

Bioinformática aplicada a la Microbiología Clínica

Isabel Cuesta y Sara Monzón

UNIDAD DE BIOINFORMÁTICA (BU-ISCIII)

Unidades Centrales Científico Técnicas – SGSAFI-ISCIII

18 Marzo 2022

Master Bioinformática aplicada a la Medicina Personalizada y la Salud



cnio stop cancer



Index

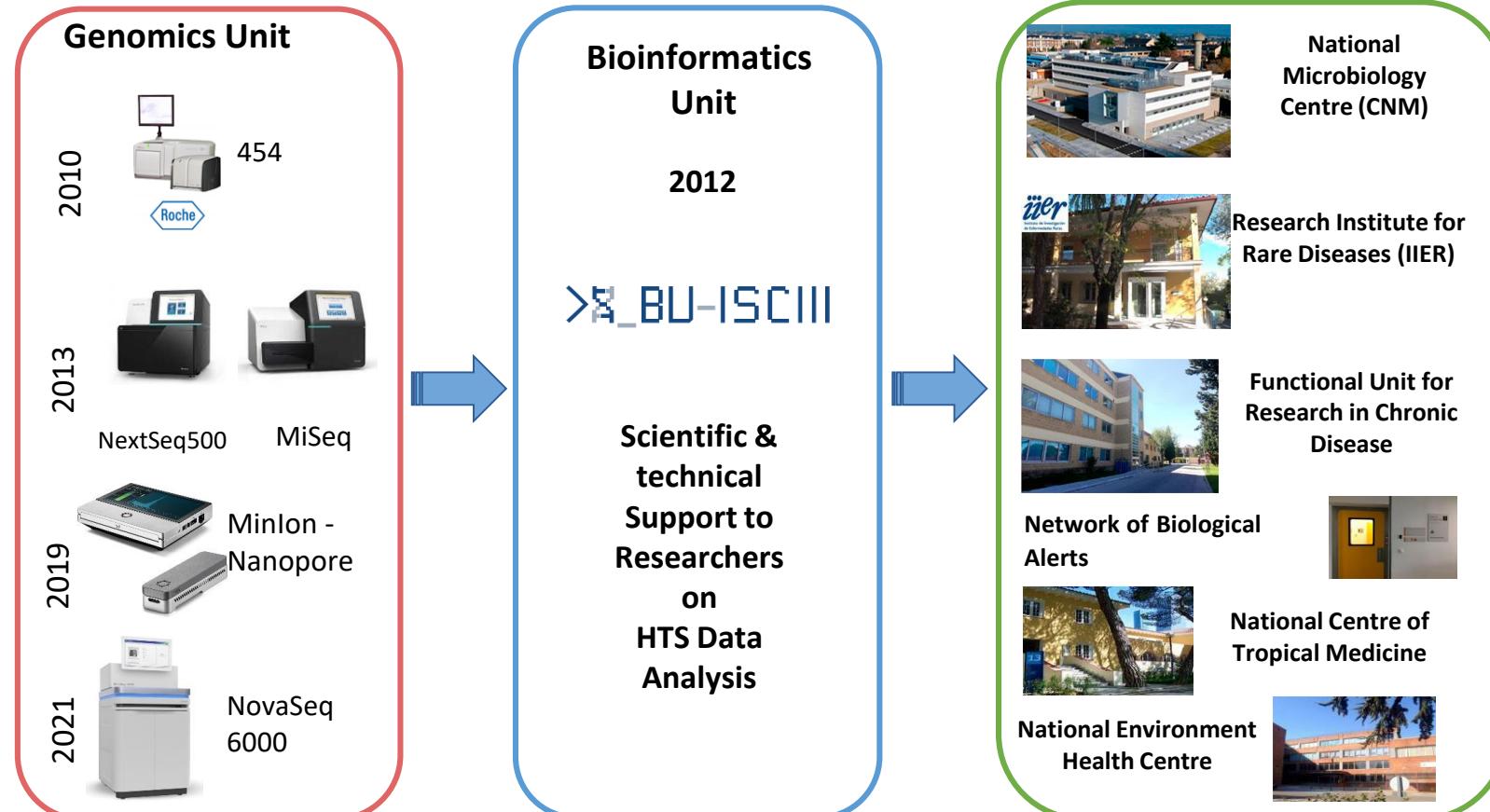
- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology
- Concepts: HTS and Outbreak investigation
- Bacterial and Viral Genome Sequencing
- Bioinformatics analysis in microbial genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

Index

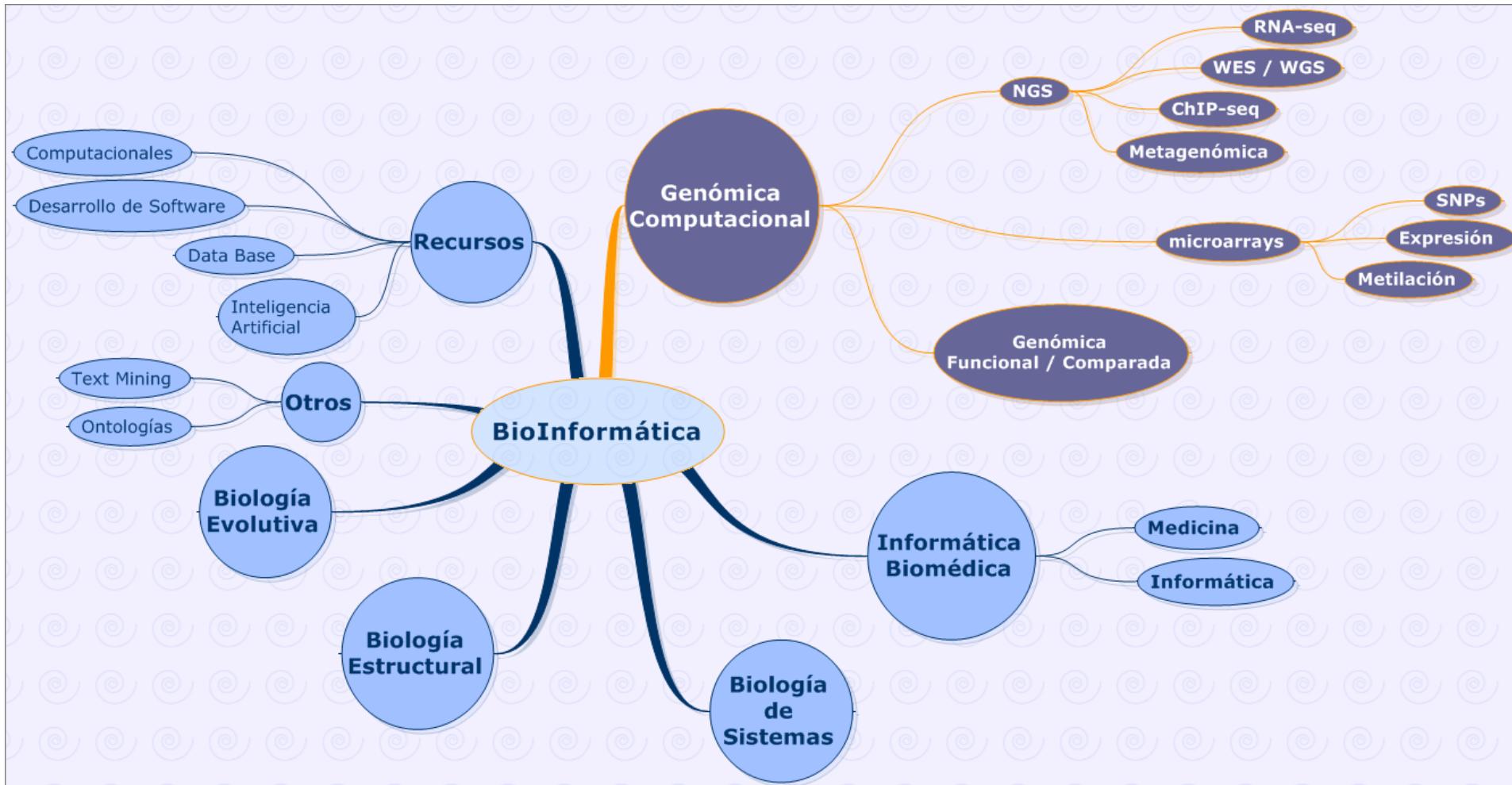
- **BU-ISCIII**

- High-throughput sequencing (HTS) applications in Microbiology
- Concepts: HTS and Outbreak investigation
- Bacterial and Viral Genomics
- Bioinformatics analysis in microbial genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

BU-ISCIII: As example of Clinical Bioinformatics Unit



BU-ISCIII Mission - Activities



- TEAM: Bioinformatician & IT
- SOURCE OF FUNDING:
 - Research Project
 - National or International Calls
 - Permanent position – Civil Servant

RRHH



COMPUTATIONAL RESOURCES & PIPELINES DEVELOPMENT



SERVICES PORTFOLIO & TRAINING



- COURSES, TFM, TFG
- DATA ANALYSIS
 - DNAseq
 - RNAseq
 - Metagenomics

- INFRASTRUCTURE
 - HPC
 - Workstations
- REPOSITORY:
 - GitHub
 - Benchmarking
 - Validation

DATA MANAGEMENT



- INFRASTRUCTURE
- DATA WORKFLOW
- LIMS

Computational Resources

- IT support: establish agreement with IT department including permission for using Linux.



Workstations (5), 4cores, 64Gb, 8TB
Server, 4-quad, 120Gb, 16TB

Data Centre (CPD-ISCIII)



HPC 320 cores, 8TB RAM, 10Gbps.
2 flexible and scalable storages,
NetApp, 70 TB and 250TB

- Reproducibility of in-silico pipelines analysis

nextflow



Singularity containers
Admin support & environment independency
Sharing code easier

 GitHub

<https://github.com/BU-ISCIII>

- TEAM: Bioinformatician & IT
- SOURCE OF FUNDING:
 - Research Project
 - National or International Calls
 - Permanent position – Civil Servant

RRHH



COMPUTATIONAL
RESOURCES &
PIPELINES
DEVELOPMENT



SERVICES
PORTFOLIO
&
TRAINING



- COURSES, TFM, TFG
- DATA ANALYSIS
 - DNAseq
 - RNAseq
 - Metagenomics

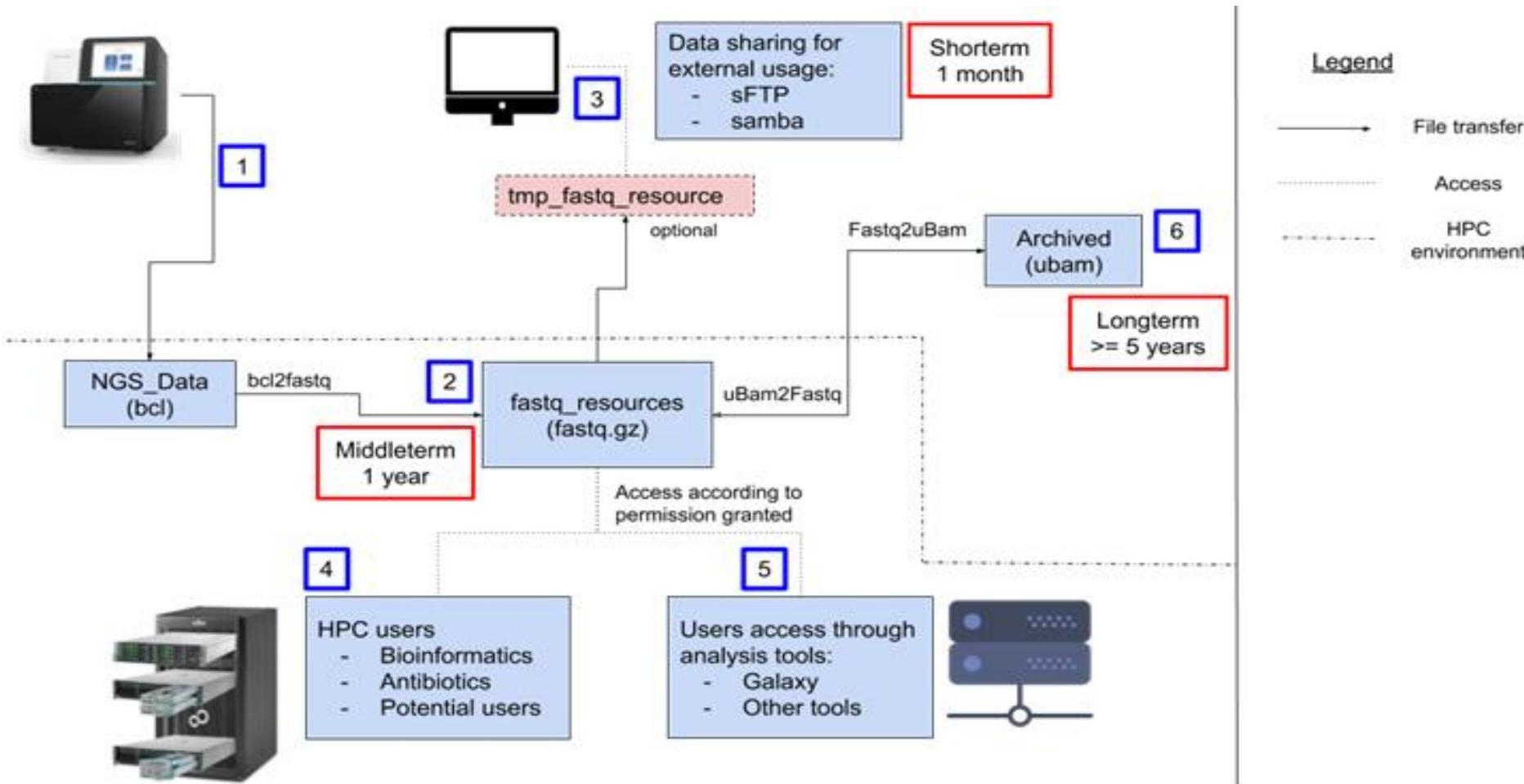
- INFRASTRUCTURE
 - HPC
 - Workstations
- REPOSITORY:
 - GitHub
 - Benchmarking
 - Validation

