro J. Sola Campoy, Biólogo y Bioinformático (Analista de Datos). Contrato Titulado Superior Servicio Antibioticos (2017-2018)
- Miguel Juliá Molina, Matemático y Bioinformático (Analista de Datos). Contrato Titulado Superior, PTA Mineco (2018-2020)
- Isabel Cuesta, Dra Biología, Bioinformática (Científico de Datos). Científico Titular de OPIS (nivel 27). Coordinador BU-ISCIII

## **Session 1.1 - Secuenciación masiva de genomas bacterianos: situación actual**

**Isabel Cuesta**

**BU-ISCIII**

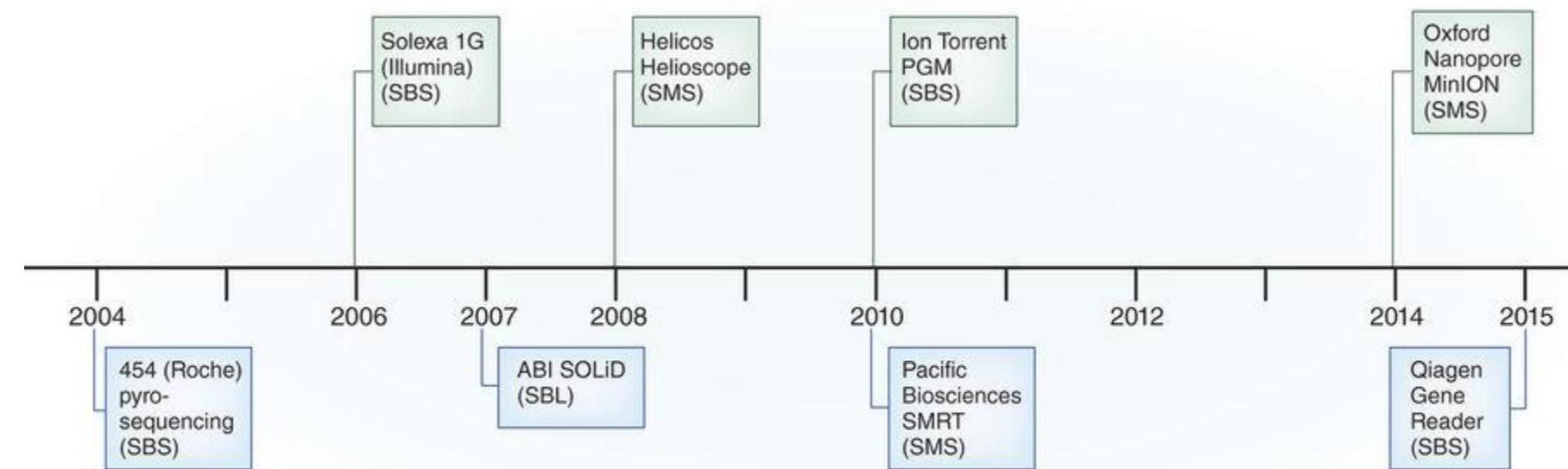
**Unidades Comunes Científico Técnicas - SGSAFI-ISCIII**

05-09 Noviembre 2018, 1<sup>a</sup> Edición  
Programa Formación Continua, ISCIII

# Index

- High throughput sequencing platforms update
- Bacterial genome sequencing, brief history
- Advantages of WGS
- Use of WGS in Europe

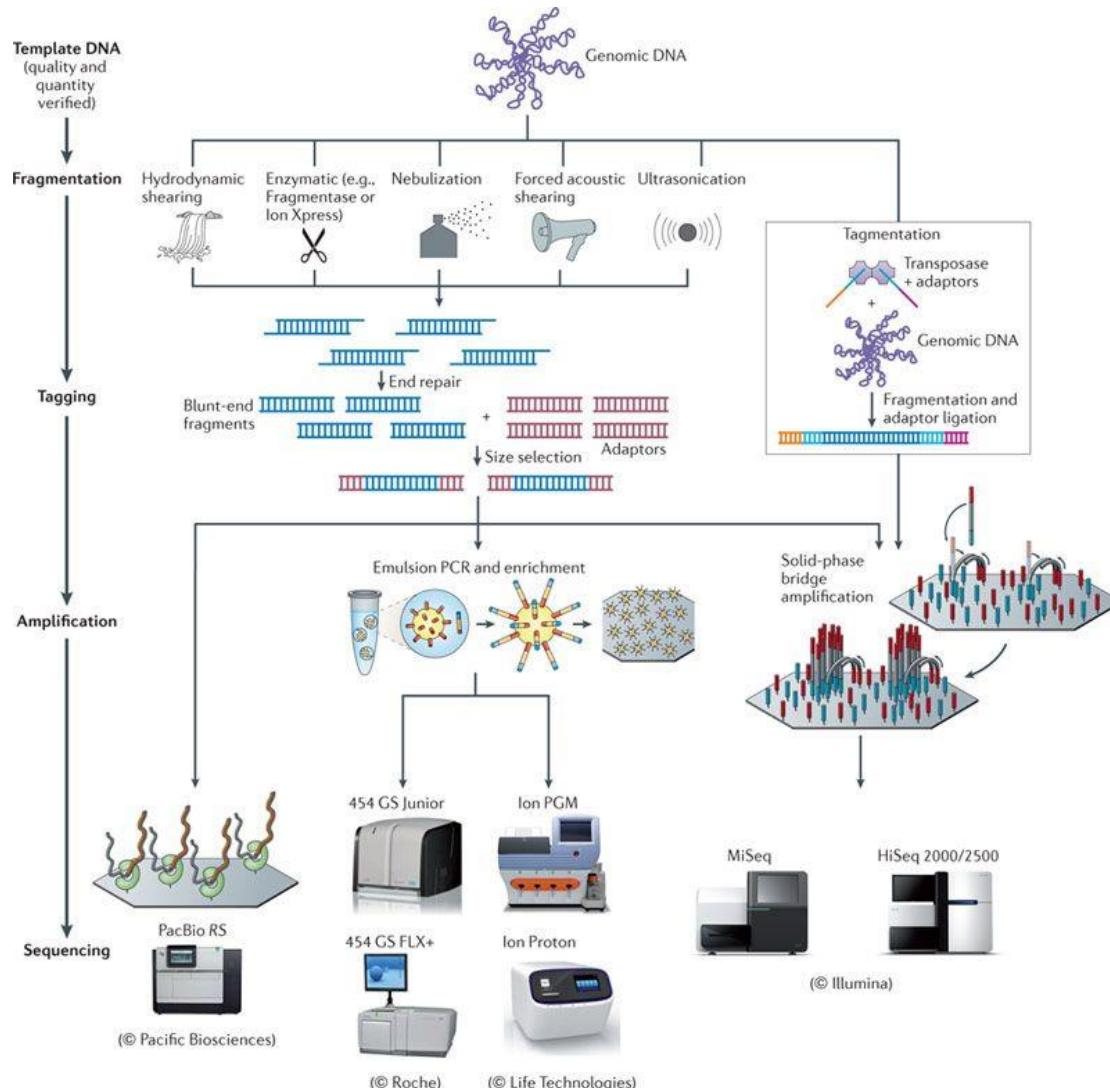
# DNA sequencing technologies 2006-2016



Mardis, Nature Protocols 2017

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

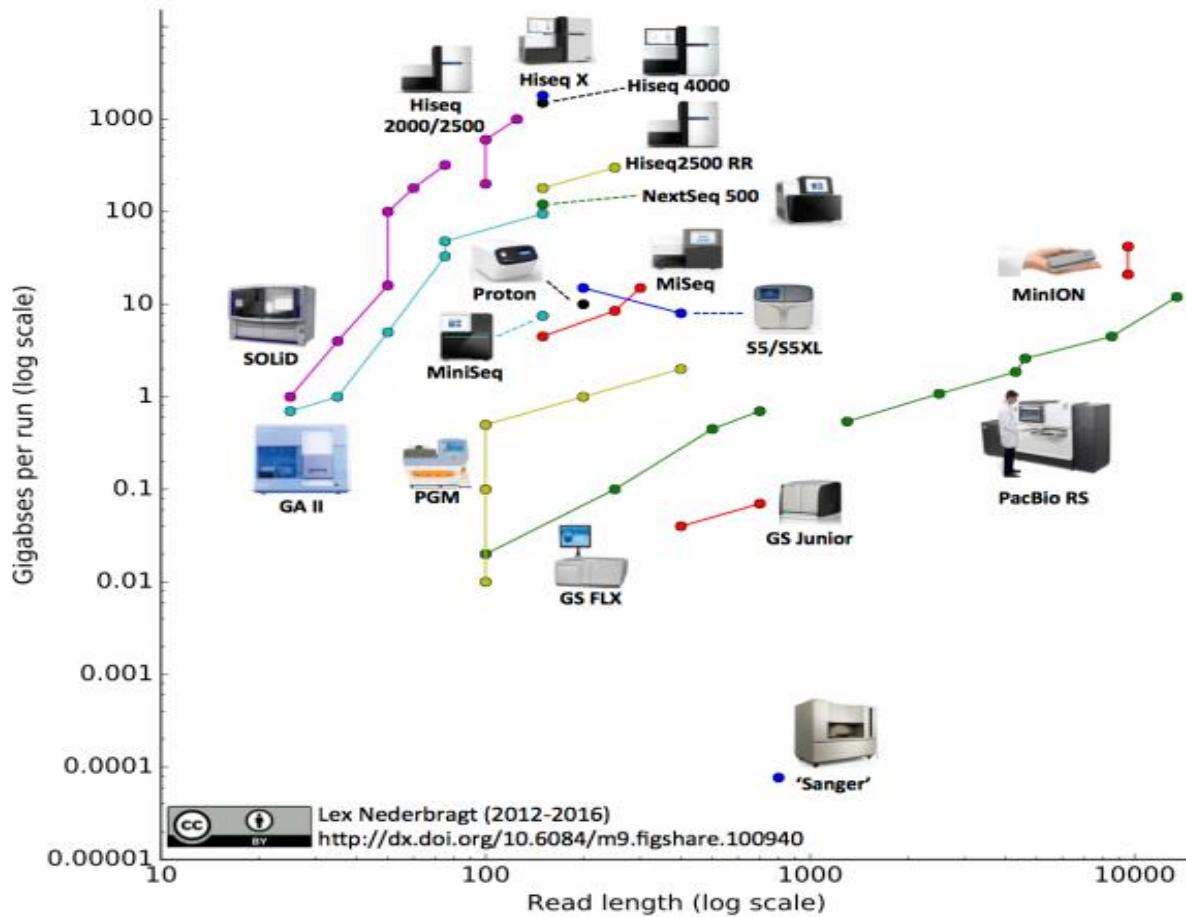
# High-throughput sequencing platforms



Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

Nature Reviews | Microbiology Loman et al, 2012

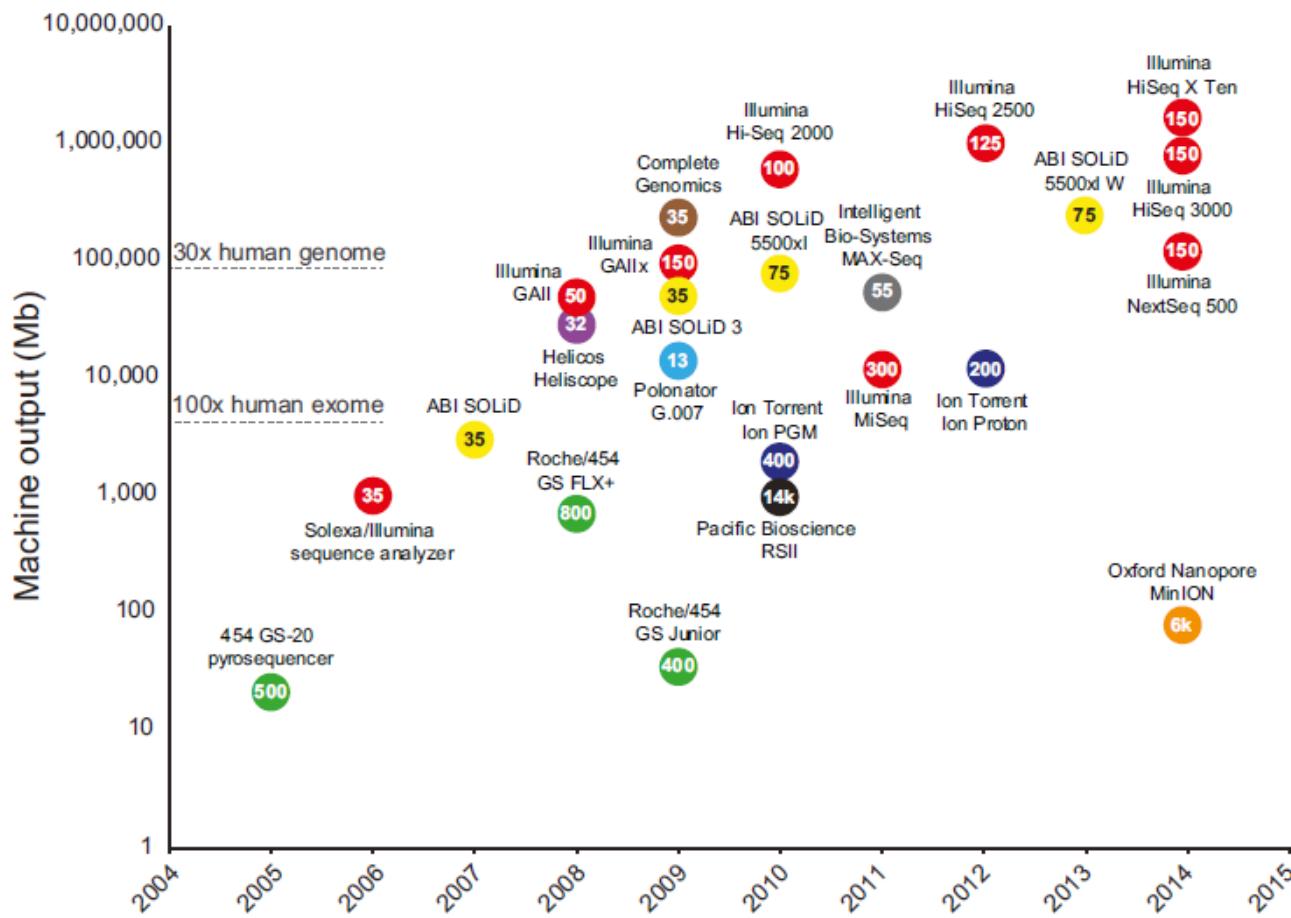
# High-Throughput Sequencing Technologies