DATA
MANAGEMENT



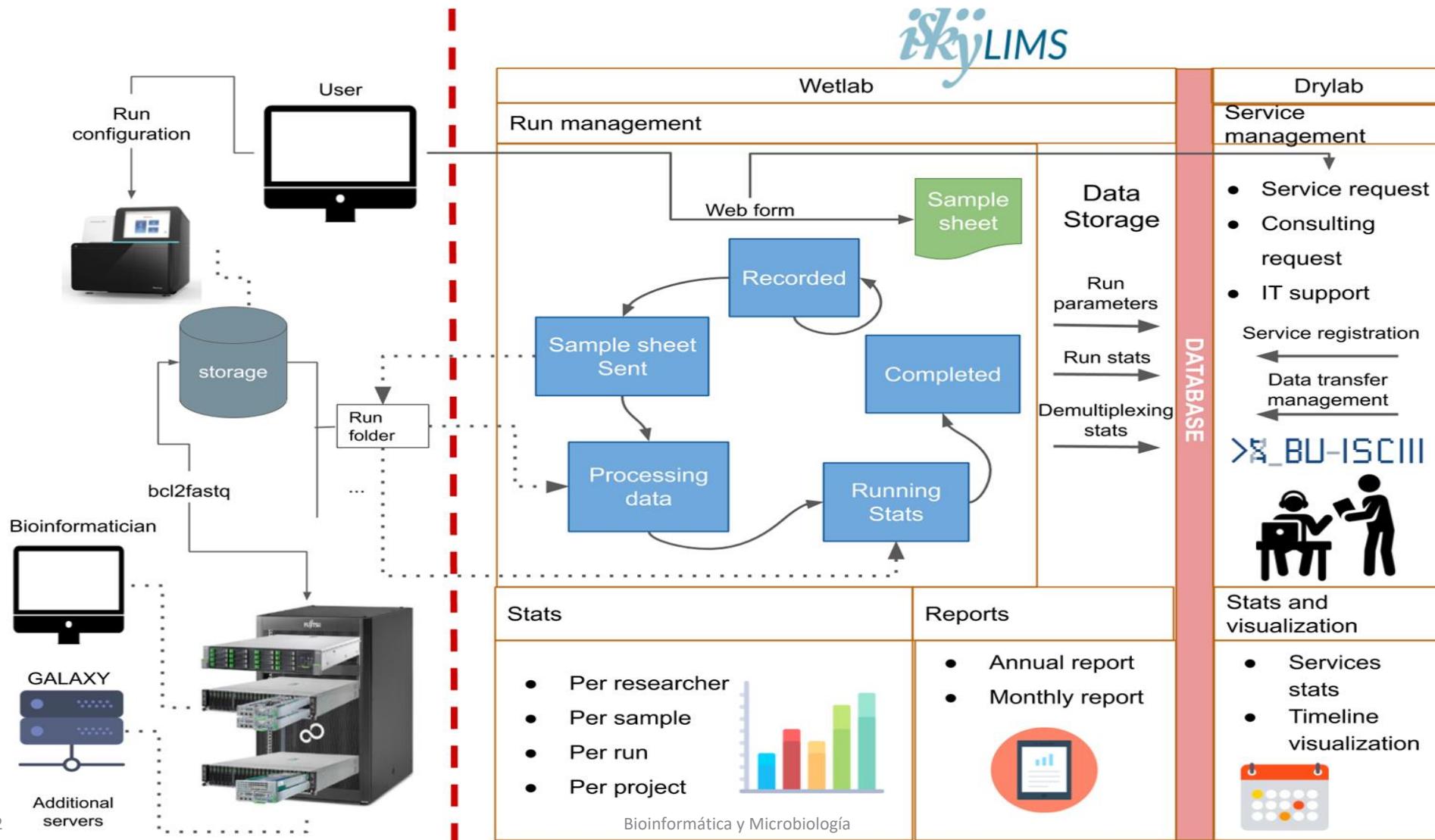
- INFRASTRUCTURE
- DATA WORKFLOW
- LIMS

Infrastructure and data management

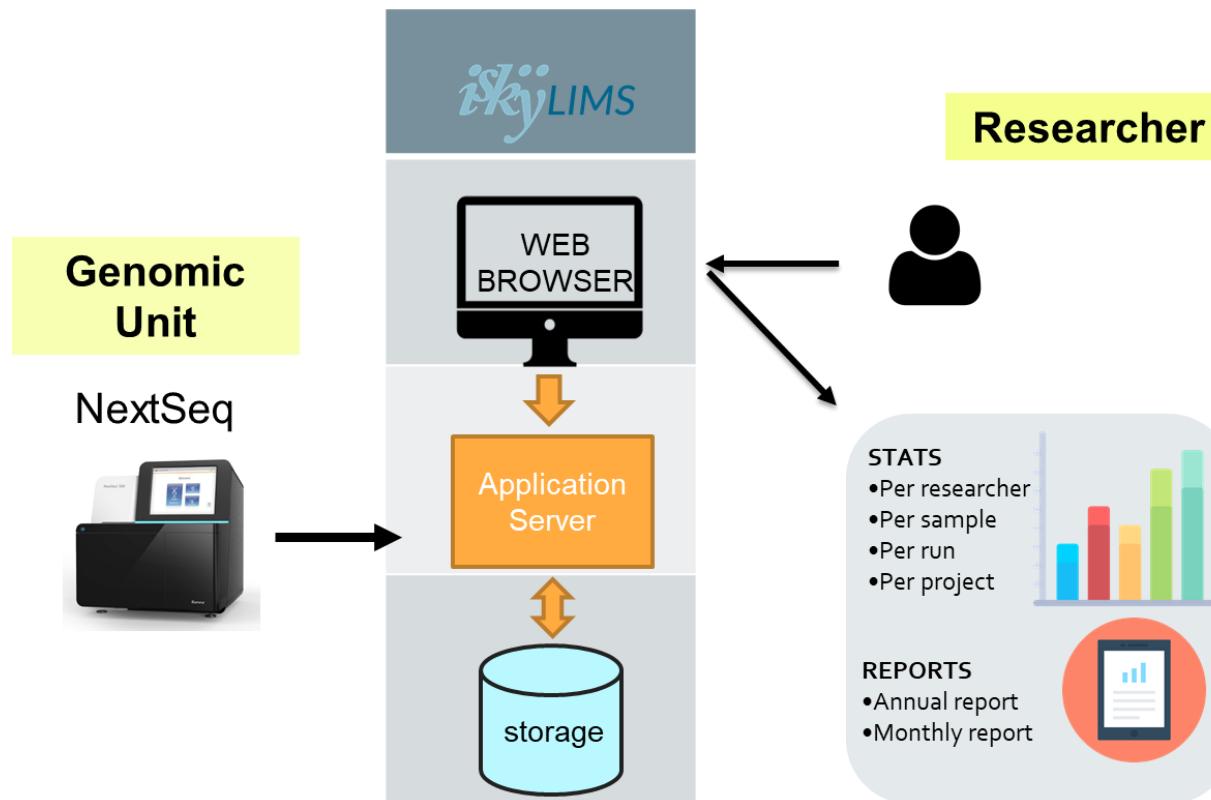


- Minimize I/O issues
- Maximize storage uses

Infrastructure and data management: LIMS



iSkyLIMS



- TEAM: Bioinformatician & IT
- SOURCE OF FUNDING:
 - Research Project
 - National or International Calls
 - Permanent position – Civil Servant

RRHH



COMPUTATIONAL RESOURCES & PIPELINES DEVELOPMENT



SERVICES PORTFOLIO & TRAINING



- COURSES, TFM, TFG
- DATA ANALYSIS
 - DNAseq
 - RNAseq
 - Metagenomics

- INFRASTRUCTURE
 - HPC
 - Workstations
- REPOSITORY:
 - GitHub
 - Benchmarking
 - Validation

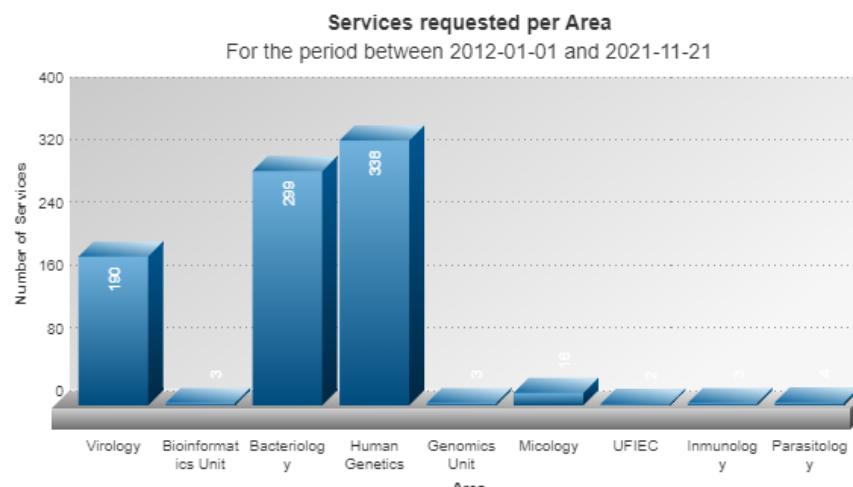
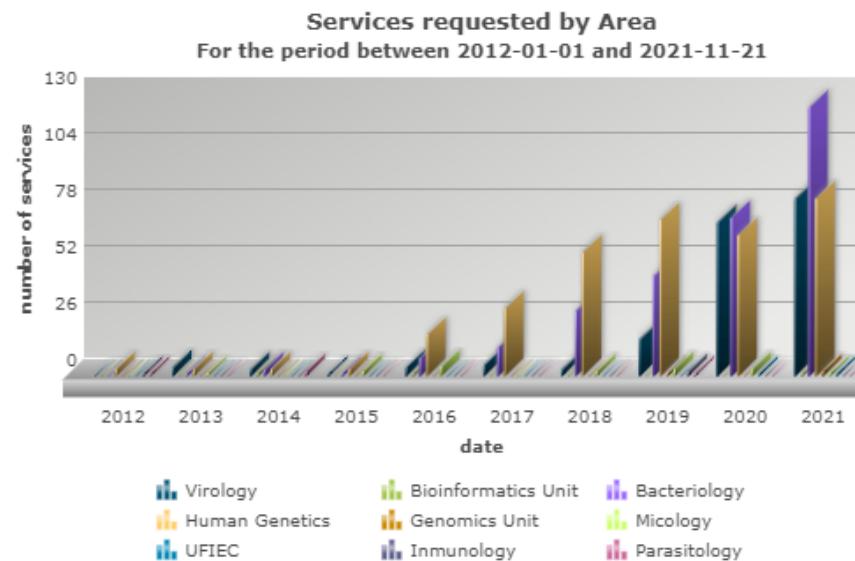
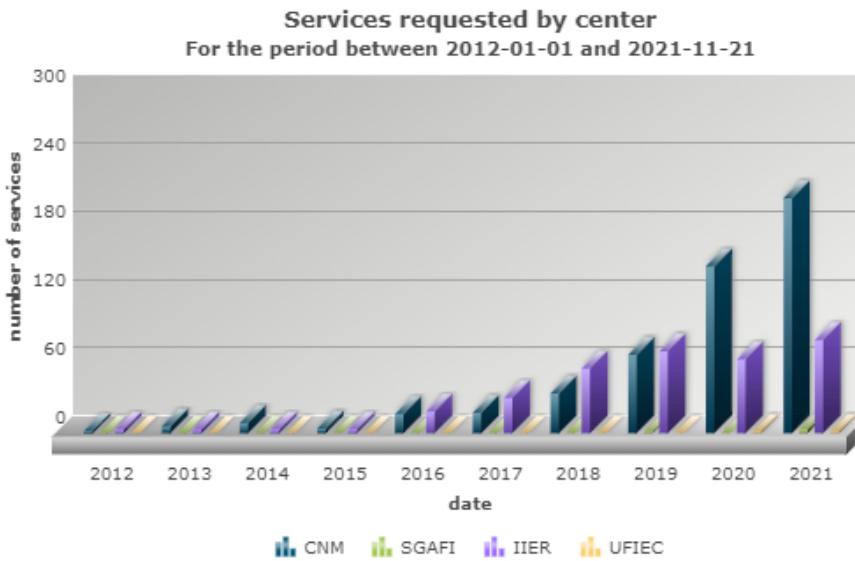
DATA MANAGEMENT