<https://flxlexblog.wordpress.com/>

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

# High-Throughput Sequencing Technologies

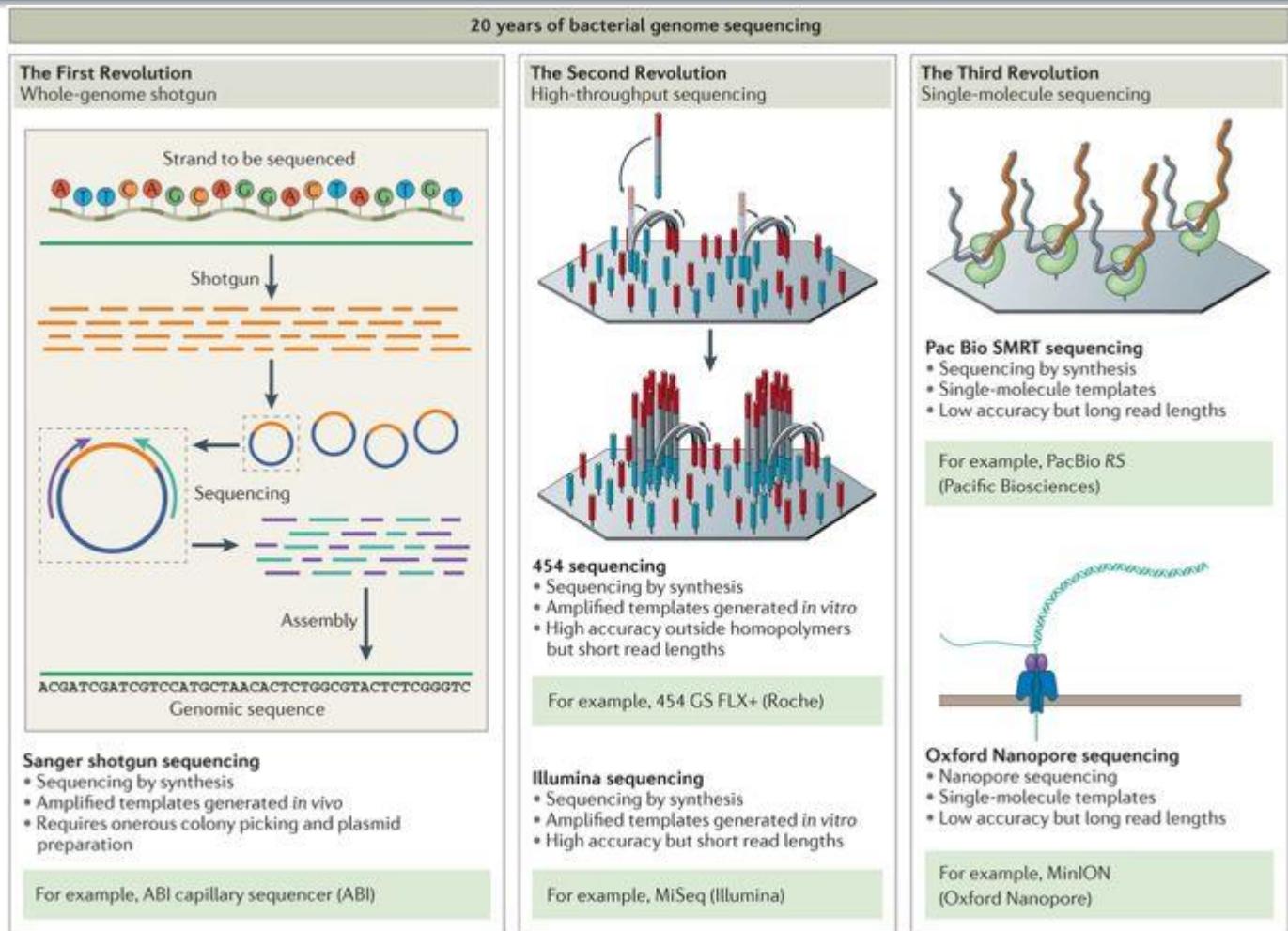


Numbers inside data points denote current read lengths.  
Sequencing platforms are color coded.

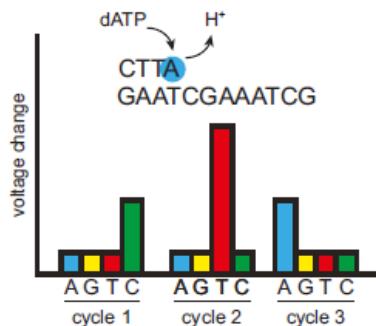
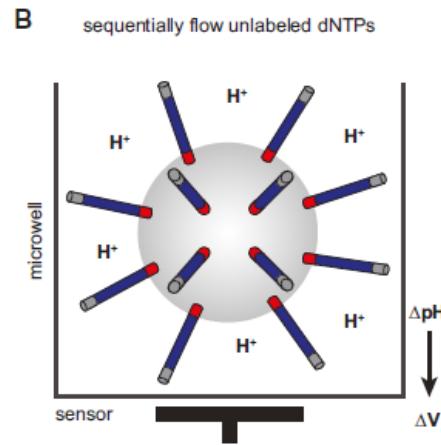
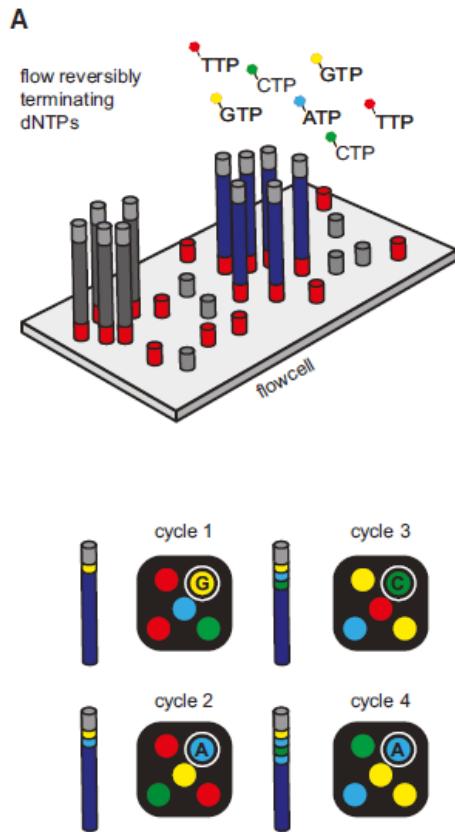
Reuter et al., Mol Cell 2015

# High-Throughput Sequencing Technologies

The three revolutions in sequencing technology that have transformed the landscape of bacterial genome sequencing



# The Second-generation Sequencing Technologies



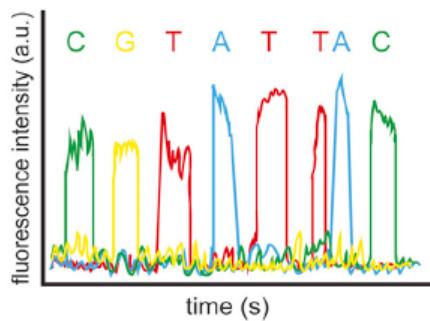
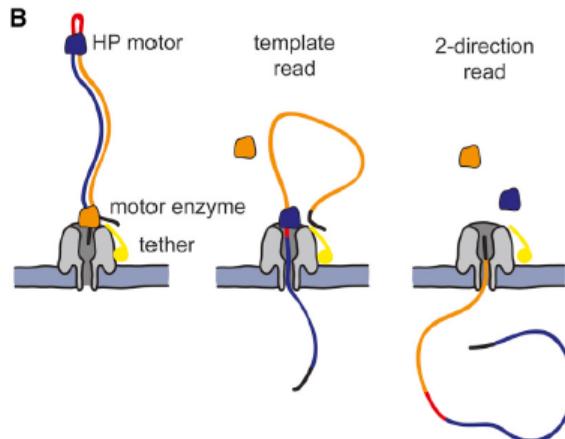
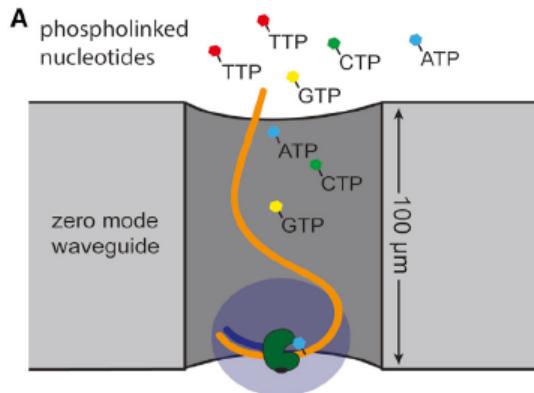
## Clonal Amplification-Based Sequencing Platforms

(A) Illumina's four-color reversible termination sequencing method.

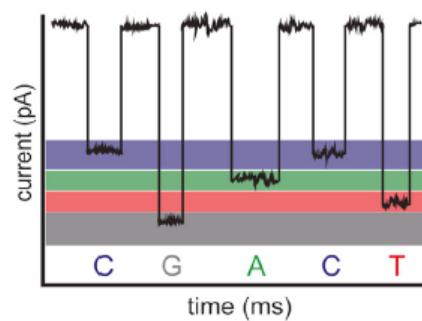
(B) Ion Torrent's semiconductor sequencing method.

# The Third-generation Sequencing Technologies

## Single Molecule Sequencing Platforms



Pacific Bioscience's SMRT sequencing



Oxford Nanopore's sequencing strategy

Reuter et al., Mol Cell 2015

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

# PacBio sequencing and its applications

Rhoads & Au, Gen Prot Bioinf 2015



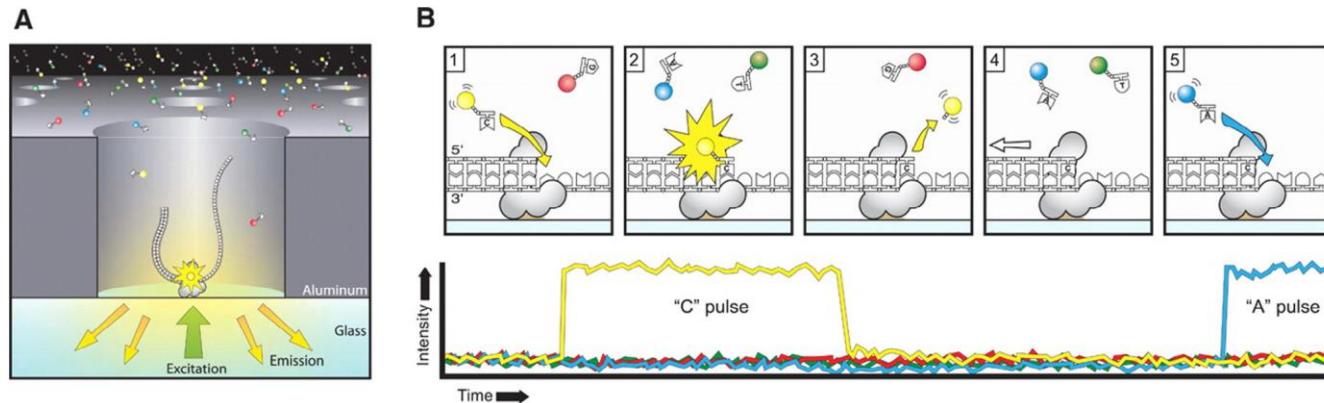
**SMRTbell template:** is a closed, single-stranded circular DNA that is created by ligating hairpin adaptors to both ends of a target dsDNA

**Sequencing by light pulses:** The replication processes in all ZMWs of a SMRTcell are recorder by a movie of light pulses, and the pulses corresponding to each ZMW can be interpreted to be a sequence of bases (**continuous long read, CLR**).

Both strands can be sequenced multiple times (passes) in a single CLR. CLR can be split to multiple reads (subreads) and CCS is the consensus sequence of multiple subreads



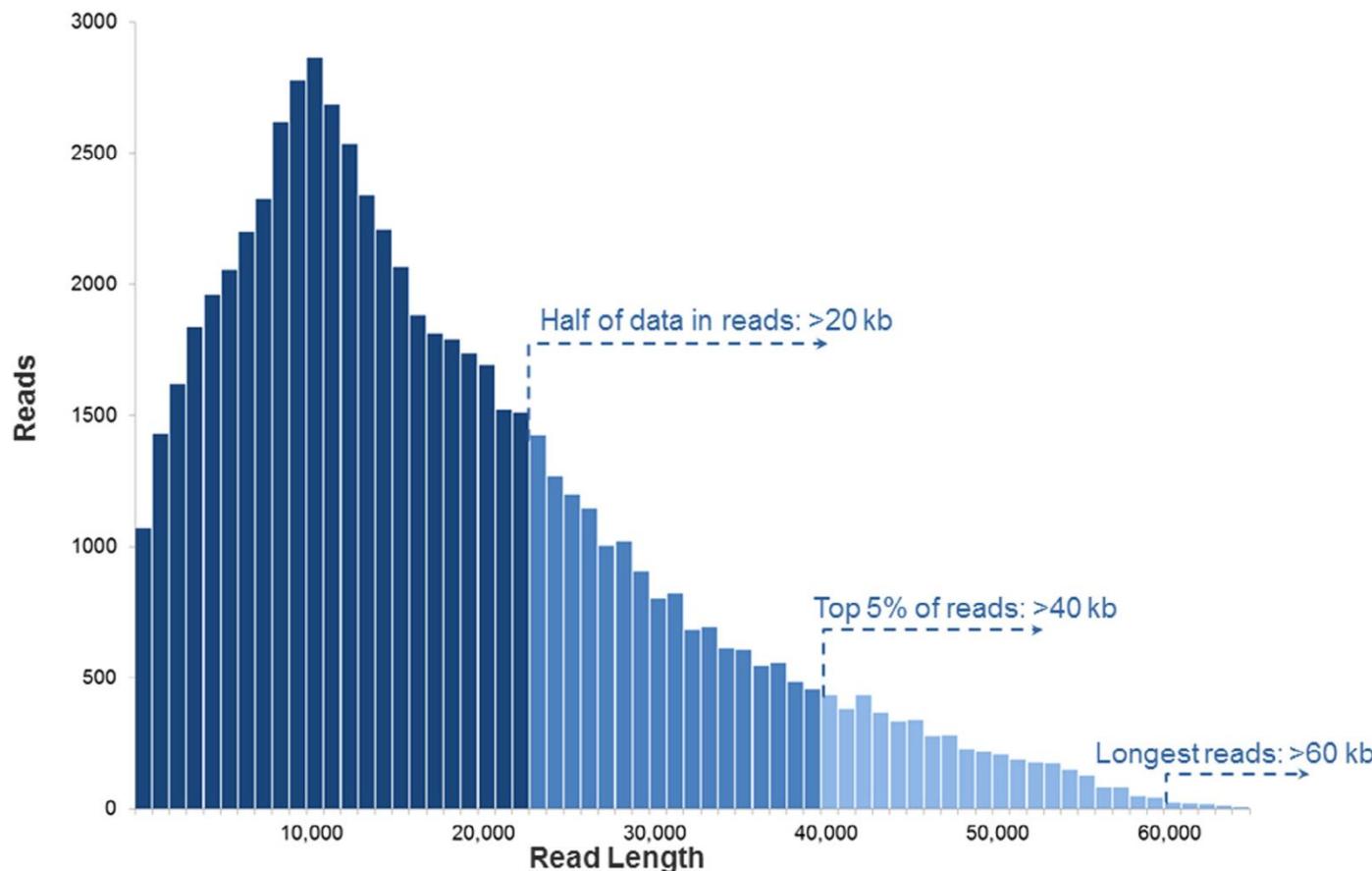
**A single SMRT cell:** this contains 150000 ZMWs (zero-mode waveguide). A SMRTbell diffuses into a ZMW.  
Approx 35000 -75000 ZMWs produce a read in a run lasting 0,5-4h resulting in 0,5-1Gb.



# PacBio sequencing and its applications

Rhoads & Au, Gen Prot Bioinf 2015

**PacBio RS II read length distribution** using P6-C4 chemistry. Data are based on a 20kb size-selected E. coli library using a 4-h movie. A SMRTcell produces 0,5-1 billion bases.



Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

# PacBio sequencing and its applications

Rhoads & Au, Gen Prot Bioinf 2015

Table 2 *De novo* genome assemblies using hybrid sequencing or PacBio sequencing alone

Species	Method	Tools	SMRT cells	Coverage	Contigs	Achievements	Ref.
<i>Clostridium autoethanogenum</i>	PacBio	HGAP	2	179×	1	21 fewer contigs than using SGS; no collapsed repeat regions ( $\geq 4$ using SGS)	[7]
<i>Potentilla micrantha</i> (chloroplast)	PacBio	HGAP, Celera, minimus2, SeqMan	26	320×	1	6 fewer contigs than with Illumina; 100% coverage (Illumina: 90.59%); resolved 187 ambiguous nucleotides in Illumina assembly; unambiguously assigned small differences in two $> 25$ kb inverted repeats	[33]
<i>Escherichia coli</i>	PacBio	PBcR, MHAP, Celera, Quiver	1	85×	1	4.6 CPU hours for genome assembly (10× improvement over BLASR)	[31]
<i>Saccharomyces cerevisiae</i>	PacBio	PBcR, MHAP, Celera	12	117×	21	27 CPU hours for genome assembly (8× improvement over BLASR); improved current reference of telomeres	[31]
<i>Arabidopsis thaliana</i>	PacBio	PBcR, MHAP, Celera	46	144×	38	1896 CPU hours for genome assembly	[31]
<i>Drosophila melanogaster</i>	PacBio	PBcR, MHAP, Celera, Quiver	42	121×	132	1060 CPU hours for genome assembly (593× improvement over BLASR); improved current reference of telomeres	[31]
<i>Homo sapiens</i> (CHM1hert)	PacBio	PBcR, MHAP, Celera	275	54×	3434	262,240 CPU hours for genome assembly; potentially closed 51 gaps in GRCh38; assembled MHC in 2 contigs (60 contigs with Illumina); reconstructed repetitive heterochromatic sequences in telomeres	[31]
<i>Homo sapiens</i> (CHM1hert)	PacBio	BLASR, Celera, Quiver	243	41×	N/A (local assembly)	Closed 50 gaps and extended into 40 additional gaps in GRCh37; added over 1 Mb of novel sequence to the genome; identified 26,079 indels at least 50 bp in length; cataloged 47,238 SV breakpoints	[32]
<i>Melopsittacus undulatus</i>	Hybrid	PBcR, Celera	3	$5.5 \times$ PacBio + $15.4 \times$ 454 = $3.83 \times$ corrected	15,328	1st assembly of $> 1$ Gb parrot genome; N50 = 93,069	[34]
<i>Vibrio cholerae</i>	Hybrid	BLASR, Bambus, AHA	195	$200 \times$ PacBio + $28 \times$ Illumina + $22 \times$ 454	2	No N's in contigs; 99.99% consensus accuracy; N50 = 3.01 Mb	[30]
<i>Helicobacter pylori</i>	PacBio	HGAP, Quiver, PGAP	8 per strain	446.5× average among strains	1 per strain	1 complete contig for each of 8 strains; methylation analysis associated motifs with genotypes of virulence factors	[35]

Note: N50, the contig length for which half of all bases are in contigs of this length or greater; MHC, major histocompatibility complex; SV, structural variation.

# PacBio sequencing and its applications

Rhoads & Au, Gen Prot Bioinf 2015

## Advantage

Closes gaps and completes genomes due to longer reads

Identifies non-SNP SVs

## Achievements

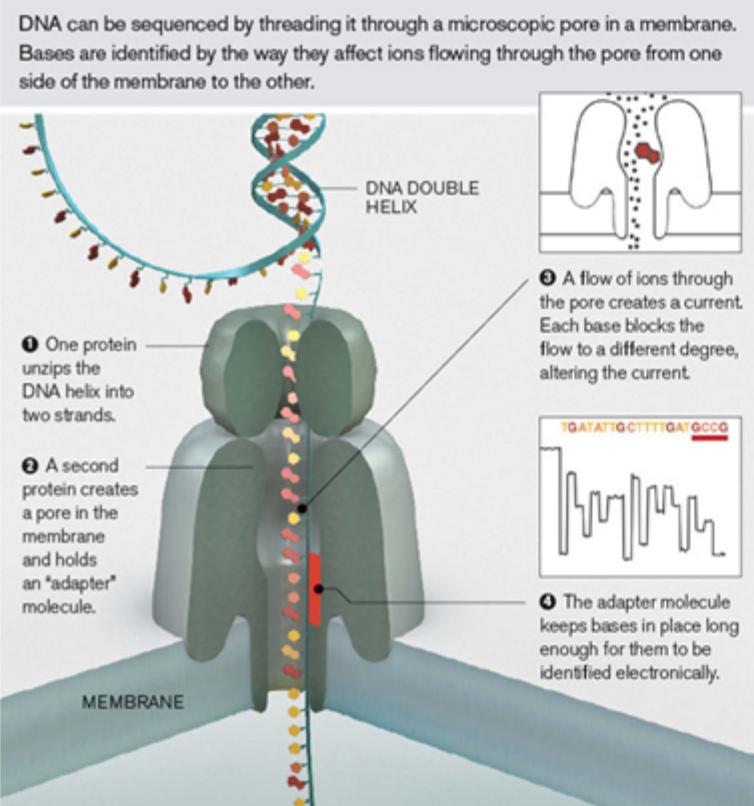
Produced highly-contiguous assemblies of bacterial and eukaryotic genomes

Discovered STRs (short tandem repeats)

## Limitations

Both strands can be sequenced several times if the lifetime of the polymerase is long enough.

# Nanopore-based fourth-generation DNA sequencing technology. ONT, Oxford Nanopore Technologies



'Strand sequencing' is a technique that passes intact DNA polymers through a protein nanopore, sequencing in real time as the DNA translocates the pore.

Nanopore sequencing also offers, for the first time, direct RNA sequencing, as well as PCR or PCR-free cDNA sequencing.

<https://nanoporetech.com/applications/dna-nanopore-sequencing>

Feng et al , Gen Prot Bioinf 2015

# Oxford Nanopore Technologies, MinION



The MinION is a portable sequencer; flow cells contain up to 512 nanopore sensors.

The Oxford Nanopore system processes the reads that are presented to it rather than generating read lengths. Sample-prep dependent, the longest read reported by a MinION user to date is >1 Mb.

Long reads confer many advantages, including simpler assembly and in the analysis of repetitive regions, phasing or CNVs.

# Oxford Nanopore Technologies



**Flongle**

**MinION**

**GridION**

**PromethION**

Long read, direct DNA/RNA/epigenetic sequencing, scalable, real time/rapid, on-demand sequencing that is easy to use and install.

<ul style="list-style-type: none"><li>✓ Your portable device for smaller, individual, rapid tests.</li><li>✓ When you don't want to multiplex samples or start a larger run.</li><li>✓ Amplicons, panels/targeted sequencing, quality testing and more.</li><li>✓ For use with MinIT or a laptop.</li></ul>	<ul style="list-style-type: none"><li>✓ Your personal sequencer, putting you in control.</li><li>✓ Whether in your lab or out in the field.</li><li>✓ Whole genomes/exomes, metagenomics, targeted sequencing, whole transcriptome (cDNA), smaller transcriptomes (direct RNA), multiplexing for smaller samples and more.</li><li>✓ For use with MinIT or a laptop.</li></ul>	<ul style="list-style-type: none"><li>✓ High throughput sequencing, in modular form (up to 5 flow cells) to be on-demand.</li><li>✓ For your lab or to offer as a service.</li><li>✓ Larger genomes or projects, whole transcriptomes (direct RNA or cDNA) or where you have larger numbers of samples and more.</li><li>✓ Compute included for real time data analysis and easy installation.</li></ul>	<ul style="list-style-type: none"><li>✓ Very high throughput sequencing, in modular form (up to 48 flow cells) to be on-demand.</li><li>✓ For your lab or as a service.</li><li>✓ Larger genomes or projects, whole transcriptomes (direct RNA or cDNA), very large numbers of samples and more.</li><li>✓ Compute included for real time data analysis and easy installation.</li></ul>
---	--	--	--

# Nanopore sequencing applications

Nanopore sequencing offers advantages in all areas of research...



Microbiology



Environmental research



Microbiome



Basic genome research



Human genetics



Cancer research



Clinical research



Plant research



Transcriptome analysis



Population genetics



Animal research

# Library preparation



Oxford Nanopore has developed VolTRAX – a small device designed to perform library preparation automatically, so that a user can get a biological sample ready for analysis, hands-free. VolTRAX is designed as an alternative to a range of lab equipment, to allow consistent and varied, automated library prep options.

## VolTRAX V2 Starter Pack

\$8,000.00

VolTRAX V2 is designed to automate all laboratory processes associated with Nanopore Sequencing from sample extraction to library preparation.

# MinIT, Analysis



Eliminating the need for a dedicated laptop  
for nanopore sequencing with MinION.

\$2400

## MinIT Specifications:

Pre-installed software: Linux OS, MinKNOW, Guppy, EPI2ME

Bluetooth and Wi-Fi enabled; you can control your experiments using a laptop, tablet or smartphone  
fastq or fast5 files are written to Onboard storage: 512 GB SSD

Processing: GPU accelerators (ARM processor 6 cores, 256 Core GPU), 8 GB RAM.

Small footprint, 290g

1 x USB 2.0 port, 1 x USB 3.0 port and 1 x Ethernet port (1 Gbit capacity)

# SmidgION, Mobile analysis



Oxford Nanopore has now started developing an even smaller device, SmidgION.

**potential applications** may include remote monitoring of pathogens in a breakout or infectious disease; the on-site analysis of environmental samples such as water/metagenomics samples, real time species ID for analysis of food, timber, wildlife or even unknown samples; field-based analysis of agricultural environments, and much more.

# PacBio sequencing and its applications

Rhoads & Au, Gen Prot Bioinf 2015

## Performance comparison of sequencing platforms of various generations

Method	Generation	Read length (bp)	Single pass error rate (%)	No. of reads per run	Time per run	Cost per million bases (USD)	Refs.
Sanger ABI 3730×1	1st	600–1000	0.001	96	0.5–3 h	500	[14,18–21]
Ion Torrent	2nd	200	1	$8.2 \times 10^7$	2–4 h	0.1	[15,25]
454 (Roche) GS FLX +	2nd	700	1	$1 \times 10^6$	23 h	8.57	[14,17,27]
Illumina HiSeq 2500 (High Output)	2nd	2 × 125	0.1	$8 \times 10^9$ (paired)	7–60 h	0.03	[9,16,26]
Illumina HiSeq 2500 (Rapid Run)	2nd	2 × 250	0.1	$1.2 \times 10^9$ (paired)	1–6 days	0.04	[9,16,26]
SOLiD 5500×1	2nd	2 × 60	5	$8 \times 10^8$	6 days	0.11	[14,24]
PacBio RS II: P6-C4	3rd	1.0–1.5 × 10 <sup>4</sup> on average	13	$3.5\text{--}7.5 \times 10^4$	0.5–4 h	0.40–0.80	[5,12,15]
Oxford Nanopore MinION	3rd	2–5 × 10 <sup>3</sup> on average	38	$1.1\text{--}4.7 \times 10^4$	50 h	6.44–17.90	[22,23]

# Characteristics, strengths and weaknesses of commonly used sequencing platforms

Table 2

Characteristics, strengths and weaknesses of commonly used sequencing platforms

Platform \ Instrument	Throughput range (Gb) <sup>a</sup>	Read length (bp)	Strength	Weakness
<i>Sanger sequencing</i>				
ABI 3500/3730	0.0003	Up to 1 kb	Read accuracy and length	Cost and throughput
<i>Illumina</i>				
MiniSeq	1.7–7.5	1×75 to ×150	Low initial investment	Run and read length
MiSeq	0.3–15	1×36 to 2×300	Read length, scalability	Run length
NextSeq	10–120	1×75 to 2×150	Throughput	Run and read length
HiSeq (2500)	10–1000	×50 to ×250	Read accuracy, throughput,	High initial investment, run
NovaSeq 5000/6000	2000–6000	2×50 to ×150	Read accuracy, throughput	High initial investment, run
<i>IonTorrent</i>				
PGM	0.08–2	Up to 400	Read length, speed	Throughput, homopolymers <sup>c</sup>
S5	0.6–15	Up to 400	Read length, speed,	Homopolymers <sup>c</sup>
Proton	10–15	Up to 200	Speed, throughput	Homopolymers <sup>c</sup>
<i>Pacific BioSciences</i>				
PacBio RSII	0.5–1 <sup>b</sup>	Up to 60 kb	Read length, speed (Average 10 kb, N50 20 kb)	High error rate and initial
Sequel	5–10 <sup>b</sup>	Up to 60 kb	Read length, speed (Average 10 kb, N50 20 kb)	High error rate
<i>Oxford Nanopore</i>				
MinION	0.1–1	Up to 100 kb	Read length, portability	High error rate, run length,

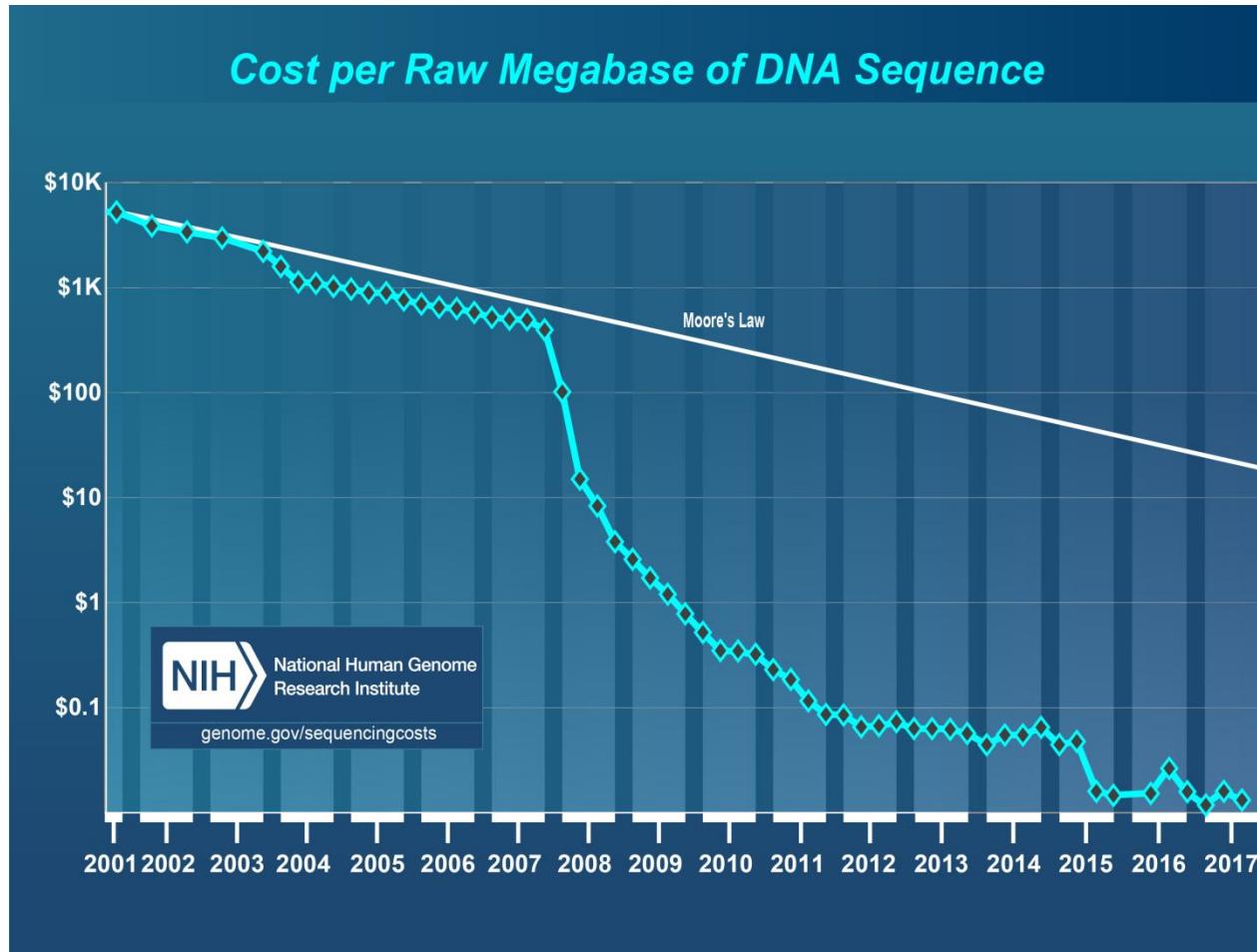
<sup>a</sup> The throughput ranges are determined by available kits and run modes on a per run basis. As an example of a 15-GB throughput, thirty-five 5-MB genomes can be sequenced to a minimum coverage of 40× on the Illumina MiSeq using the v3 600 cycle chemistry.