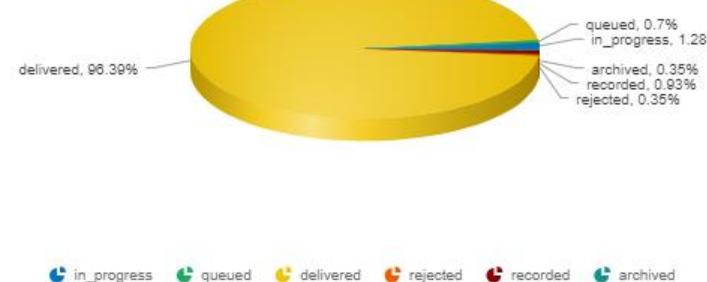
- INFRASTRUCTURE
- DATA WORKFLOW
- LIMS

Number of services: 2012 – 2021

>x_BU-ISCIII

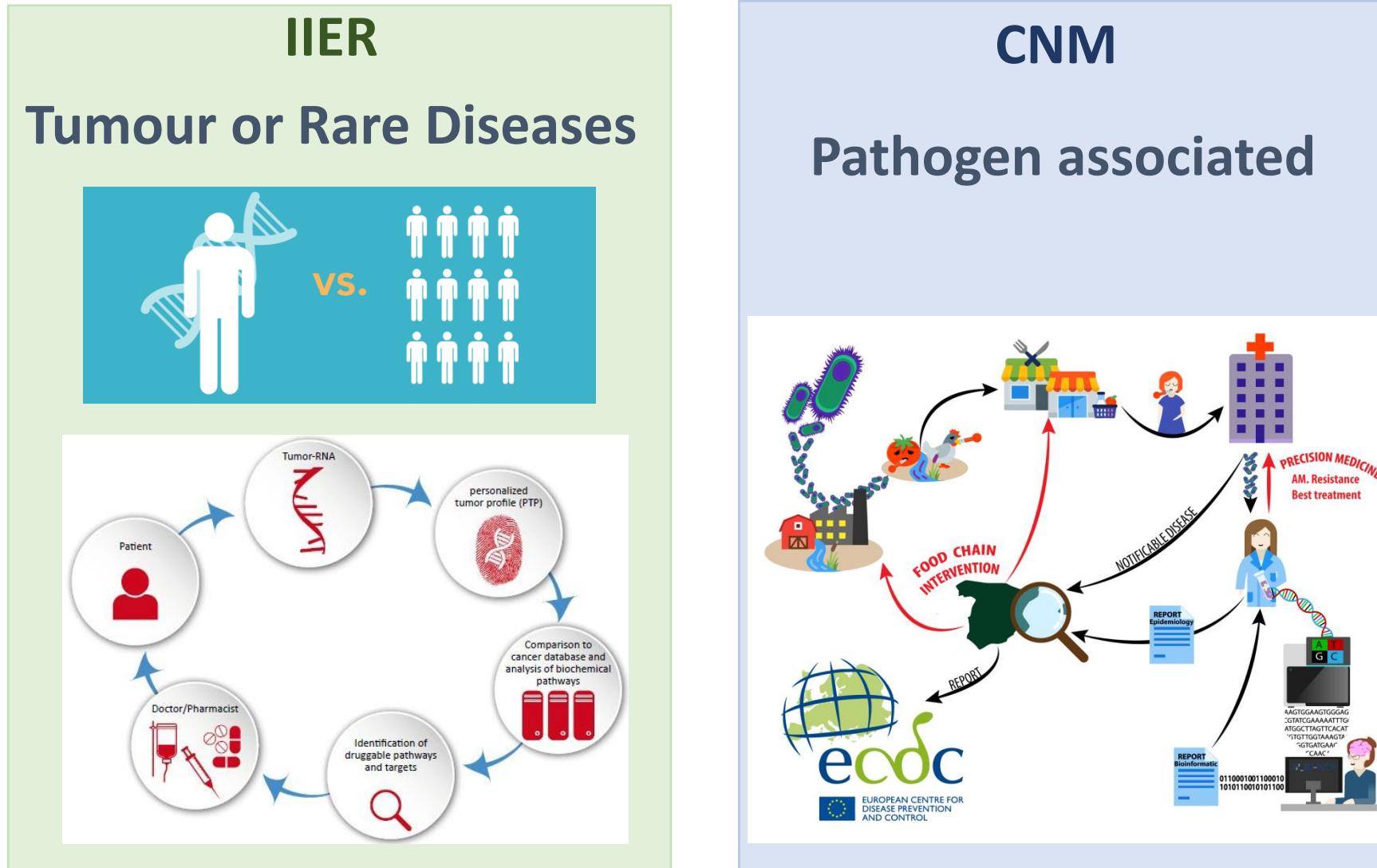


Status of Requested Services
For the period between 2012-01-01 and 2021-11-21

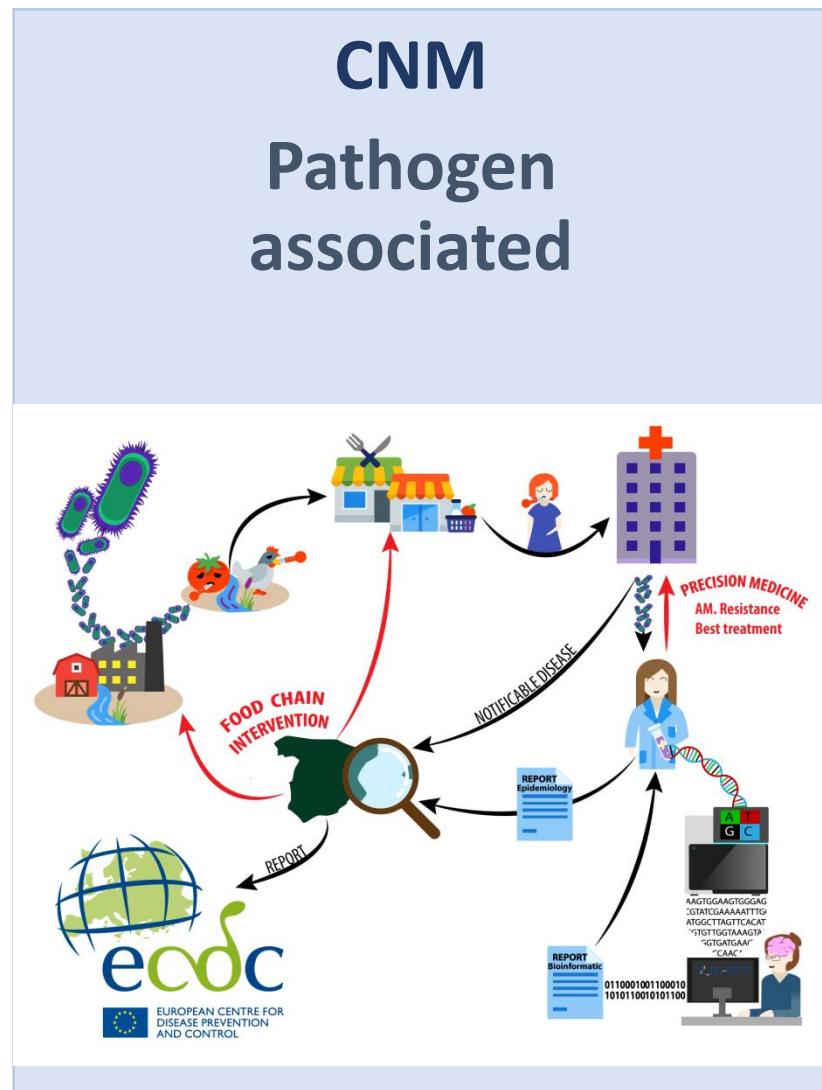


Services:
512 CNM – 341 IIER
– 3 UFIEC
57 Researchers

Clinical Bioinformatics - Precision Medicine



Research - Clinical Bioinformatics - Precision Medicine



AESI 2017-2019 BU-ISCIII – Genómica

AESI 2019-2021 BU-ISCIII - Genómica

AESI 2018 – 2021 PLATAFORMA DE
BIOINFORMATICA ISCIII-TransBioNet

METAGENOMICS EQAE

Special Pathogens Unit,
P. Anda, R. Escudero, I. Jado



GMI – HTS Standards, Databases
Data Sharing and Guidelines



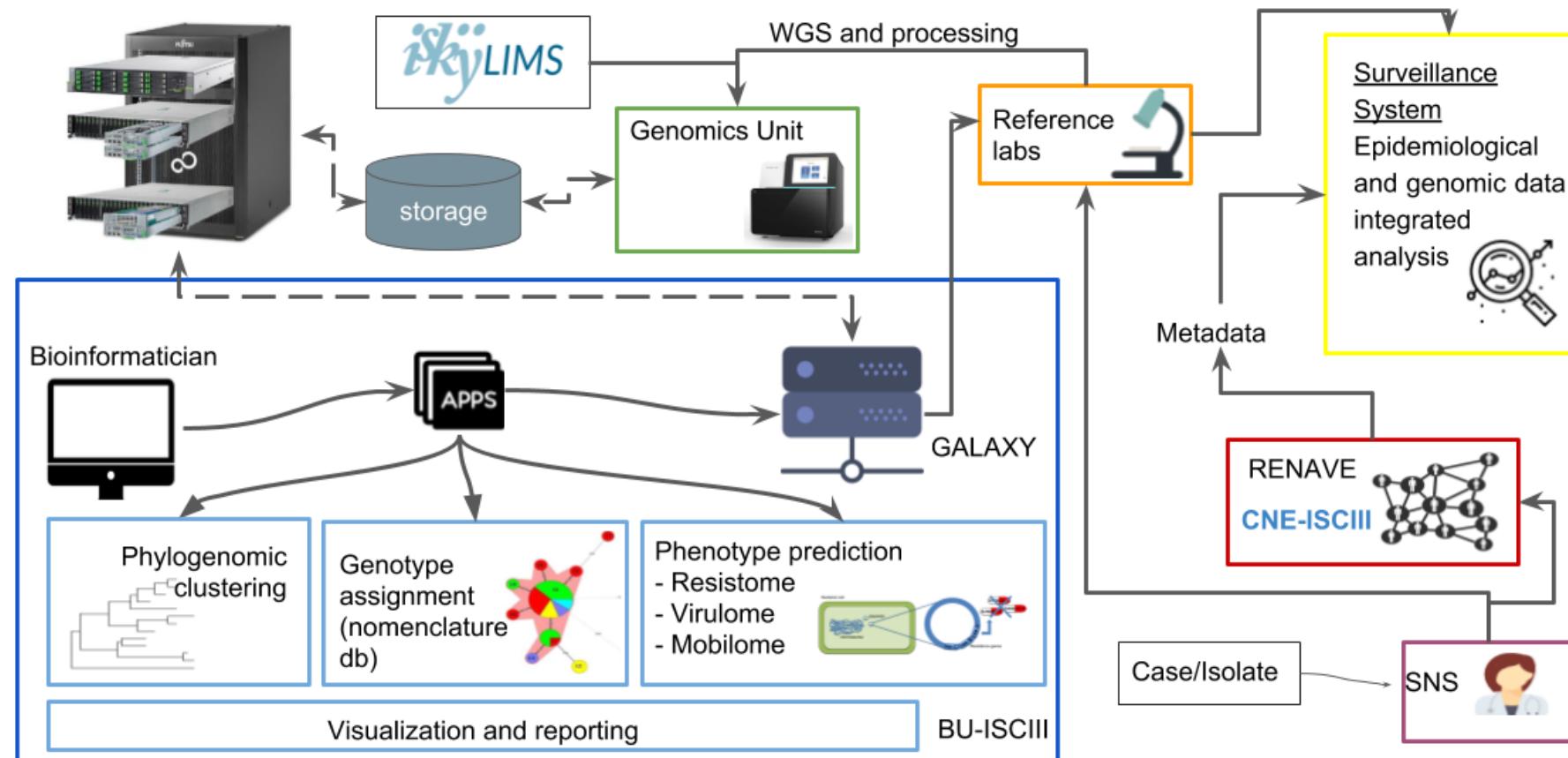
GMI – UNSGM PT for detection of biological threats by genomic analysis – **AESI 2019**

COMPARE Food Metagenomic Ring Trial



HTS Platform for Microbial Diagnostic and Surveillance based on HTS

Following ECDC roadmap for integration of molecular and genomic typing into European-level surveillance and epidemic preparedness



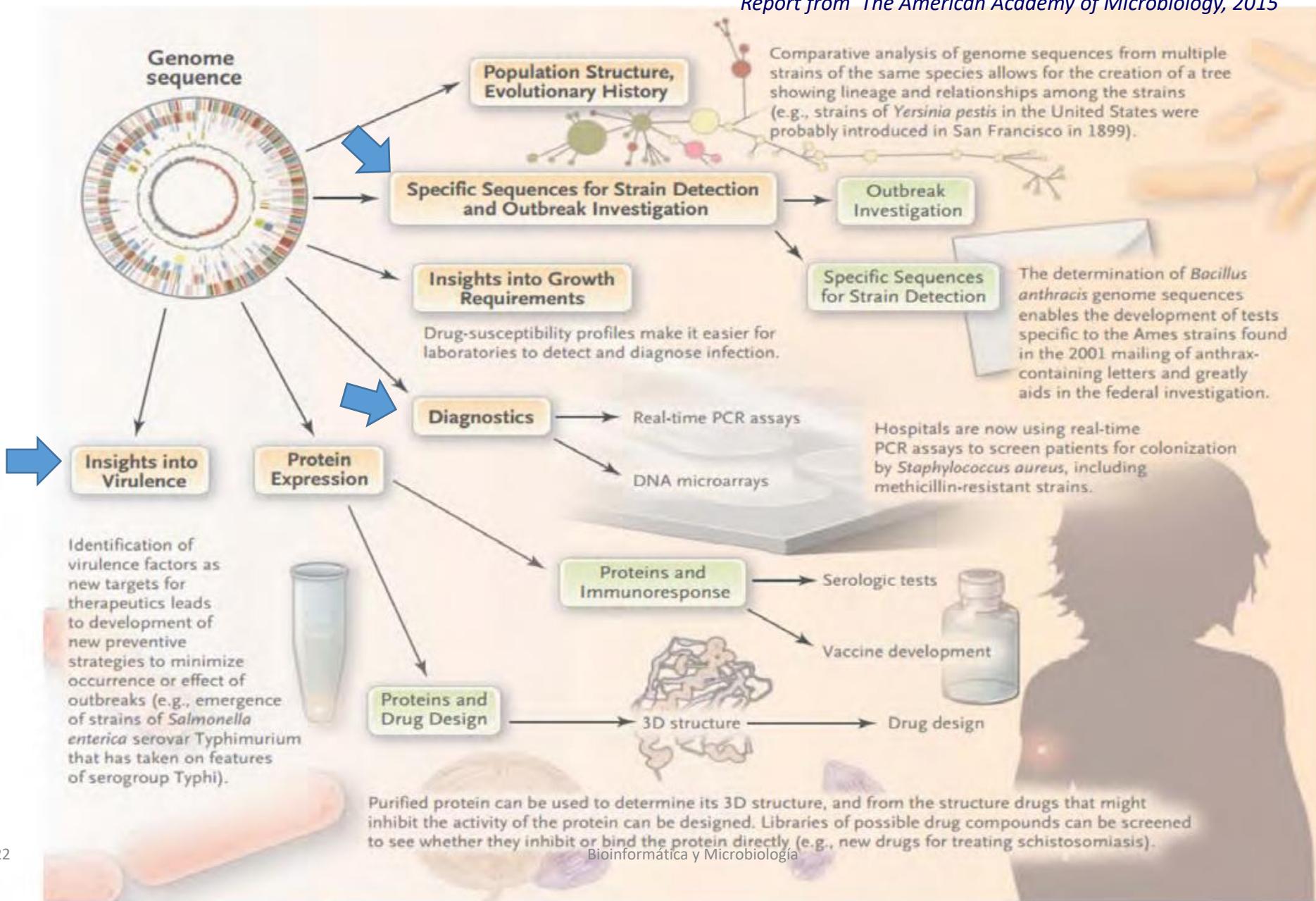
Index

- BU-ISCIII
- **High-throughput sequencing (HTS) applications in Microbiology**
 - Concepts: HTS and Outbreak investigation
 - Bacterial and Viral Genomics
 - Bioinformatics analysis in microbial genomics
 - Viralrecon: SARS-CoV-2 genome reconstruction software