<sup>b</sup> Per one single-molecule real-time cell.

<sup>c</sup> Results in increased error rate (increased proportion of reads containing errors among all reads) which in turn results in false-positive variant calling.

Besser et al., Clin Micr Infect, 2018

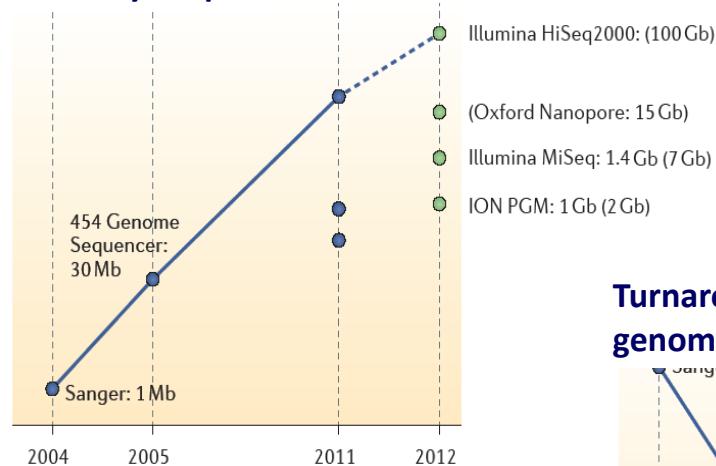
# Sequencing cost



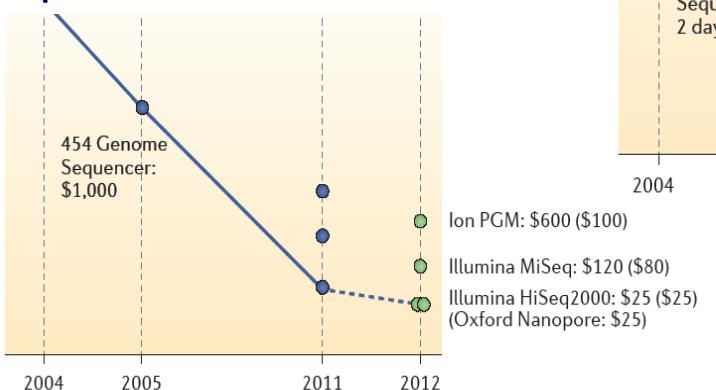
Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

# Sequencing platforms in Microbiology

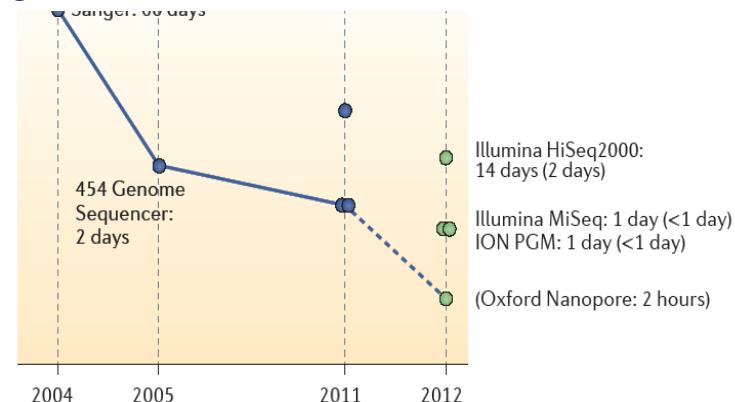
**Raw daily output**



**Cost per Mb assembled sequence**



**Turnaround time: bacterial genome**

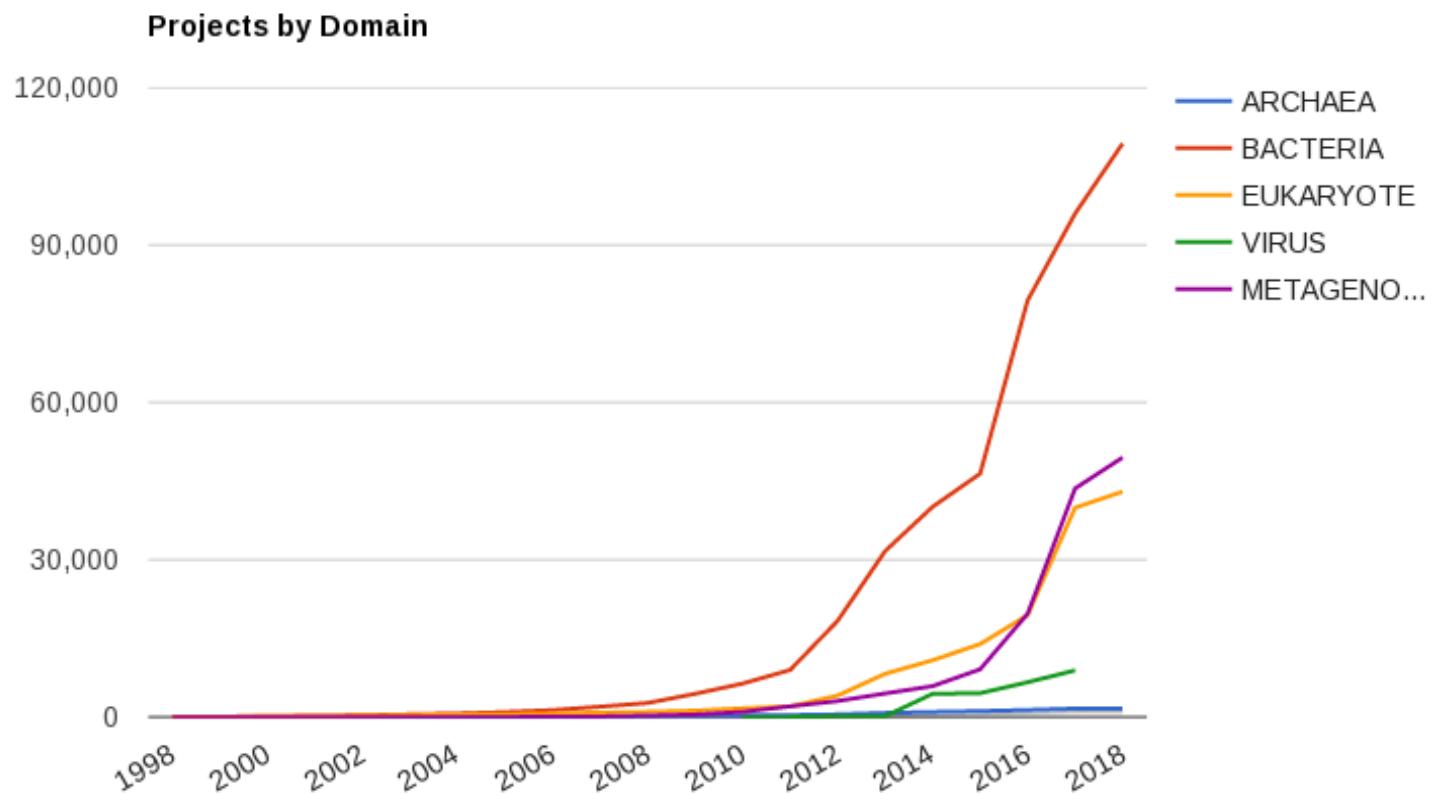


2

# Sequencing projects

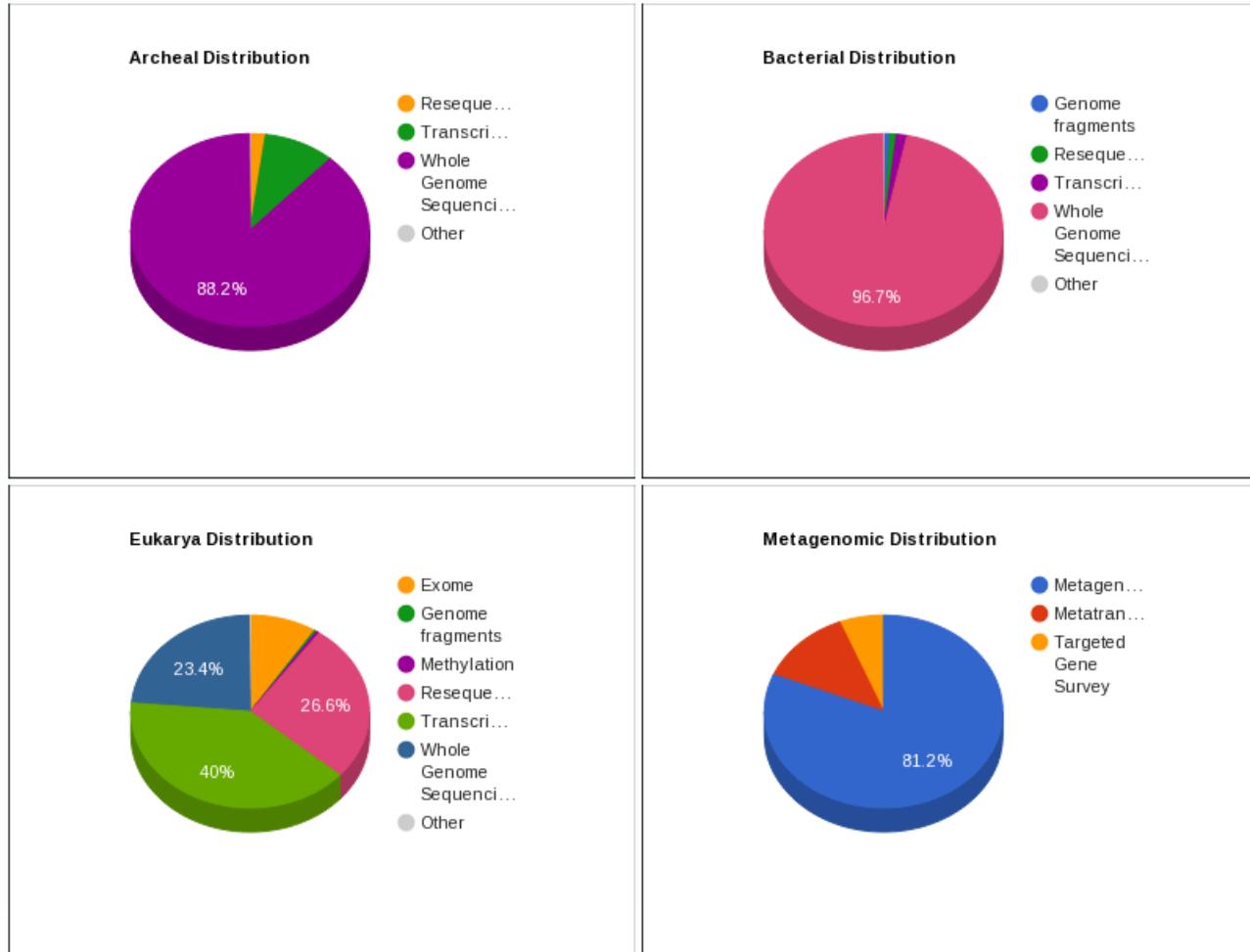
<https://gold.jgi.doe.gov/>

## GOLD, Genome Online DataBase



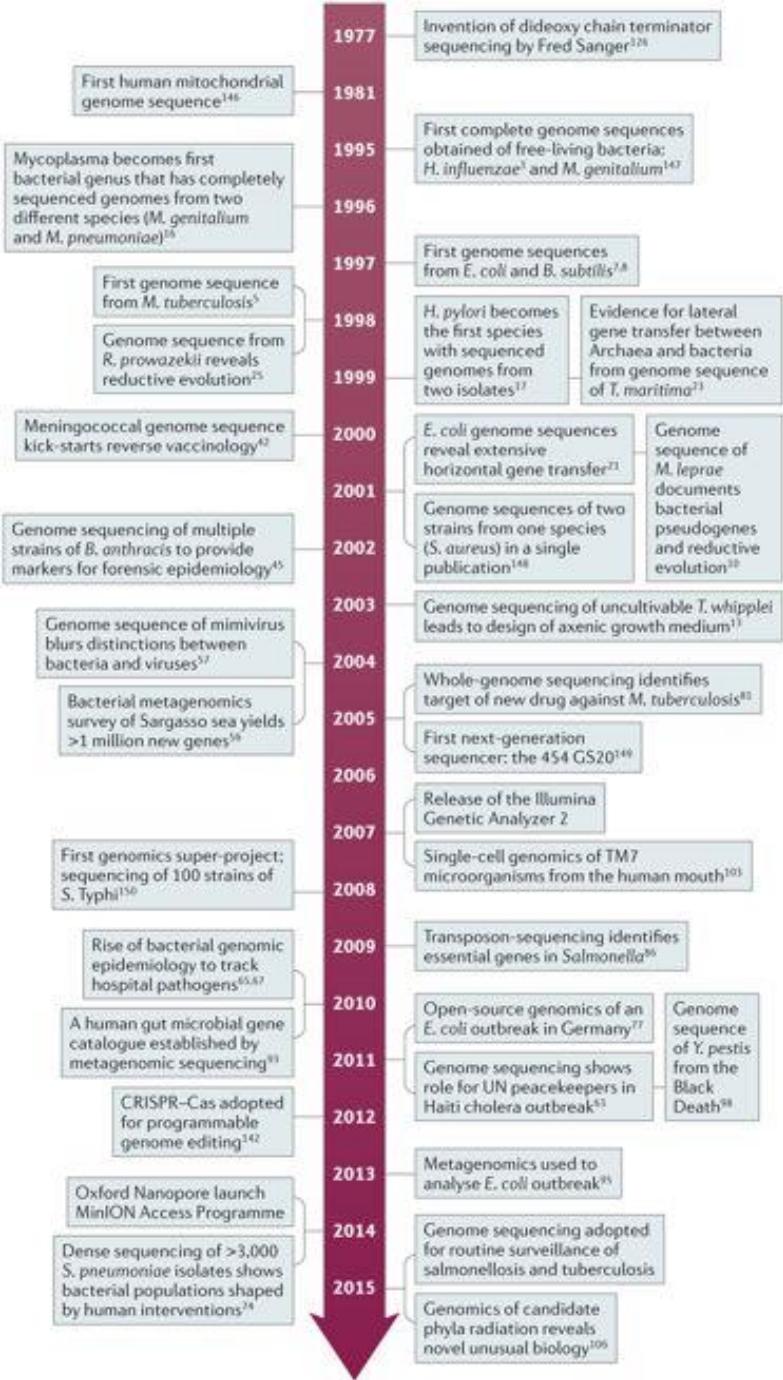
# Organisms distribution of Sequencing projects

GOLD Project Distributions



Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

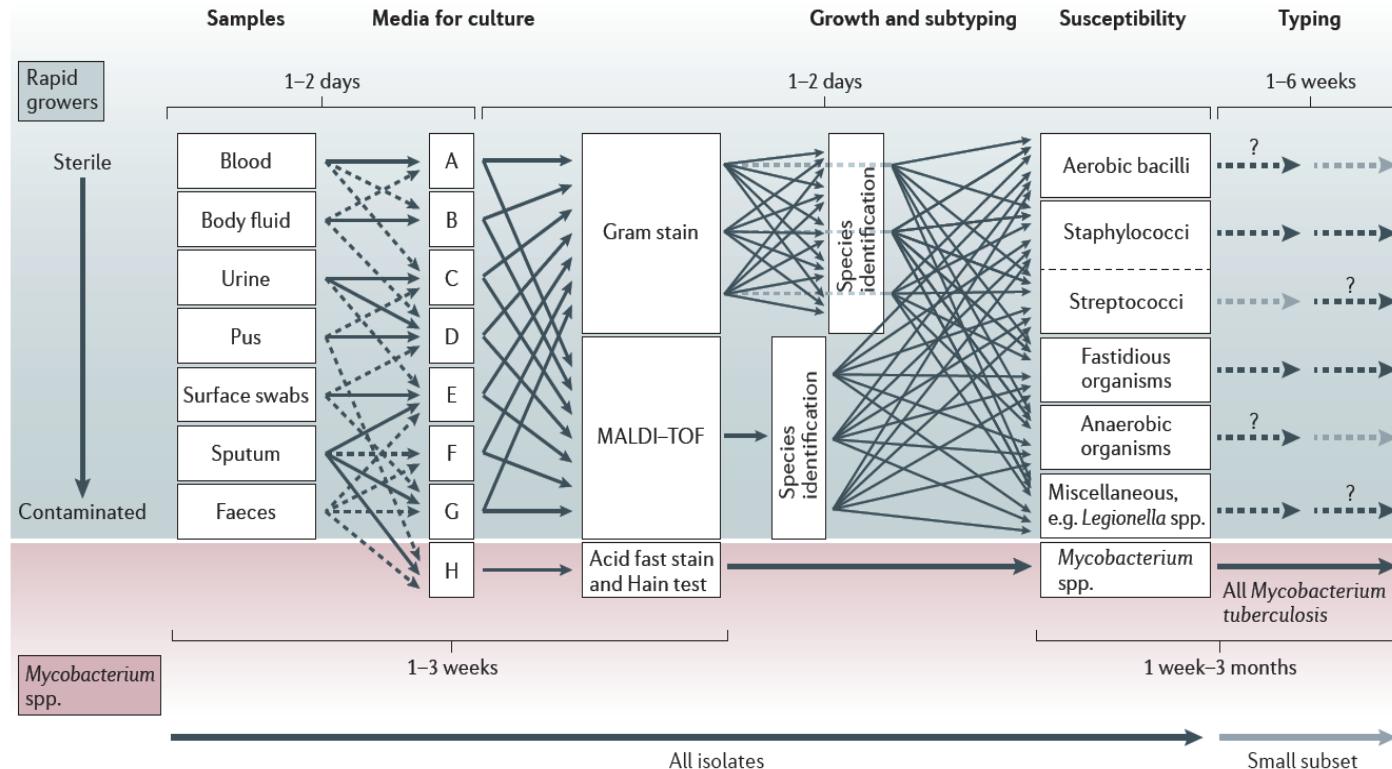
# Brief history of the major events that have shaped the sequencing and analysis of bacterial genomes in the past two decades



Iniciación de genomas bacterianos:  
 herramientas y aplicaciones

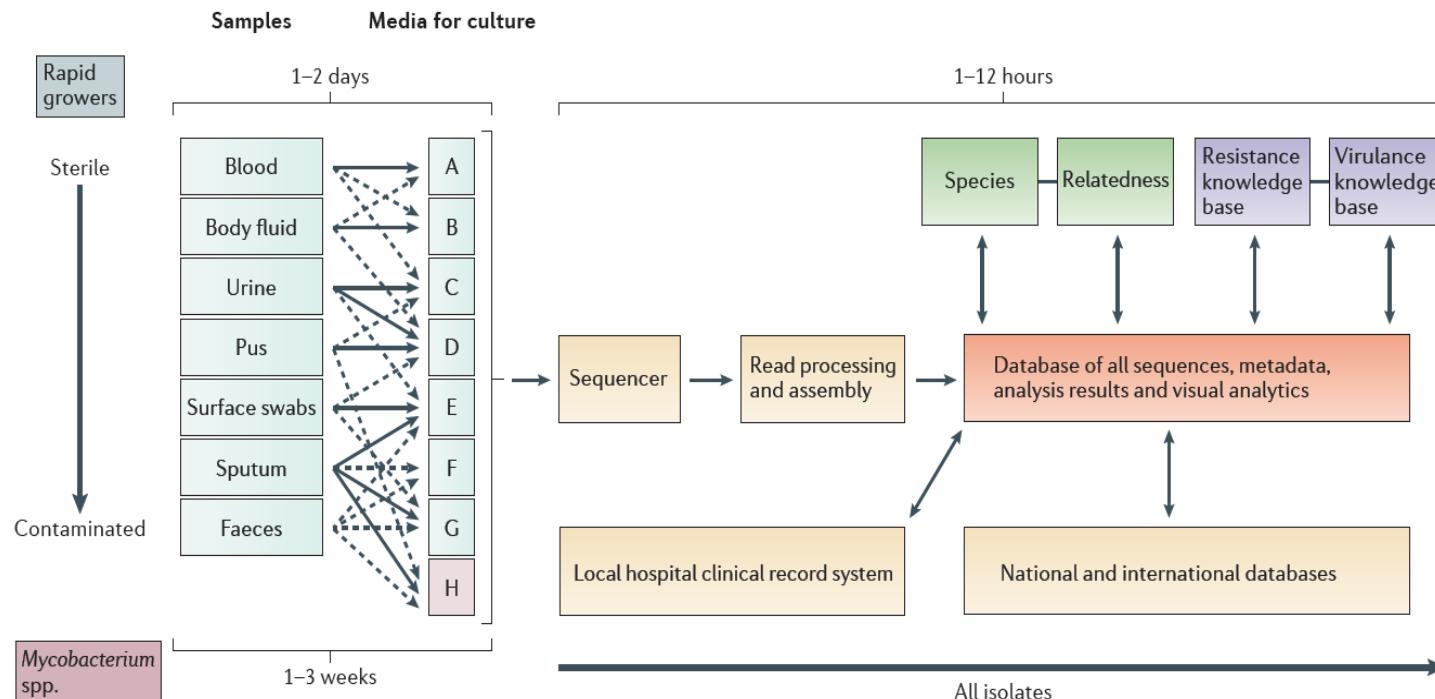
# Workflow for processing samples for bacterial pathogens

Didelot et al., Nature Genet Review 2012, 13:601-612



Ongoing developments in DNA-sequencing technologies are likely to affect the diagnosis and monitoring of all pathogens, including viruses, bacteria, fungi and parasites.

# The diagnostic and clinical applications of bacterial WGS

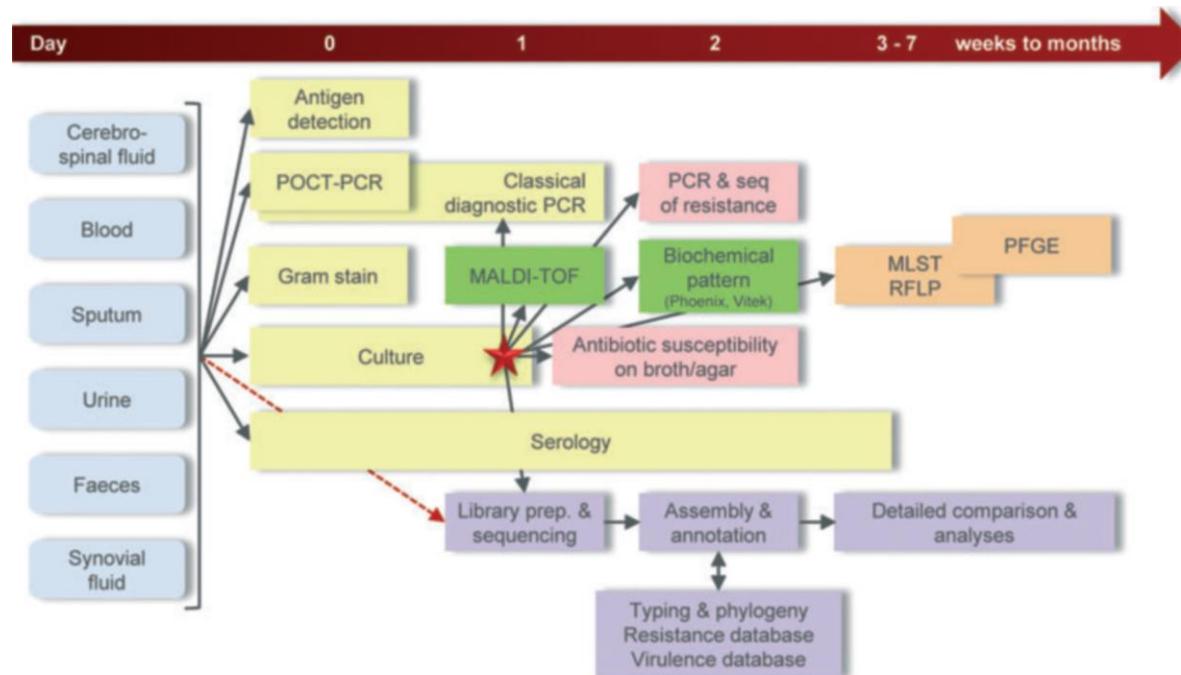


Didelot et al., Nature Genet Review 2012, 13:601-612

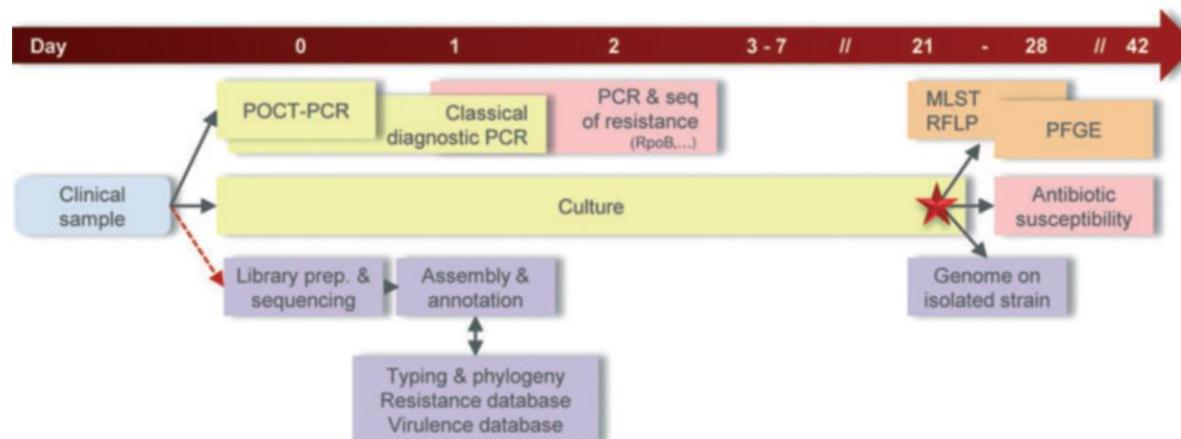
# Schematic representation of the timeline for the processing of clinical samples with classical pathogens