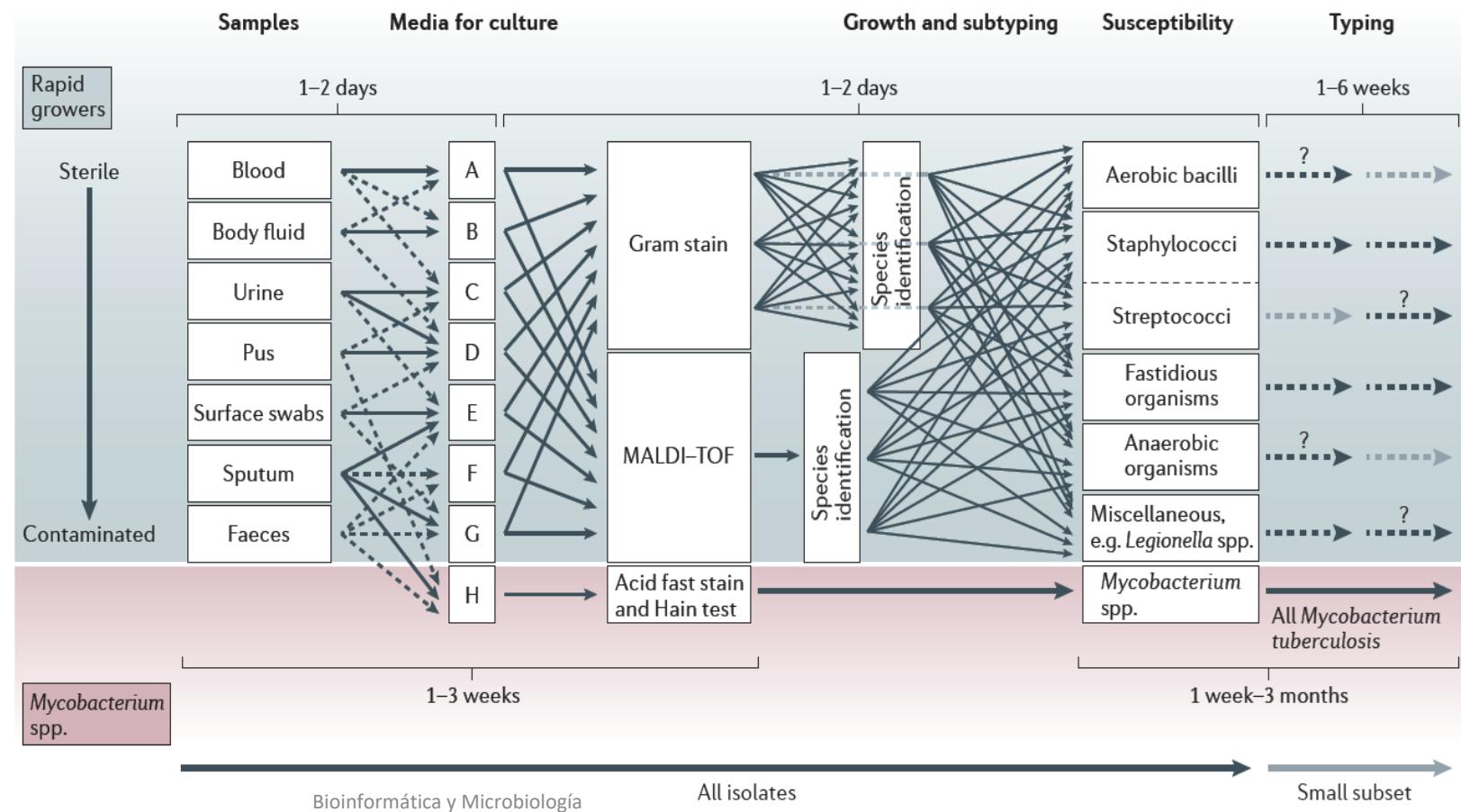
Microbial genomics

>X_BU-ISCIII

Report from The American Academy of Microbiology, 2015

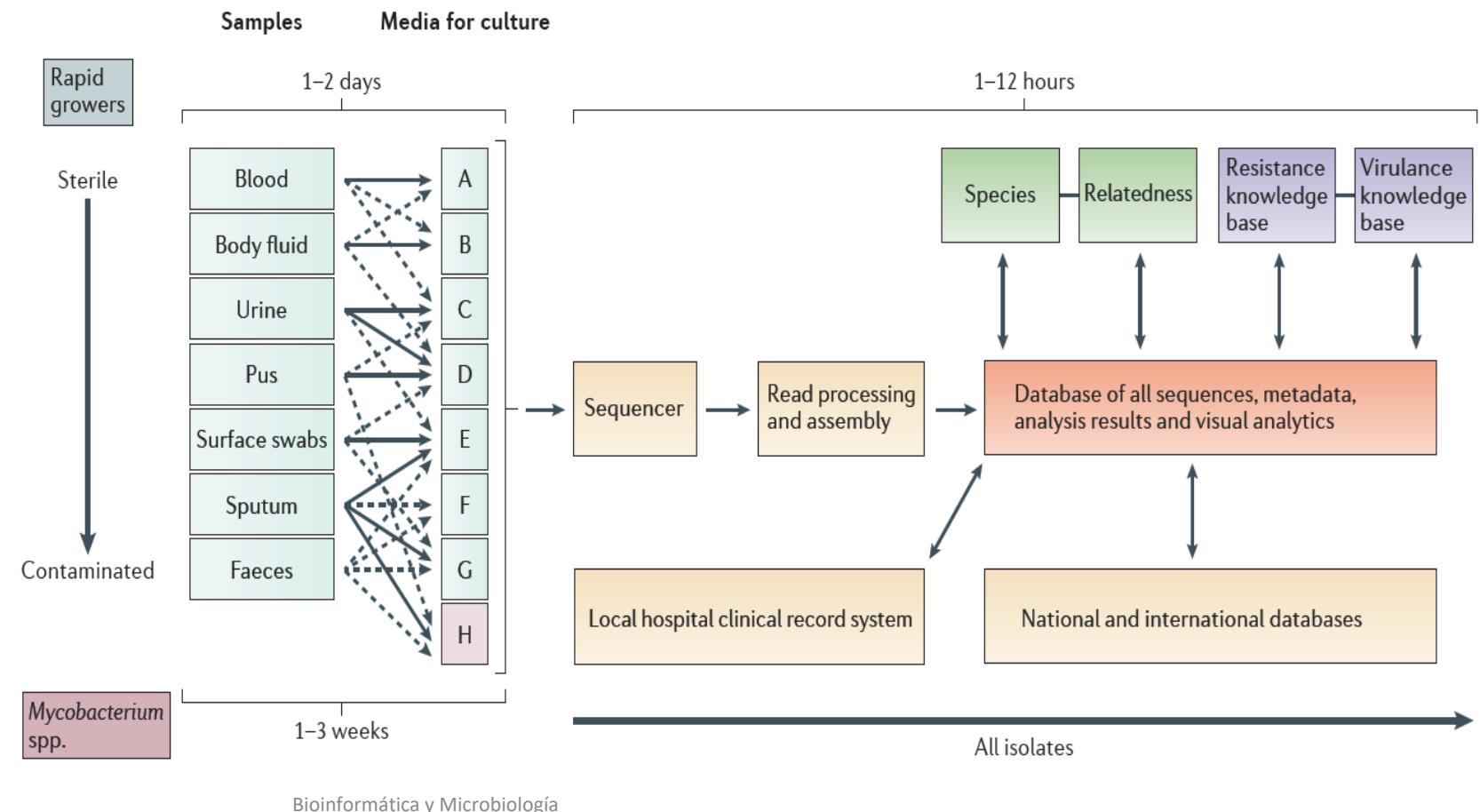


Classic techniques vs Whole Genome sequencing



ECDC roadmap and international commitment

Classic techniques vs Whole Genome sequencing



Foodborne outbreak identification “Crisis del pepino”

2011

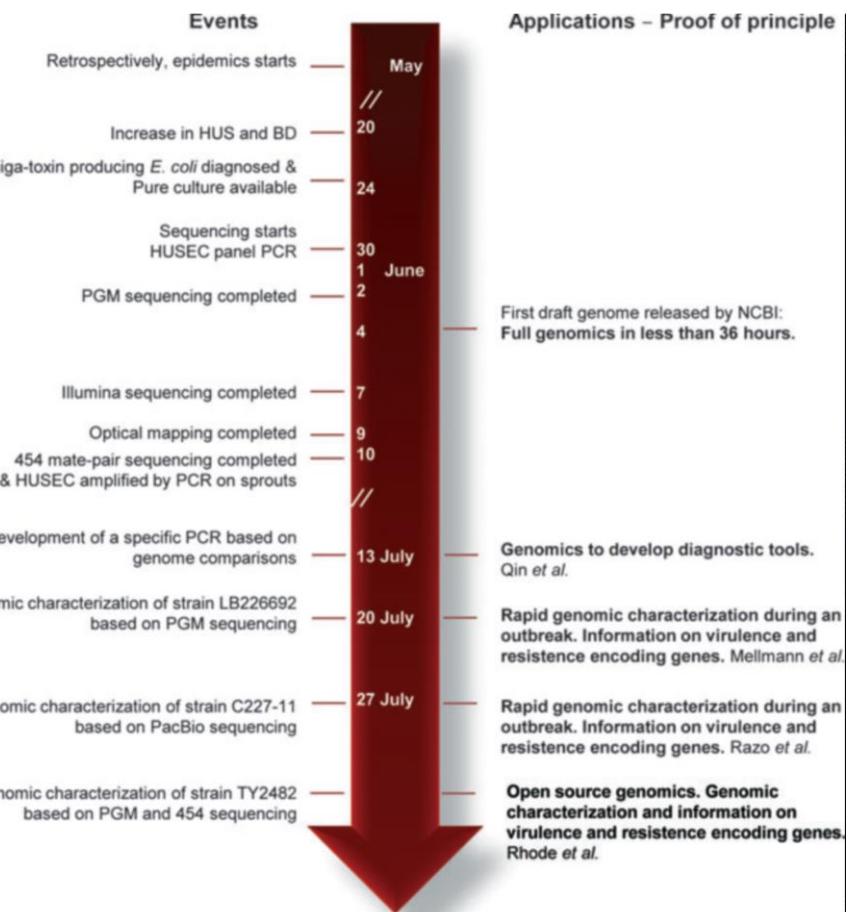
- Mayo 24 Primera muerte en Alemania
- 26 Alemania acusa a los pepinos españoles
- 30 Prohibición de importaciones de verduras de España y Alemania
- 31 Laboratorios alemanes desmienten oficialmente que los pepinos españoles sean el foco de infección
- Junio 10 Resolución de la crisis

Secuenciación Genoma



Bioinformática y Microbiología

The Escherichia coli O104:H4 epidemics: event timeline and major outputs



Bertelli and Greub, Clin Microb and Infect, 2013

Andalusian Listeria Outbreak

Actualización de información sobre el brote de intoxicación alimentaria causado por *Listeria monocytogenes*.

Publica: Agencia Española Seguridad alimentaria y Nutrición
Fecha: 29 agosto 2019
Sección: Seguridad Alimentaria

Jueves 29 de agosto de 2019, 12.00 horas

ACTUALIZACIÓN EN RELACIÓN CON LA DISTRIBUCIÓN DE PRODUCTOS RELACIONADOS CON LA ALERTA.

La Agencia Española de Seguridad Alimentaria y Nutrición (AESAN) recomienda a las personas que tengan en su domicilio algún producto de la marca "La Mechá" se abstengan de consumirlo. Si se dispone del producto se debe devolver al punto de compra y, de no ser posible, desecharlo.

Brote de listeriosis: sube el número de afectados y se apunta a la falta de higiene en la carne como causa

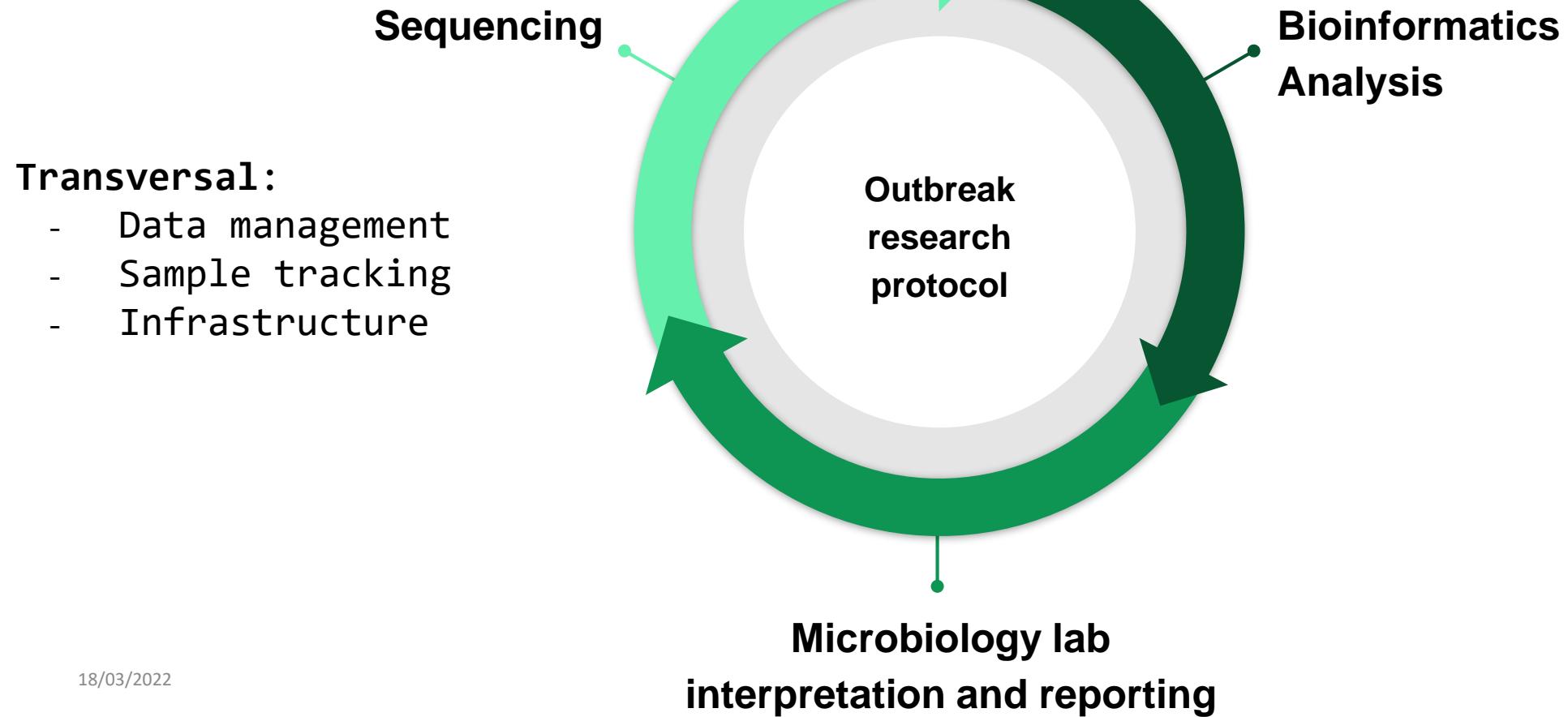
EFE 25.08.2019

- Tres nuevos casos, en Sevilla y Cádiz, dejan el número de personas afectadas en Andalucía en 192.
- [La carne con listeria de la marca blanca se vendió en los municipios de Sevilla.](#)
- La empresa que vendió la marca blanca de Magrudis dice que cumple los protocolos.



- Meat “La Mechá”. Margulis S.L.
- 250 cases related.
- Meat “"La Montanera del Sur". INCARYBE S.L”, suspicion. (Cádiz)
- Meat “Sabores de Paterna” (Málaga)

Andalusian Listeria Outbreak

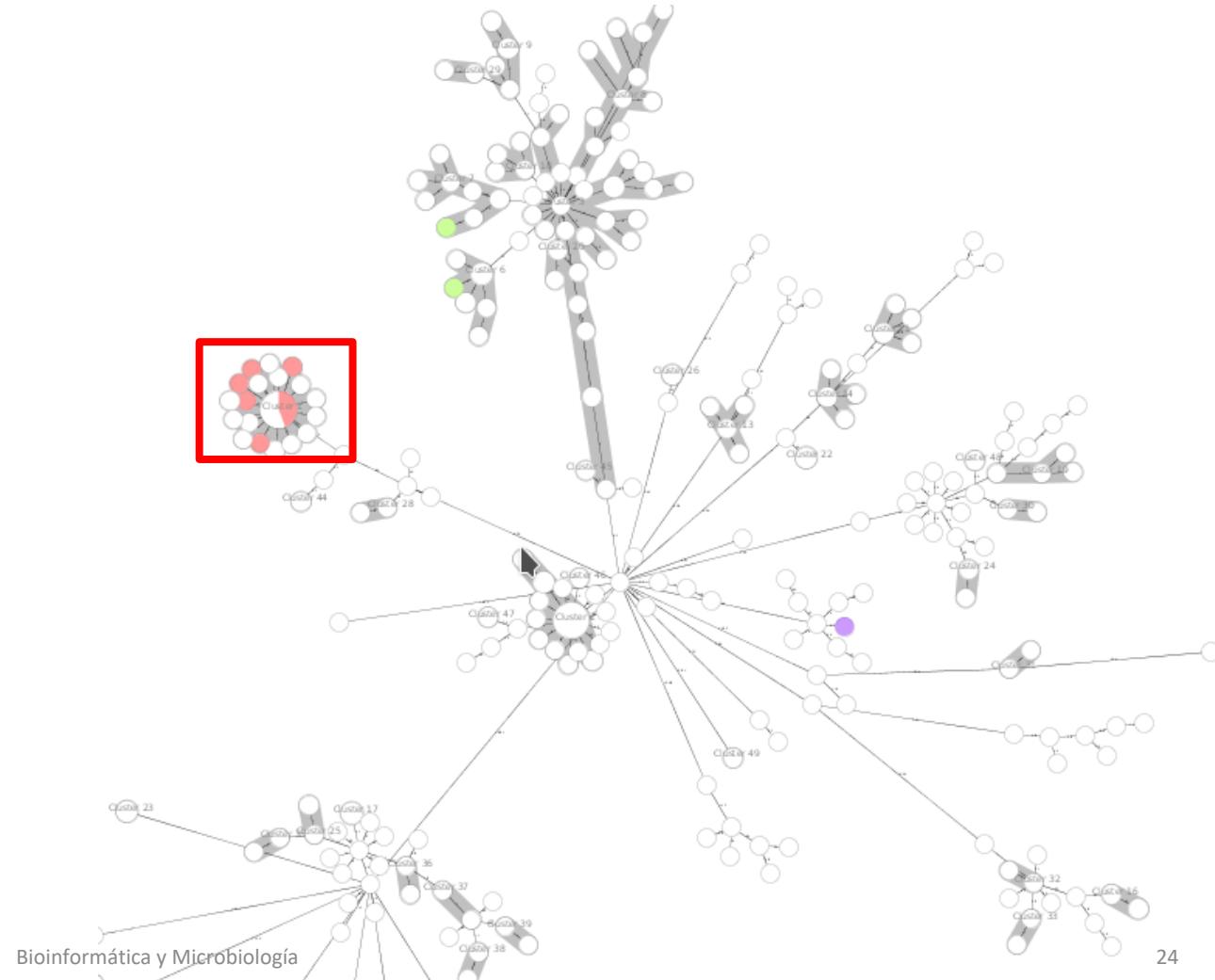


Andalusian Listeria Outbreak

- 625 listeria samples already sequenced
 - 258 suspected to be related to the outbreak (mid august to mid september)

Results:

- 233 related to the outbreak, confirmed to be caused by the meat “La Mechá”
 - 25 sporadic cases not related to the outbreak.



Pathogen discovery: new virus – SARS-CoV-2

Deep Meta-Transcriptomic Sequencing



bronchoalveolar lavage fluid (BALF)



Meta-transcriptomic library

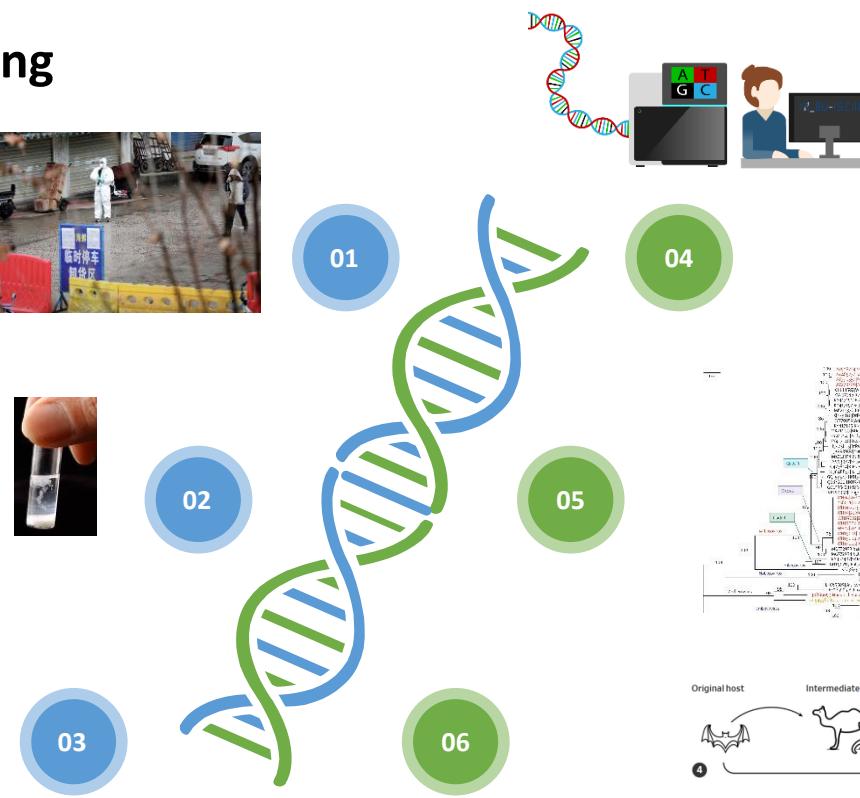
2x150 MiniSeq

56,565,928 sequences reads

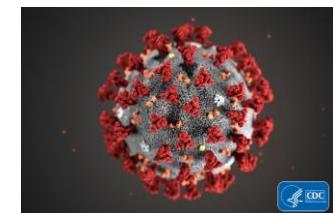


03

02



Wu et al., Nature 2020

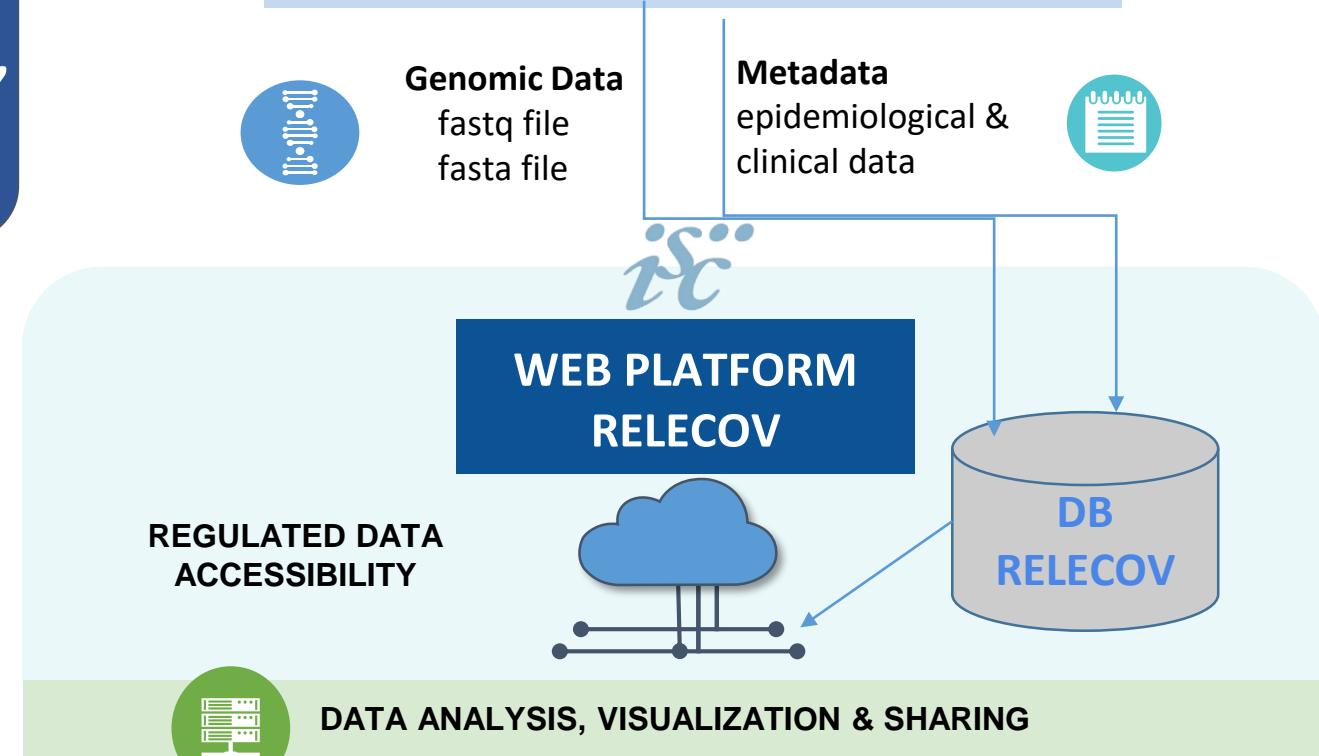


De novo-assembled - Megahit
384,096 Contigs
Screened for potential aetiological agents
The longest 30,474 nt