Bertelli and Greub, Clin Microb and Infect, 2013

(a) classical pathogens

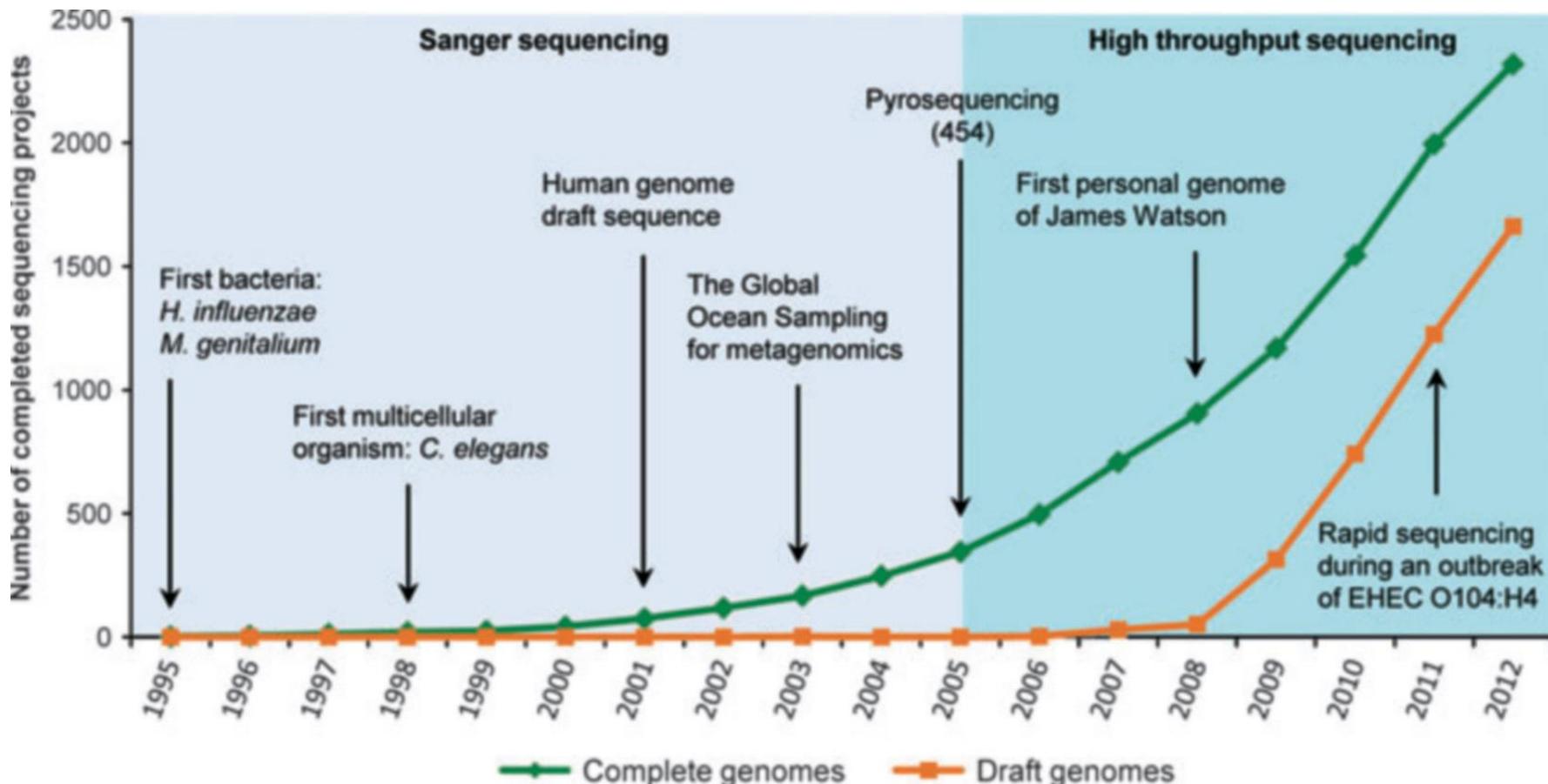


(b) slow-growing bacteria such as *Mycobacterium tuberculosis*



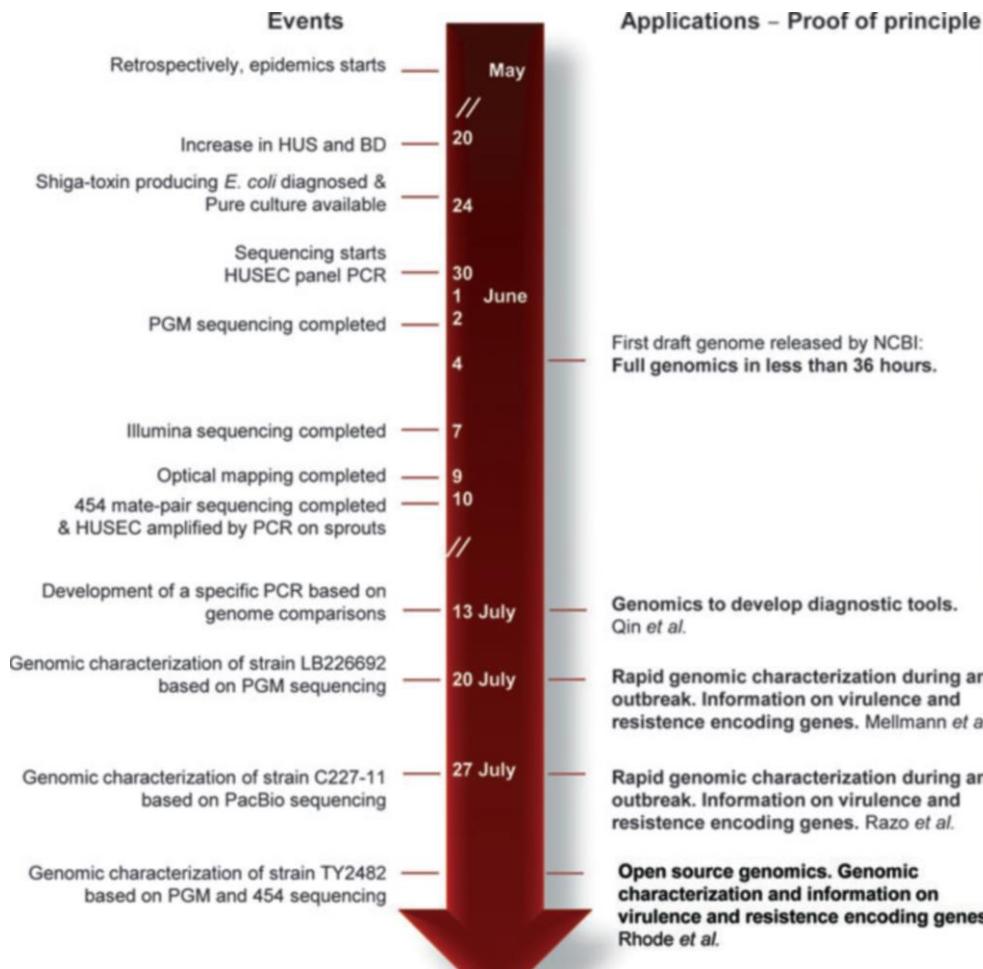
# Milestones in whole genome sequencing

Bertelli and Greub, Clin Micro and Infect, 2013



# The Escherichia coli 0104:H4 epidemics: event timeline and major outputs

Bertelli and Greub, Clin Microb and Infect, 2013



Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

# The idea that via a *One Health* approach infectious diseases could be better controlled and prevented



Global Microbial Identifier



# Global Microbial Identifier



The GMI network consists of approximately 228 experts members from 43 countries, including clinical-, food-, and public health microbiologists and virologists, bioinformaticians, epidemiologists, representatives from funding agencies, data hosting systems, and policy makers from academia, public health, industry, governments.



Technical University of Denmark



World Health Organization



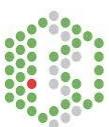
World Organisation for Animal Health



National Institute for Public Health  
and the Environment  
Ministry of Health, Welfare and Sport



EMBL-EBI



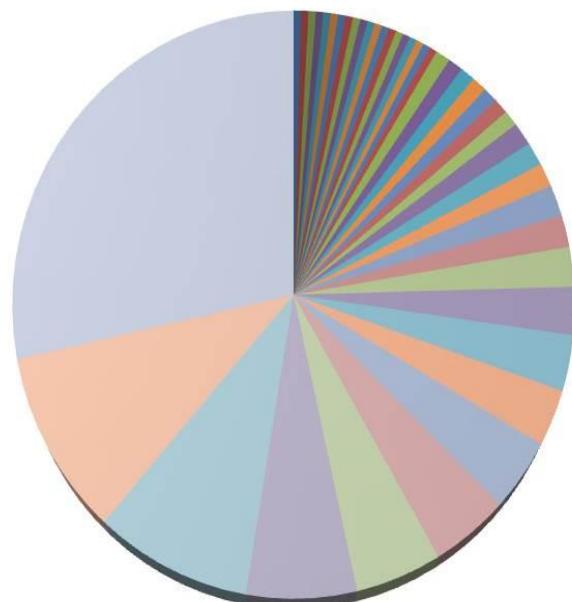
05/11/2018

Secuenciación  
de la  
microbiota  
humana y aplicaciones

/A\_DU-IDATIII

50

43 Countries represented



- Brazil 1
- Bangladesh 1
- Cameroon 1
- Egypt 1
- Ghana 1
- Iran 1
- Israel 1
- Japan 1
- Kenya 1
- Luxembourg 1
- Mauritius 1
- Philippines 1
- Thailand 1
- Senegal 1
- Tunisia 1
- Serbia 1
- Slovenia 1
- Finland 1
- Greece 1
- Chile 2
- Singapore 2
- Tanzania 2
- Malaysia 2
- Norway 2
- Poland 2
- Sweden 2
- India 3
- Germany 3
- Mexico 3
- Nigeria 4
- Switzerland 4
- Belgium 5
- Australia 6
- France 7
- Italy 7
- Canada 9
- The Netherlands 10
- Spain 11
- China 14
- Denmark 20
- United Kingdom 23
- United States 23

# Global Microbial Identifier



GMI envisions a global system of DNA genome databases for microbial and infectious disease identification and diagnostics.

Such a system will benefit those tackling individual problems at the frontline, clinicians, veterinarian, etc., as well as policy-makers, regulators, and industry.

By enabling access to this global resource, a professional response on health threats will be within reach of all countries with basic laboratory infrastructure.

# Global Microbial Identifier



Work group 1: Political challenges, outreach and building a global network

Work group 4: Repository and storage of sequence and meta-data

Work group 3: Analytical approaches

Work group 4: Ring trials and quality assurance

# Global Microbial Identifier, Proficiency test



## 1. Proficiency Testing for bacterial WGS, 2012

an end-user survey of current capabilities, requirements and priorities

## 2. Proficiency Test Pilot, 2014

Wet lab and Dry lab

*Escherichia coli, Staphilococcus aureus and Salmonella typhimurium*

## 3. Full Proficiency Test, 2015

*Escherichia coli, Staphilococcus aureus and Salmonella tiphimurium*

## 4. Full Proficiency Test, 2016

Wet lab and Dry lab

*Campylobacter coli and C. jejuni, Listeria monocytogenes and klebsiella pneumoniae*

## 5. Full Proficiency Test, 2017

Wet Lab and Dry Lab

*Escherichia coli, Staphilococcus aureus and Salmonella tiphimurium*

# Full Proficiency Test, 2015

## Dry Lab, Conclusions



- ❖ 2015 GMI PT highlight the diversity of bioinformatics tools that are being employed around the world to analyze WGS data of bacteria that are of importance to public health and food safety.
- ❖ These methods do not produce the same data objects (variant positions and SNP matrices) from which phylogenetic trees (topologies) are inferred.
- ❖ However, the topologies clustered samples quite similarly (>93% participants clustered samples correctly).
- ❖ A vast majority of labs would reach similar conclusions.
- ❖ Individual centers will be able to define sensible thresholds for determining clusters of isolates.
- ❖ A standardized approach will likely emerge within which thresholds will be decided upon that will facilitate congruence among center-specific pipelines in the conclusions that are reached.

<http://www.globalmicrobialidentifier.org/workgroups/gmi-proficiency-test-reports>



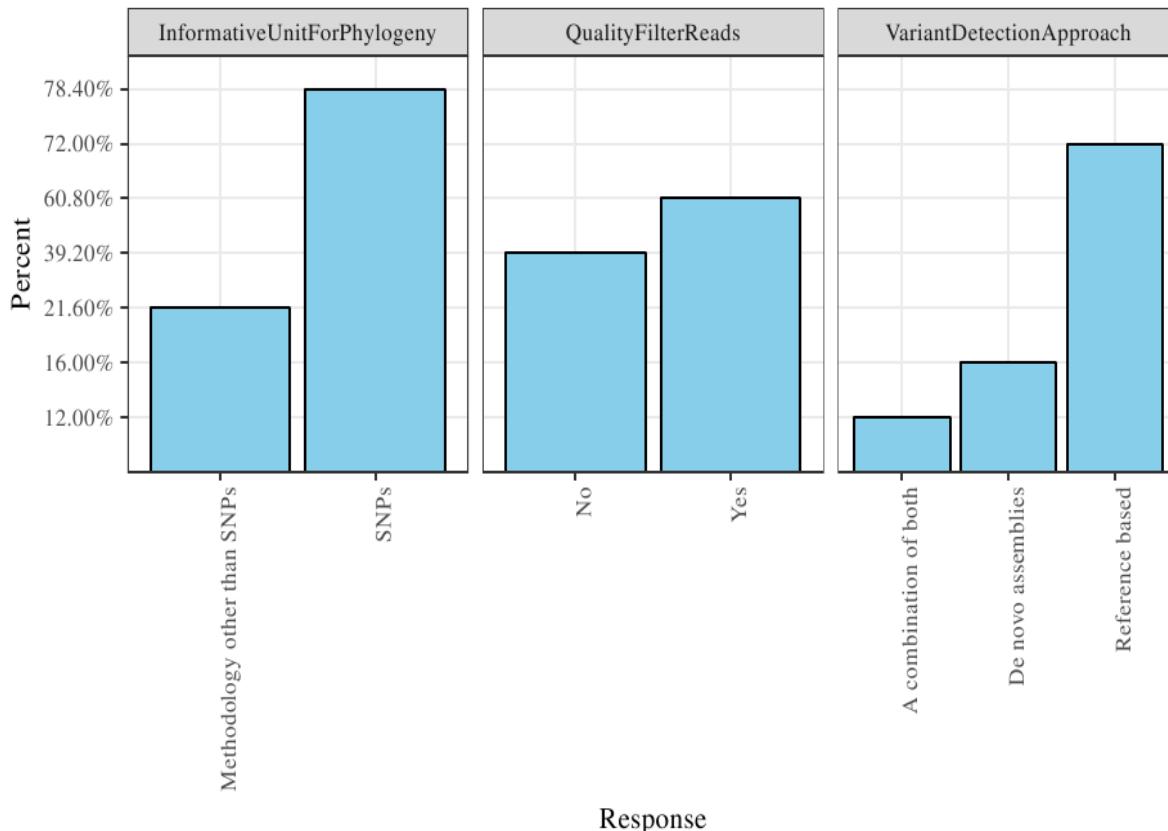
# Full Proficiency Test, 2016

## Dry Lab

### Diversity of the Methods Being Used

Figures 1-3. Charts illustrating the diversity of methods and practices employed for detecting variant from WGS data.

Figure 1. Responses to question about filtering and detection methods



<http://www.globalmicrobialidentifier.org/workgroups/gmi-proficiency-test-reports>

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

05/11/2018

> BU-ISCIII

43

# Full Proficiency Test, 2016

## Dry Lab



Global Microbial Identifier

Lab	CJ	KP	LM
GMI100	1851	126	1652
GMI104	728	20	63
GMI105	55	80	71
GMI106	69	555	235
GMI107	1879	95	146
GMI110	39	1416	77
GMI115	1644	NA	107
GMI116	2084	178	1811
GMI117	747	NA	NA
GMI118	1268	NA	77
GMI66	1516	91	98
GMI67	1516	91	101
GMI70	1696	126	1616
GMI71	NA	NA	294
GMI72	478	87	NA
GMI73	205	87	77
GMI75	1492548	4626433	2566491
GMI77	728	20	63
GMI80	1734	126	144
GMI82	1149	99	NA
GMI83	1619756	5581932	2941727
GMI84	1205	113	699
GMI85	43	97	85
GMI88	1260	167	233
GMI90	NA	105	1422
GMI92	1680	120	957
GMI93	728	253	63
GMI95	648	NA	NA
GMI96	728	20	63
GMI98	1620929	5582195	2941547

Lab	Cluster1
GMI100_CJ	TRUE
GMI104_CJ	TRUE
GMI105_CJ	TRUE
GMI106_CJ	TRUE
GMI107_CJ	TRUE
GMI110_CJ	TRUE
GMI115_CJ	TRUE
GMI116_CJ	TRUE
GMI117_CJ	TRUE
GMI118_CJ	TRUE
GMI66_CJ	TRUE
GMI70_CJ	TRUE
GMI74_CJ	TRUE
GMI75_CJ	TRUE
GMI77_CJ	TRUE
GMI79_CJ	TRUE
GMI80_CJ	TRUE
GMI81_CJ	TRUE
GMI82_CJ	TRUE
GMI84_CJ	TRUE
GMI88_CJ	TRUE
GMI88_CJ	TRUE
GMI92_CJ	TRUE
GMI93_CJ	TRUE
GMI95_CJ	TRUE
GMI96_CJ	TRUE
GMI97_CJ	TRUE
GMI98_CJ	TRUE

Lab	Cluster1	Cluster2
GMI100_KP	TRUE	TRUE
GMI104_KP	TRUE	FALSE
GMI105_KP	TRUE	TRUE
GMI106_KP	FALSE	FALSE
GMI107_KP	TRUE	TRUE
GMI110_KP	FALSE	FALSE
GMI116_KP	TRUE	TRUE
GMI66_KP	TRUE	TRUE
GMI70_KP	TRUE	TRUE
GMI72_KP	TRUE	TRUE
GMI73_KP	TRUE	TRUE
GMI77_KP	TRUE	FALSE
GMI79_KP	FALSE	FALSE
GMI80_KP	TRUE	TRUE
GMI81_KP	TRUE	FALSE
GMI82_KP	TRUE	TRUE
GMI84_KP	TRUE	TRUE
GMI88_KP	TRUE	TRUE
GMI90_KP	TRUE	TRUE
GMI92_KP	TRUE	TRUE
GMI93_KP	TRUE	TRUE
GMI96_KP	TRUE	FALSE
GMI97_KP	FALSE	FALSE
GMI98_KP	FALSE	FALSE

Lab	Cluster1	Cluster2
GMI100_LM	TRUE	TRUE
GMI102_LM	FALSE	TRUE
GMI104_LM	TRUE	TRUE
GMI105_LM	TRUE	TRUE
GMI106_LM	TRUE	TRUE
GMI107_LM	TRUE	TRUE
GMI110_LM	TRUE	FALSE
GMI115_LM	TRUE	TRUE
GMI116_LM	TRUE	TRUE
GMI118_LM	TRUE	TRUE
GMI66_LM	TRUE	TRUE
GMI70_LM	TRUE	TRUE
GMI71_LM	TRUE	TRUE
GMI72_LM	TRUE	TRUE
GMI73_LM	TRUE	TRUE
GMI74_LM	FALSE	TRUE
GMI75_LM	FALSE	TRUE
GMI77_LM	TRUE	TRUE
GMI81_LM	FALSE	TRUE
GMI82_LM	TRUE	TRUE
GMI84_LM	TRUE	TRUE
GMI88_LM	TRUE	TRUE
GMI89_LM	TRUE	TRUE
GMI90_LM	FALSE	TRUE
GMI92_LM	TRUE	TRUE
GMI93_LM	TRUE	TRUE
GMI96_LM	TRUE	TRUE
GMI97_LM	TRUE	FALSE
GMI98_LM	FALSE	TRUE

<http://www.globalmicrobialidentifier.org/www.globalmicrobialidentifier.org/proficiency-test-reports>

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

## Summary and Key Findings

- A total of 215 results files were submitted (fasta or newick tree) (Table 1).
- Not surprisingly, participants differed in how they quality filtered (Fig. 1) and the methods they used to analyze the datasets (Figs. 2 & 3, Table 2).
- The number of positions within the fasta matrices differed greatly (Table 3).
- Despite differences in the size of the matrices and, in some cases, relative differences among samples, the majority of participants created trees that contained the same clusters of isolates (Tables 4 - 6).

<http://www.globalmicrobialidentifier.org/workgroups/gmi-proficiency-test-reports>

Secuenciación de genomas bacterianos:  
herramientas y aplicaciones

05/11/2018

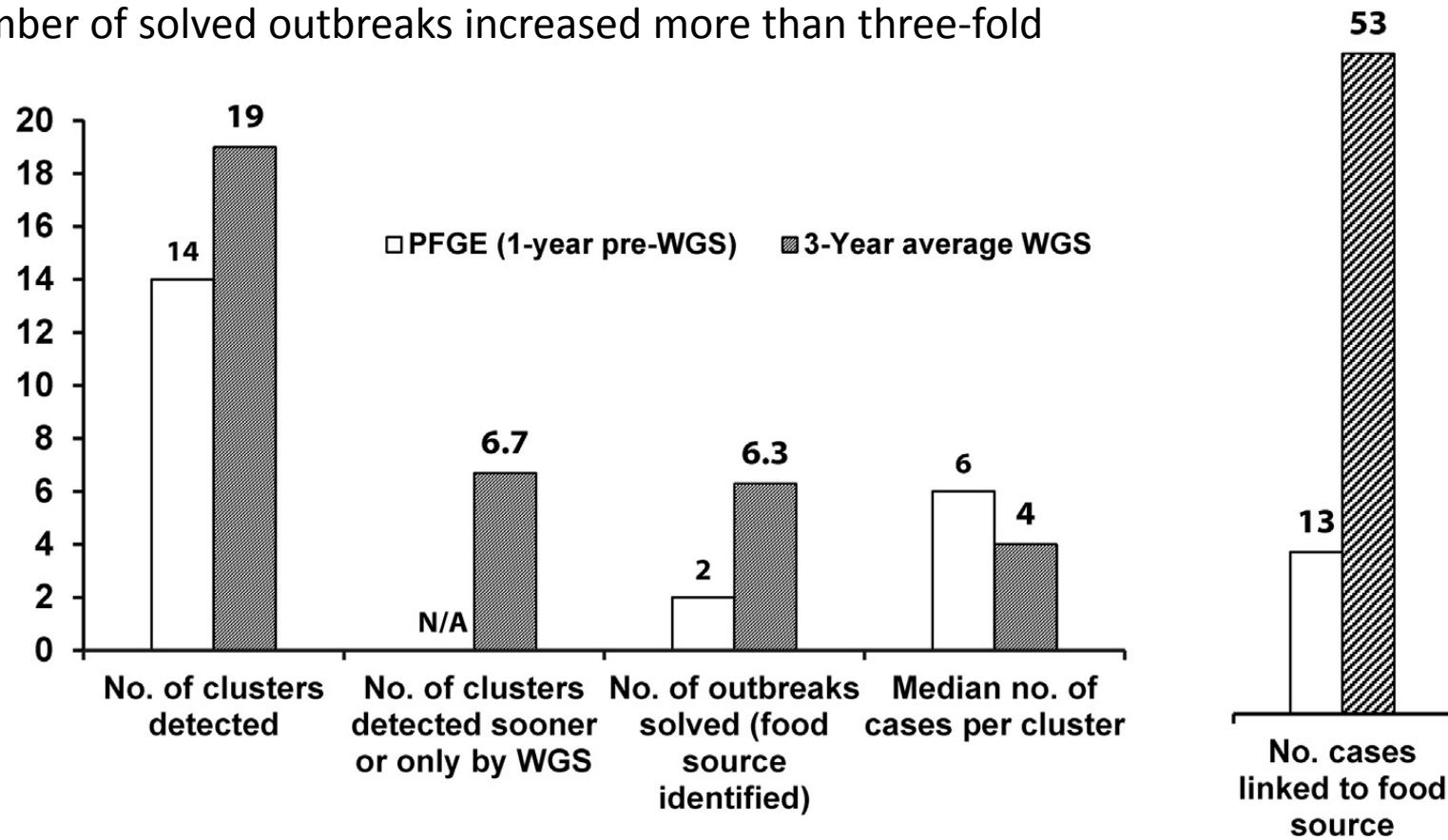
>x\_BU-ISCIII

45

# Early data from surveillance of listeriosis in the USA

Besser et al., Clin Micr Infect, 2018

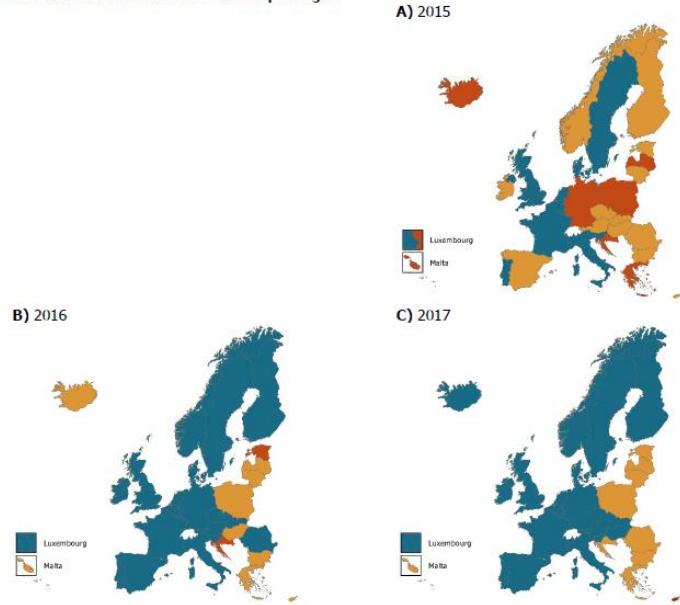
The number of outbreaks detected increased 36% after implementation of real-time WGS based surveillance, and likewise the number of solved outbreaks increased more than three-fold



**WGS provides higher resolution and accuracy** than classical molecular typing methods, such as PFGE or MLVA, contributing to a better understanding of infectious disease and drug resistance transmission patterns and thereby improving the effectiveness of interventions for their control.

# ECDC technical report: Monitoring the use of wgs in infectious disease surveillance in Europe 2015-2017

**Figure 1.** National public health reference laboratories use of WGS-based typing for national surveillance of at least one human pathogen



**Figure 3.** Number of EU/EEA countries using WGS-based typing as first or second-line method for routine surveillance and outbreak investigations in National Public Health Reference Laboratories by disease group and pathogen, 2017

**Foodborne pathogens**  
*Listeria monocytogenes*  
*Salmonella enterica*  
*Shiga toxin-producing E. coli (STEC)*

**Antimicrobial-resistant pathogens**  
 Carbapenemase-producing Enterobacteriaceae (CPE)  
 Antibiotic resistant *Neisseria gonorrhoeae*  
 Multidrug-resistant *Mycobacterium tuberculosis*

**Vaccine-preventable pathogens**  
 Invasive *Neisseria meningitidis*  
 Human Influenza virus

Number of countries by use of WGS-typing in 2017 or being planned by 2019, per pathogen						
4	8	3	6	2	9	5
4	10		9	2		8
7	8	5	2			7
2	4	12	2	3		7
1	10	7	6	1	5	
10	7	3	4		6	
4	9	2	8		7	
6	14	2	5	3		

- Legend
- No information
  - No, it is not used for public health operations or planned by 2019
  - No, it is not used for public health operations or planned by 2019
  - Only for outbreak investigations
  - Routine surveillance and outbreak investigations - Second-line typing
  - Routine surveillance and outbreak investigations - First-line typing

# ECDC technical report: Monitoring the use of wgs in infectious disease surveillance in Europe 2015-2017

Figure 6. EU/EEA national public health reference laboratories' use of NGS technologies for typing (routine surveillance and outbreak investigation), as of July 2017

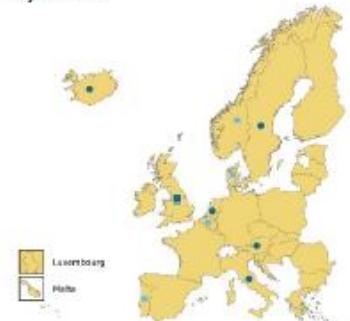
A) Listeria monocytogenes



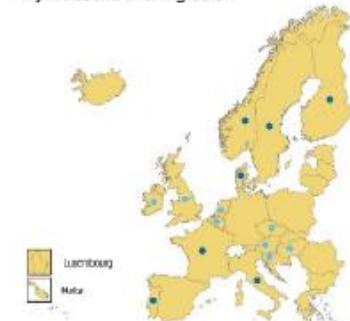
B) Salmonella enterica



A) MDR-TB



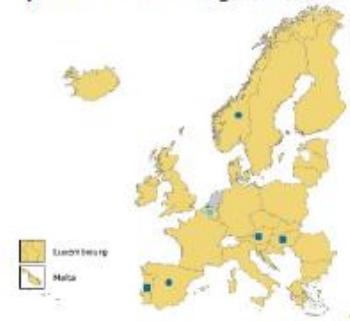
B) Neisseria meningitidis



C) STEC



C) Antibiotic-resistant *N. gonorrhoeae*



D) Carbapenemase-producing Enterobacteriaceae



E) Human influenza virus



Yellow: Participation in the survey; grey: No participation in the survey; sampling frame (comprehensive – circle and sentinel/subset of samples – square) and typing scheme (first line – dark blue and second line – light blue)

# ECDC technical report: Monitoring the use of wgs in infectious disease surveillance in Europe 2015-2017

**Table 1.** Number of EU/EEA countries with one or more national public health reference laboratories having access to next-generation sequencing (NGS) technologies for routine public health operations, by technology and instrument used, 2017

NGS technology	Instrument	Foodborne pathogens			Antimicrobial-resistant pathogens		Vaccine-preventable diseases	
		<i>L. monocytogenes</i>	<i>S. enterica</i>	STEC	CPE	AR-N. gonorrhoeae	MDR-TB	<i>N. meningitidis</i>
Illumina	HiSeq series	3	3	3	2	1	2	3
	HiSeq X series						1	
	MiniSeq	1	2	3	2		4	1
	MiSeq series	12	10	7	7	10	7	13
	NextSeq	2	2	2	3	1	3	2
Ion Torrent	S4							1
	S5	1	1	1	1		1	
	S5 XL							1
	PGM		1	1		1		1
	Proton						1	1
Oxford Nanopore Technologies	MinION		1	1	2		1	1
Pacific Biosciences-PacBio	PacBio RS II		1		2			
Other not specified	-		3	2	2	1	1	1

**Table 2.** Bioinformatics tools used by the National Public Health Reference Laboratories using WGS-based typing for surveillance and outbreak investigations of foodborne pathogens, July 2017\*

	Number of EU/EEA countries		
Tools used for sequence analysis	<i>L. monocytogenes</i> (n=14)	<i>S. enterica</i> (n=7)	STEC (n=9)
Commercial software	9	4	4
Open source software	4	3	5
In-house suite of customised tools	4	2	2

# ECDC technical report: Monitoring the use of wgs in infectious disease surveillance in Europe 2015-2017

**Table 3. Number of EU/EEA countries using WGS-based typing for surveillance and outbreak investigations in the national public health reference laboratories and respective typing scheme, sampling frame, bioinformatics analysis, and raw data storage practice by pathogen, 2017**

	2017	Foodborne pathogens			Antimicrobial resistant pathogens			Vaccine preventable pathogens	
		<i>L. monocytogenes</i>	<i>S. enterica</i>	STE C	CP E	AR- <i>N. gonorrhoeae</i>	MDR TB	Human influenza virus	<i>N. meningitidis</i>
Number of countries using WGS for routine surveillance and outbreak investigations		14	7	9	10	6	10	8	15
Typing scheme	First-line WGS	9	5	8	7	5	6	3	7
	Second-line WGS	5	2	1	3	1	4	5	8
Sampling frame	Continuous comprehensive	12	6	8	7	2	9	-	15
	Sentinel/ subset of case samples	2	1	1	3	4	1	8	-
Bioinformatic analysis *	cgMLST	12	6	5	6	4	5	-	12
	SNP	7	5	5	5	2	7	-	5
	Resistome prediction	4	5	7	8	4	6	-	3
	wgMLST	5	3	3	2	2	2	-	2
	Virulome/ mobilome prediction	4	2	9	5	1	-	-	1
	MLST prediction	12	6	8	3	-	-	-	2
	Serogroup prediction	7	6	9	1	-	-	-	2
	NG-MAST	-	-	-	-	3	-	-	-
	Speciation	-	1	1	-	-	3	-	1
	Hemagglutinin and neuraminidase sequence prediction	-	-	-	-	-	-	4	-
	Phylogenetic relationship	-	1	1	1	-	-	7	1
	Identification of specific point mutations	-	1	1	-	-	-	6	1
	rMLST	-	-	-	-	-	-	-	5
	MLST+ <i>porA</i> VR1 and VR2+ <i>fetA</i>	-	-	-	-	-	-	-	12
	Vaccine antigen prediction	-	-	-	-	-	-	-	9
	Other not specified	-	-	-	1	-	3	3	1
Raw sequence data storage *	Dedicated closed database(s)	13	5	7	10	6	10	6	12
	Publicly available database(s)	1	2	2	-	1	1	2	3

05/11/2 \* Not mutually exclusive

## Conclusions

This emerging mainstream practice should enable **pan-European WGS-derived data exchange in the medium-term**, subject to **harmonisation** of sequence analysis pipelines for output compatibility, agreement on international WGS derived type nomenclature and development of **secure and efficient international data sharing** and management platforms.