89.1% identity
Closely related to a bat SARS-like coronavirus

SARS-CoV-2 GENOMICS SPAIN, RELECOV

REFERENCES LABORATORY NETWORK AUTONOMOUS COMMUNITIES



Support and Standards definition

- Needs and capacities identification.
 - Sequencing protocols and bioinformatics pipelines analysis
 - Support and training to the labs network.
 - Documents and guidelines definition
- 18/03/2022

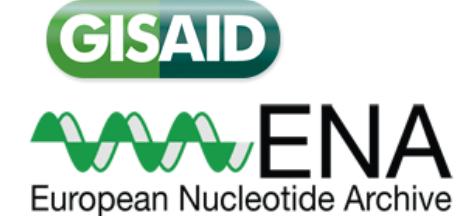
Bioinformatics analysis

- QC metrics
- Consensus genome
- Variants and Lineage report
- Viral Quasispecies
- Phylogenetic and phylodynamic analysis
- Stats and visualization results

National-wide web surveillance System

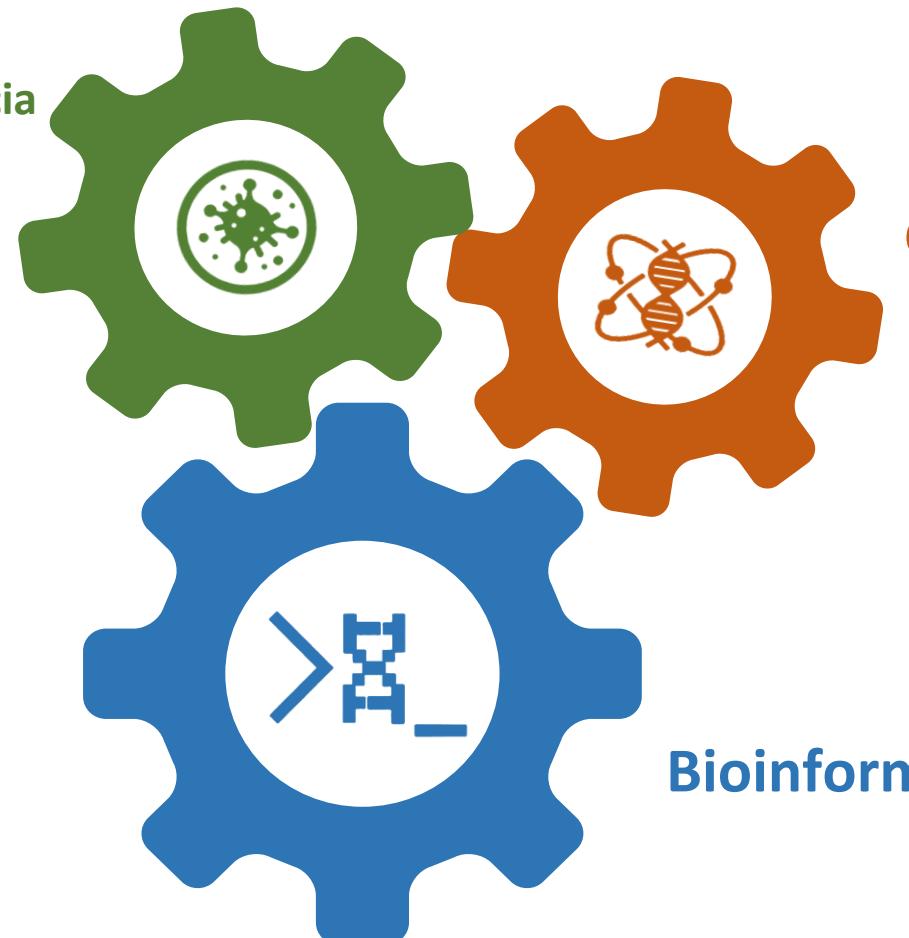
- Detection and variant surveillance.
- **Nextstrain Spain**
- **outbreak.info**
- **COG-UK Consortium**

International sharing



M2: Organization of circuits between hospitals and regional laboratories

Microbiología
Laboratorios de Referencia



Genómica

Bioinformática

Bioinformática y Microbiología

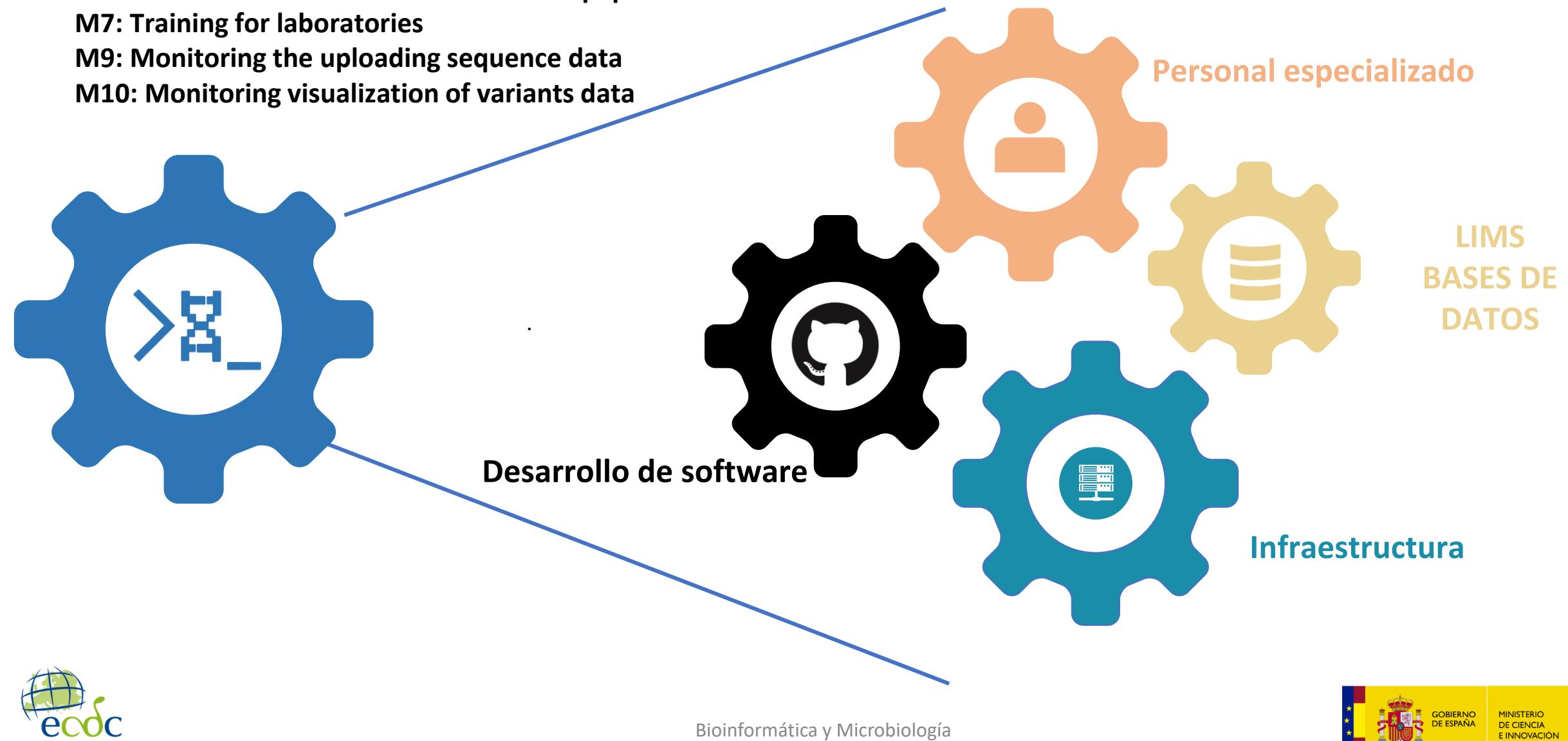
Milestones

M4: Installation and verification of new equipment

M7: Training for laboratories

M9: Monitoring the uploading sequence data

M10: Monitoring visualization of variants data



Milestones



Bioinformática y Microbiología



Incorporar la Genomica



Dotar de recursos



Personal especializado



Formación

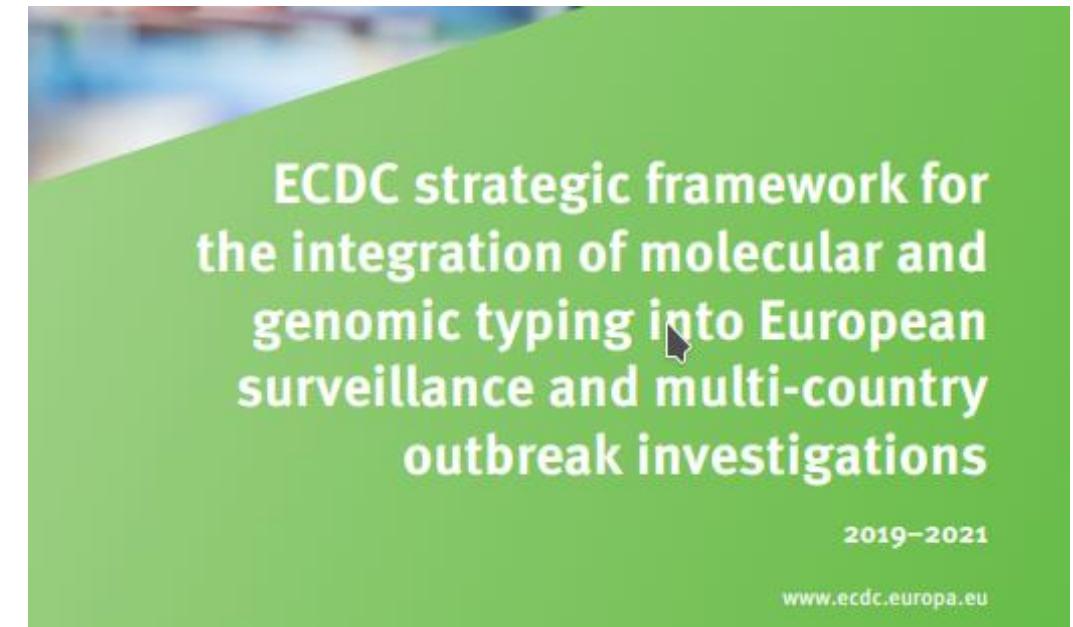
One Health approach, infectious diseases could be better controlled and prevented



ECDC roadmap and international commitment



EUROPEAN CENTRE FOR
DISEASE PREVENTION
AND CONTROL



- **Operationalisation of EU-wide WGS-based surveillance systems in the near term:** start implementation of WGS-based surveillance for *Listeria monocytogenes*, *Neisseria meningitidis*, Carbapenemase-producing *Enterobacteriaceae* and antibiotic-resistant *Neisseria gonorrhoeae*; 2018

Spanish National Microbiology Center (CNM)



Mission: Provide support to the National Health System and the different Spanish Regions in the diagnosis and control of infectious diseases. In order to fulfill this mission it acts as Reference center offering a series of scientific activities:

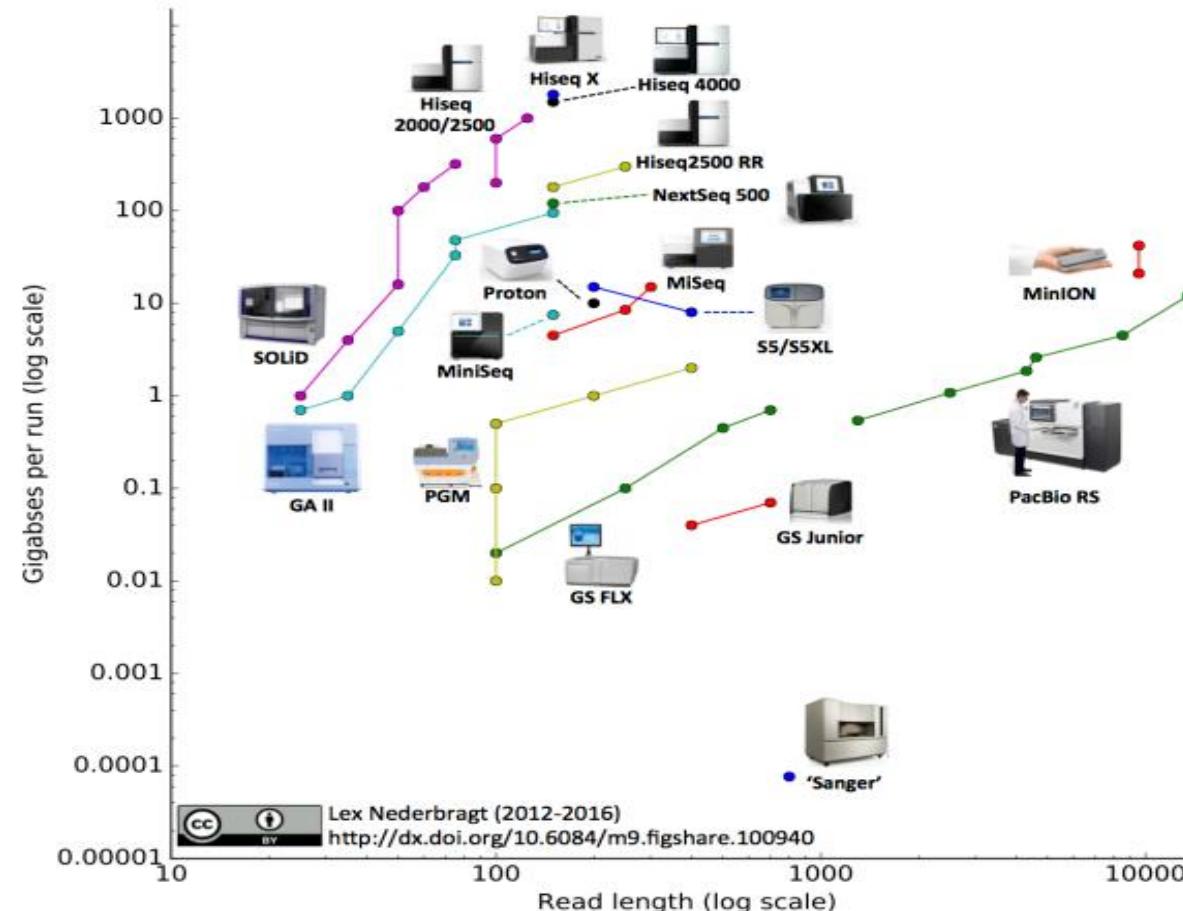
- Diagnosis
- **Surveillance** →
- Infectious diseases research
- Training

Outbreak research:
Molecular source
detection

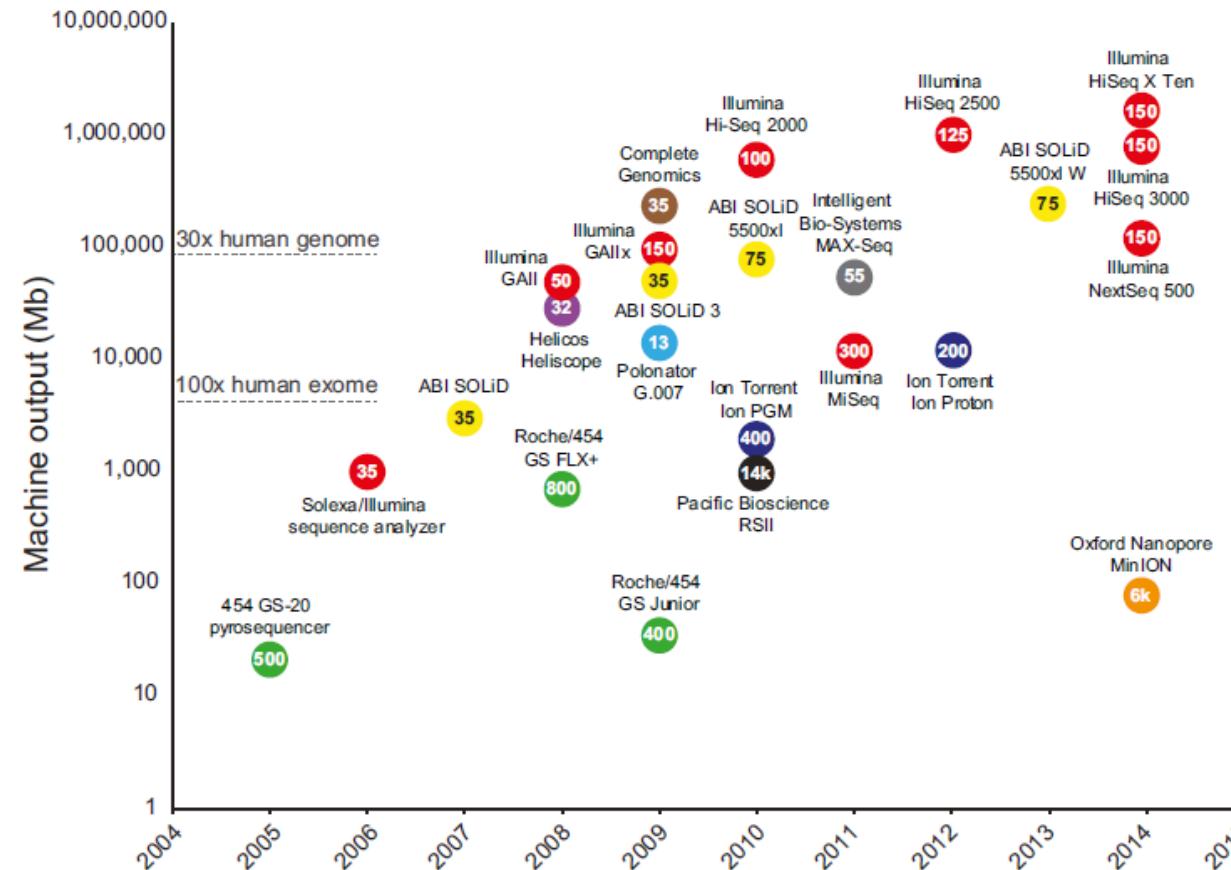
Index

- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology
- **Concepts: HTS and Outbreak investigation**
- Bacterial and Viral Genomics
- Bioinformatics analysis in microbial genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

High-Throughput Sequencing Technologies



High-Throughput Sequencing Technologies



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

Reuter et al., Mol Cell 2015

PREPARACIÓN LIBRERÍA, estrategias

SECUENCIACIÓN GENOMA, EXOMA, TRANSCRIPTOMA

1. Sin amplificación
2. Amplificación con PCR
3. Sondas captura

- Tamaño de fragmento
- Longitud de la lectura
- Single o Paired-end
- Número de bases por muestra
- Profundidad de cobertura x

SECUENCIACIÓN GENOMAS

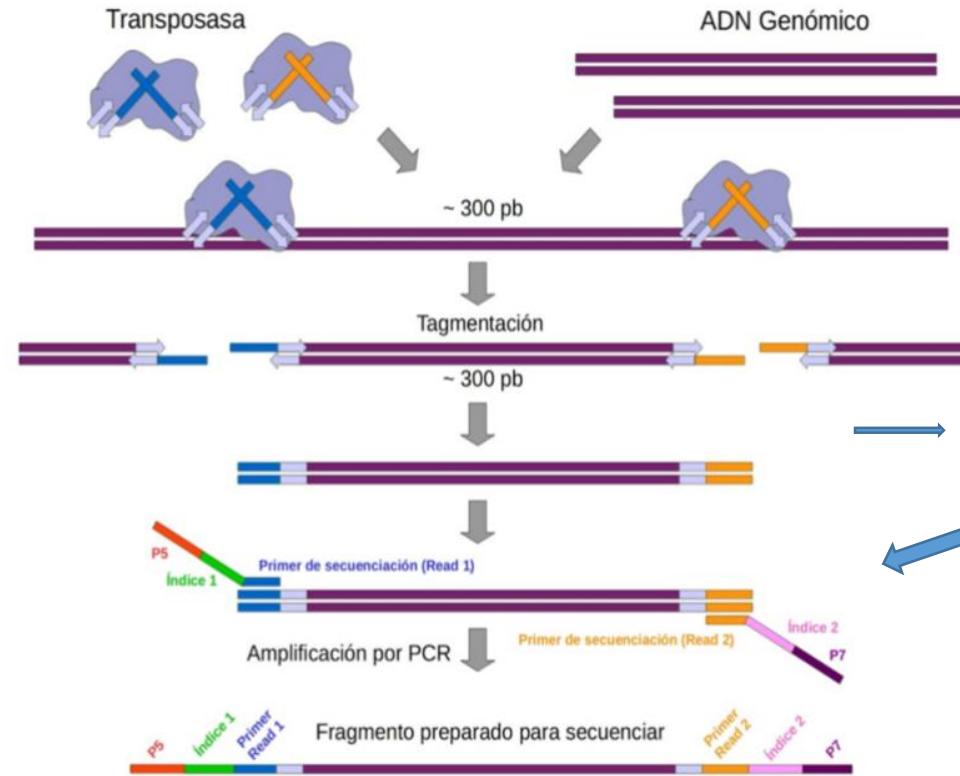
1. Metagenómica

IDENTIFICACIÓN MICROORGANISMOS

1. Metataxonomía

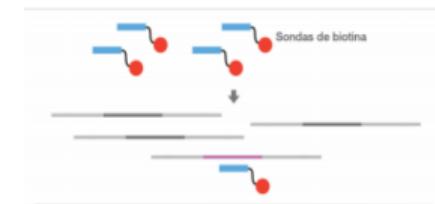
PREPARACIÓN LIBRERÍA

ENZIMÁTICA FÍSICA



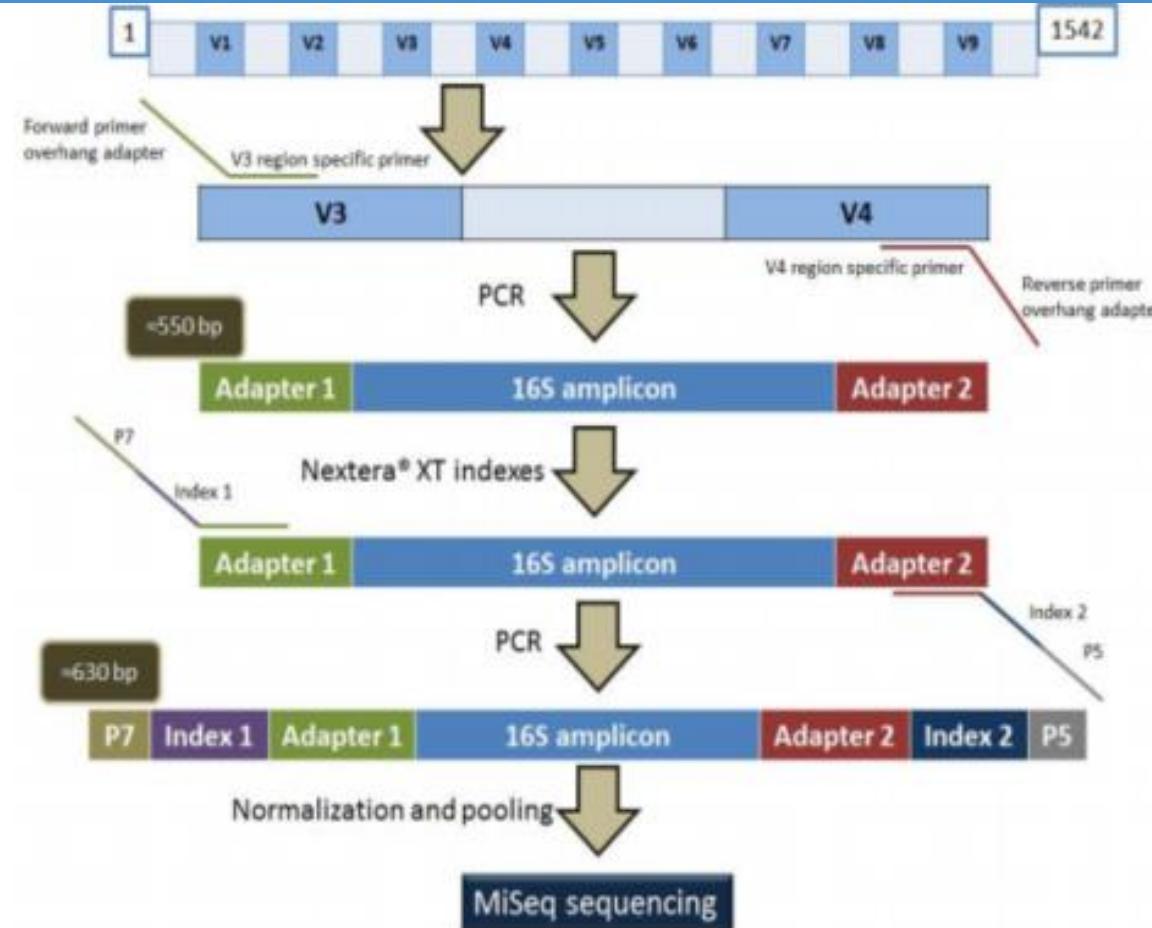
RNA -> cDNA

**ENRIQUECIMIENTO:
PCR
CAPTURA SONDAS**



Guia Práctica Genómica https://www.uv.es/varnau/GM_Cap%C3%ADtulo_2.pdf

PREPARACIÓN LIBRERÍA, rRNA 16S, caracterización microbiota



Sequencing concepts

Depth of coverage

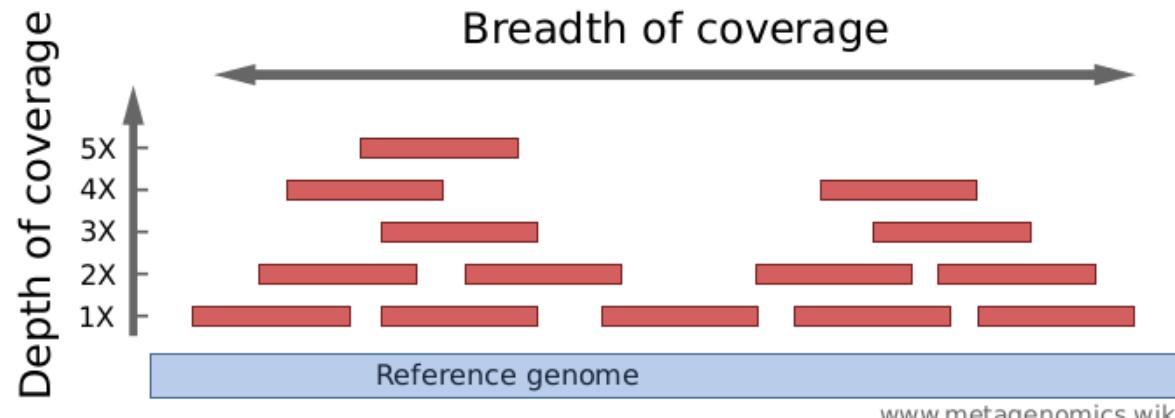
How strong is a genome "covered" by sequenced fragments (short reads)?

Per-base coverage is the average number of times a base of a genome is sequenced. The coverage depth of a genome is calculated as the number of bases of all short reads that match a genome divided by the length of this genome. It is often expressed as 1X, 2X, 3X,... (1, 2, or, 3 times coverage).

Breadth of coverage

How much of a genome is "covered" by short reads? Are there regions that are not covered, even not by a single read?

Breadth of coverage is the percentage of bases of a reference genome that are covered with a certain depth. For example: 90% of a genome is covered at 1X depth; and still 70% is covered at 5X depth.