Current bottlenecks mainly relate to development of expertise in **epidemiological-WGS data integrative analysis** and access to user-friendly international nomenclature

# Skills needed to translate WGS data into public health action

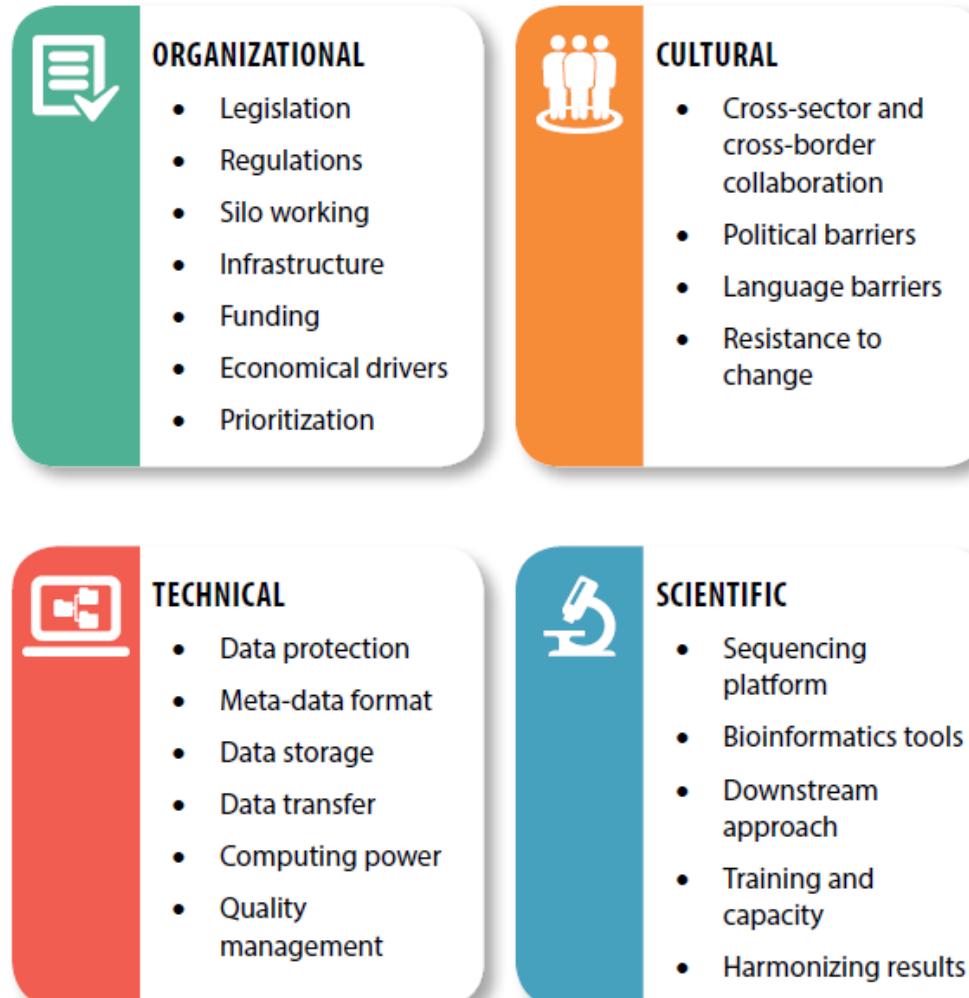


World Health Organization

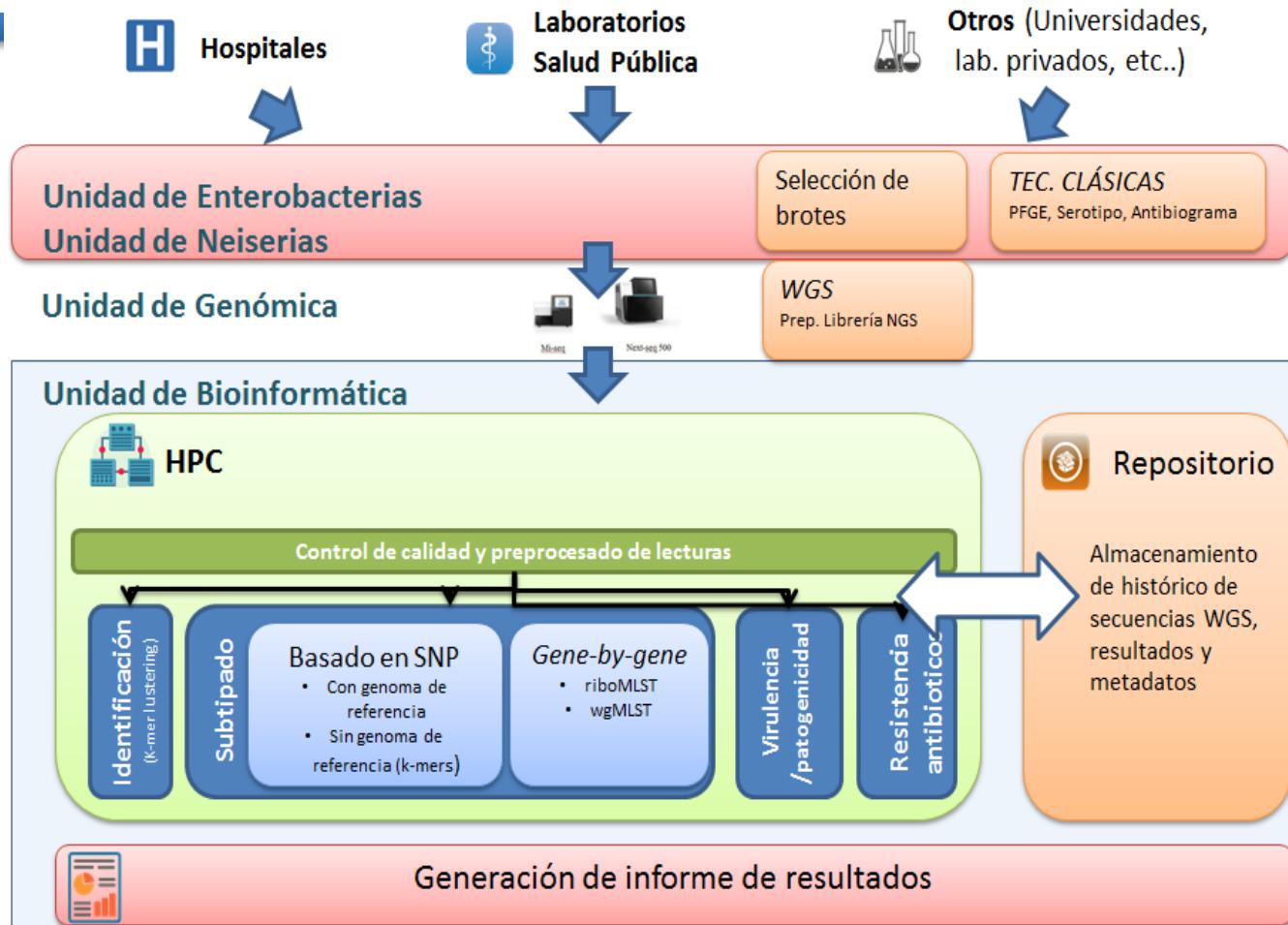
Bioinformatician	Epidemiologist	Microbiologist
Algorithms for genome mapping, assembly and comparisons	Epidemiology of communicable diseases	Microbiological diagnostics
Inferences from genomic data	Statistical analysis	Subtyping of pathogens
Genomic data handling and processing	Case-control studies	Pathogen genomics and evolution
Genome data visualization and integration	Health data linkage	Access to culture collections with epidemiological context
	Risk assessment and communication	

FIGURE 2.1

## Challenges of coordinating WGS for integrated food chain surveillance

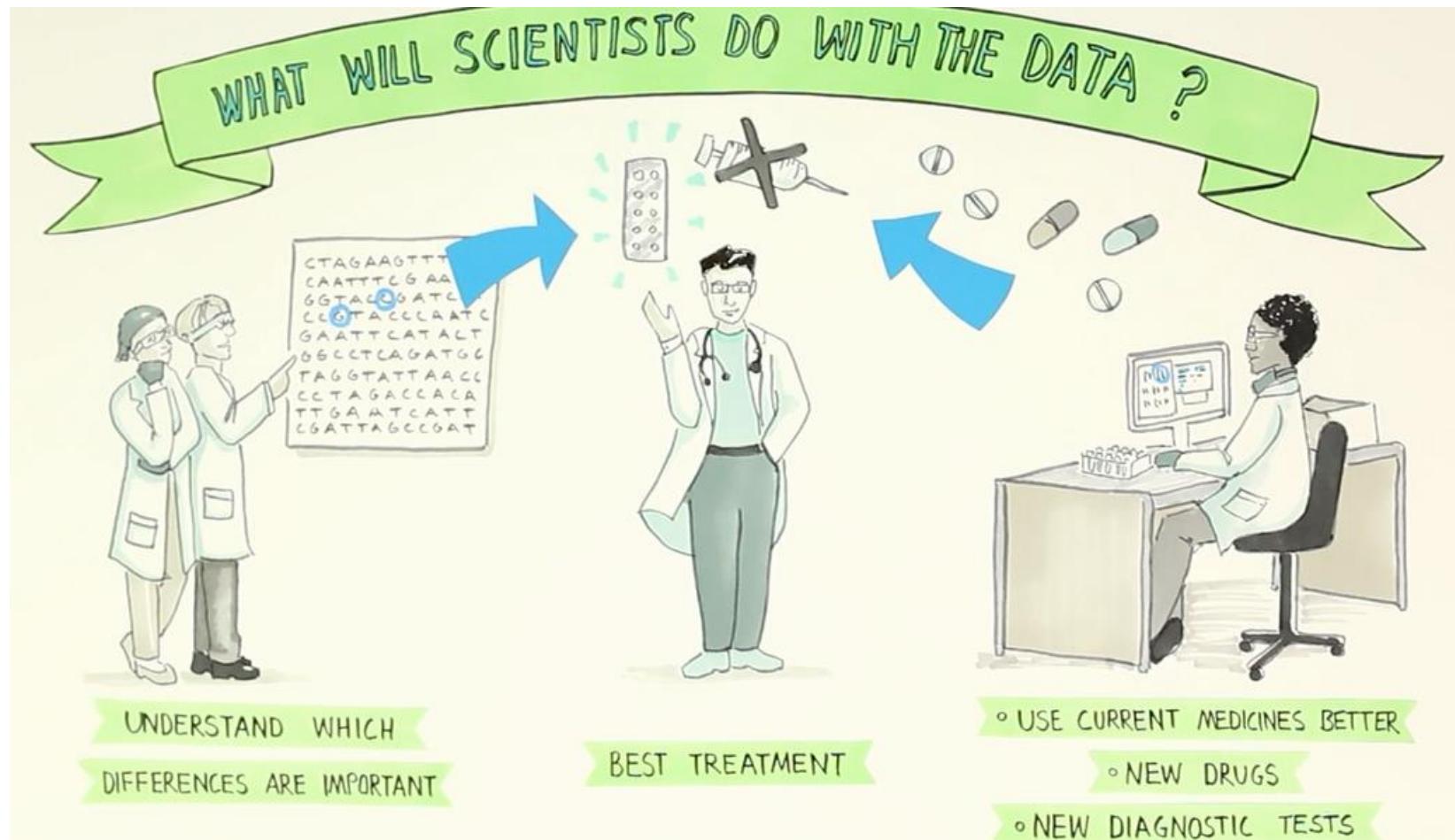


# ISCIII Project Ongoing, 2017-2019

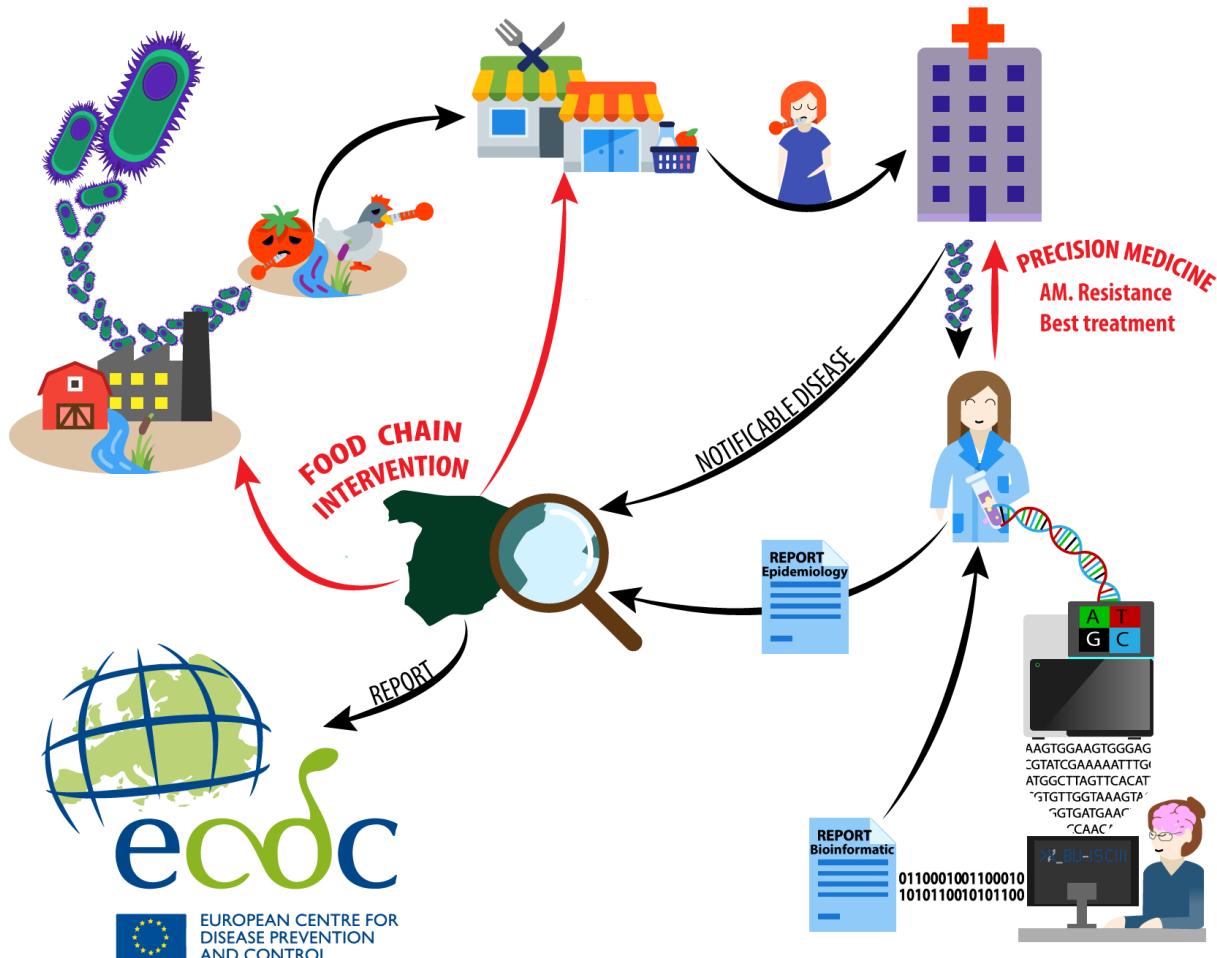


# Precision Medicine

<https://labiotech.eu/features/genome-sequencing-review-projects/>



# WGS for infectious diseases



Secuenciación de genomas bacterianos:  
herramientas y aplicaciones



GOBIERNO  
DE ESPAÑA  
MINISTERIO  
DE CIENCIA, INNOVACIÓN  
Y UNIVERSIDADES

Thanks for your attention!

Questions???