Index

- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology. Examples
- Concepts: HTS and **Outbreak investigation**
- Bacterial and Viral Genomics
- Bioinformatics analysis in microbial genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

Outbreak definition and Typing methods: DNA-based methods

A disease **OUTBREAK** is the occurrence of disease cases in excess of normal expectancy.

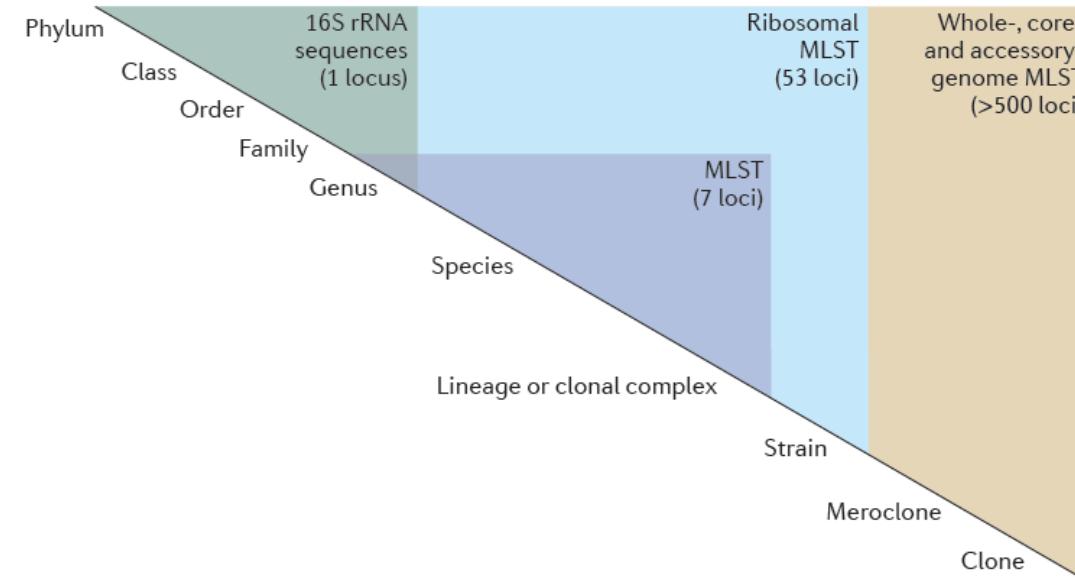
Bacterial identification and characterization at subspecies level is commonly known as **Microbial Typing**. Currently, these methodologies are fundamental tools in Clinical Microbiology and bacterial population genetics studies to track outbreaks and to study the dissemination and evolution of virulence or pathogenicity factors and antimicrobial resistance

Several typing methods have been used in outbreak detection and epidemiological surveillance ranging from **phenotypic methods to fragment based methods and sequence based methods**.

WHAT IS MOLECULAR TYPING?

Molecular typing is a way of identifying specific strains of microorganisms, such as bacteria or viruses, by looking at their genetic material. It is mainly used in outbreak investigation as pinpoint the **source of foodborne outbreaks**. It can also be used to identify which microorganisms are: Most virulent and cause serious diseases, resistant to antibiotics, or able to survive and multiply.

Sequence data for taxonomy and typing



Different levels of sequence information can be associated with different taxonomic levels.

The need for higher-resolution characterization of isolates has led to the development of a wide range of strain-typing methods

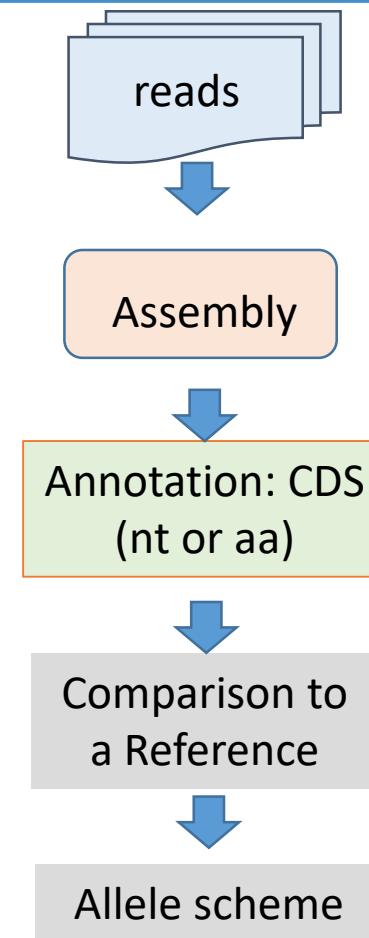
Concepts

Core genome: the number of shared features in a pool of genomes. Shared genes among multiple strains are mostly related to house-keeping genes or central metabolic processes, most of the structural information and main genotypic features. **Orthologues** (sequences have common ancestor and have split due to speciation event) in all genomes of bacteria belonging to the same taxa

Accessory genome or adaptative genome: includes genes conferring adaptive advantages to the strain in order to survive in a specific environment. In most cases, these factors are linked to antibiotic resistance, virulence, capsular serotype, adaptation, and might reflect the organisms predominant lifestyle.

Pangenome: The term “pan-genome” refers to pan (from Greek πᾶν, whole) and genome (genome) referring to the inclusion of the core and the dispensable genome.

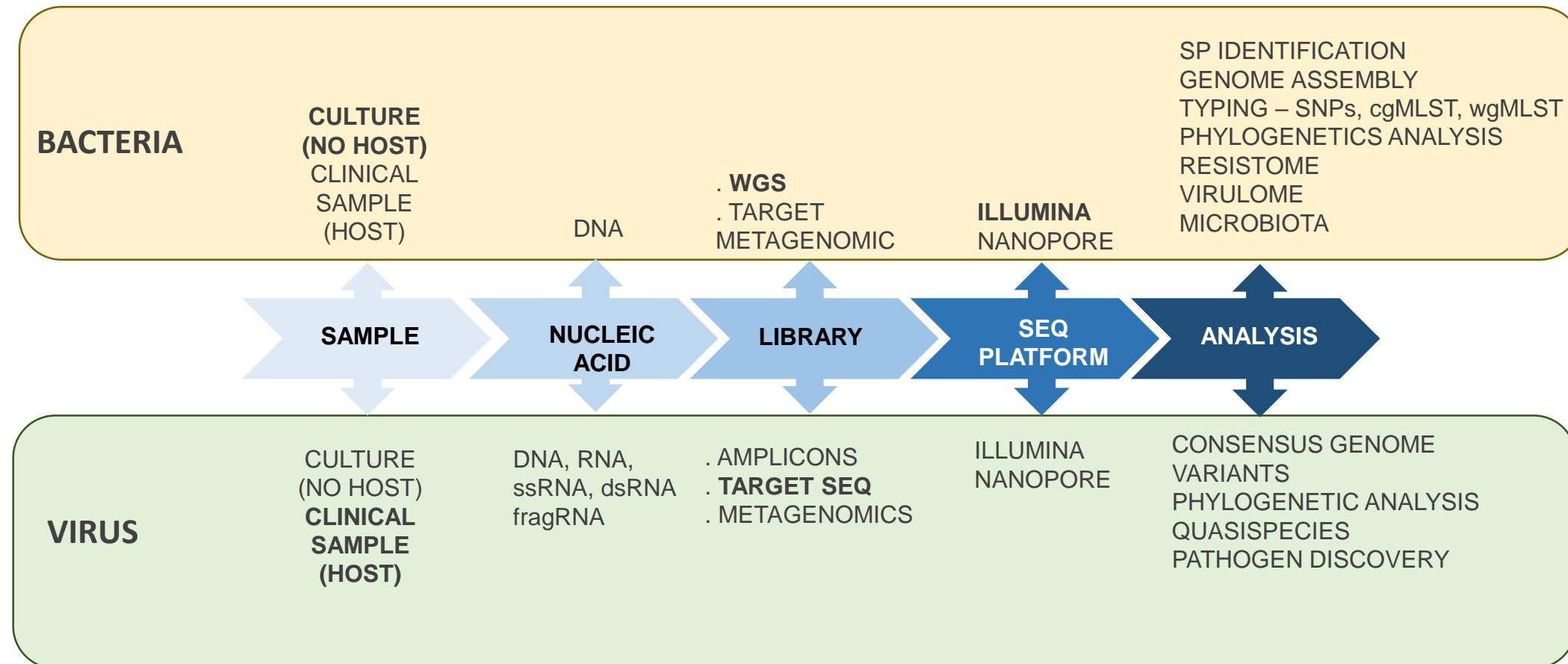
General analytical process for cgMLST / wgMLST



Index

- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology. Examples
- Concepts: HTS and Outbreak investigation
- **Bacterial and Viral Genome Sequencing**
- Bioinformatics analysis in microbial genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

Bacterial and viral Genome Sequencing



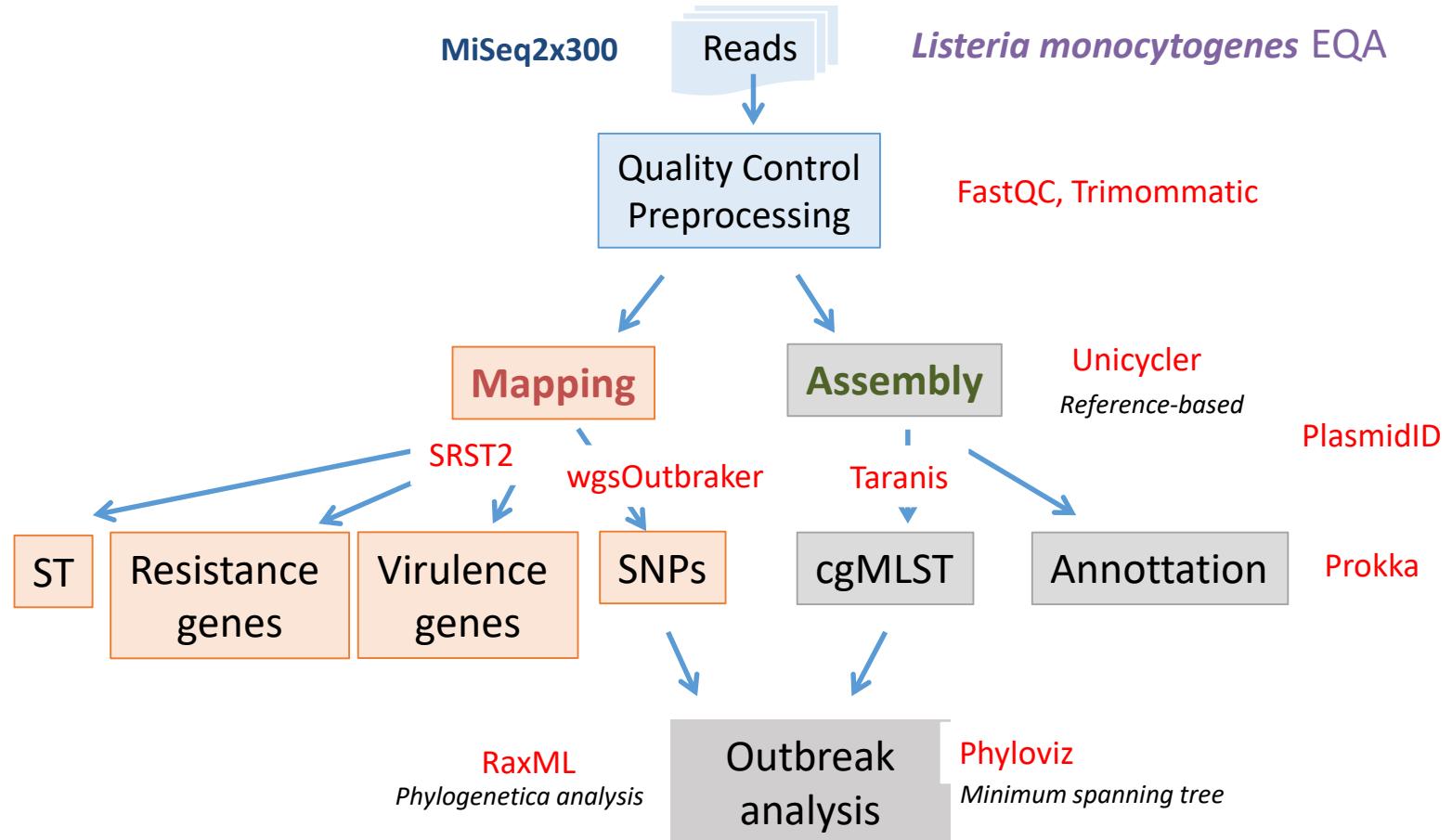
Index

- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology. Examples
- Concepts: HTS and Outbreak investigation
- Bacterial and Viral Genome Sequencing
- **Bioinformatics analysis in microbial genomics**
- Viralrecon: SARS-CoV-2 genome reconstruction software

Bioinformatics analysis in microbial genomics

- SPECIE IDENTIFICATION
 - WGS - Kmers analysis
 - TARGET METAGENOMIC, rRNA - MICROBIOTA
- ASSEMBLY GENOME
 - de NOVO or REFERENCE -BASED
 - cgMLST, wgMLST - MINIMUM SPANING TREE
 - METAGENOMIC - HOMOLOGY -BASED
- VARIANT CALLING
 - REFERENCE GENOME SELECTION
 - HAPLOID GENOME
 - LOW FREQUENCY VARIANT - QUASISPECIES
 - SNPs MATRIX - PHYLOGENETIC ANALYSIS
- STRUCTURAL AND FUNCTIONAL ANNOTATION
 - RESISTOME, VIRULOME, SEQUENCE-TYPE

Workflow example



Software disponible - VARIANT CALLING

- CFSAN SNP Pipeline

Extracción de SNPs de alta calidad de aislados relacionados

<http://snppipeline.readthedocs.io/en/latest/>

- GATK, modo haploide
- Samtools
- VarScan
- Snippy

Identificación de variantes haploides y construcción de filogenia usando core genome SNPs

<http://github.com/tseemann/snippy>

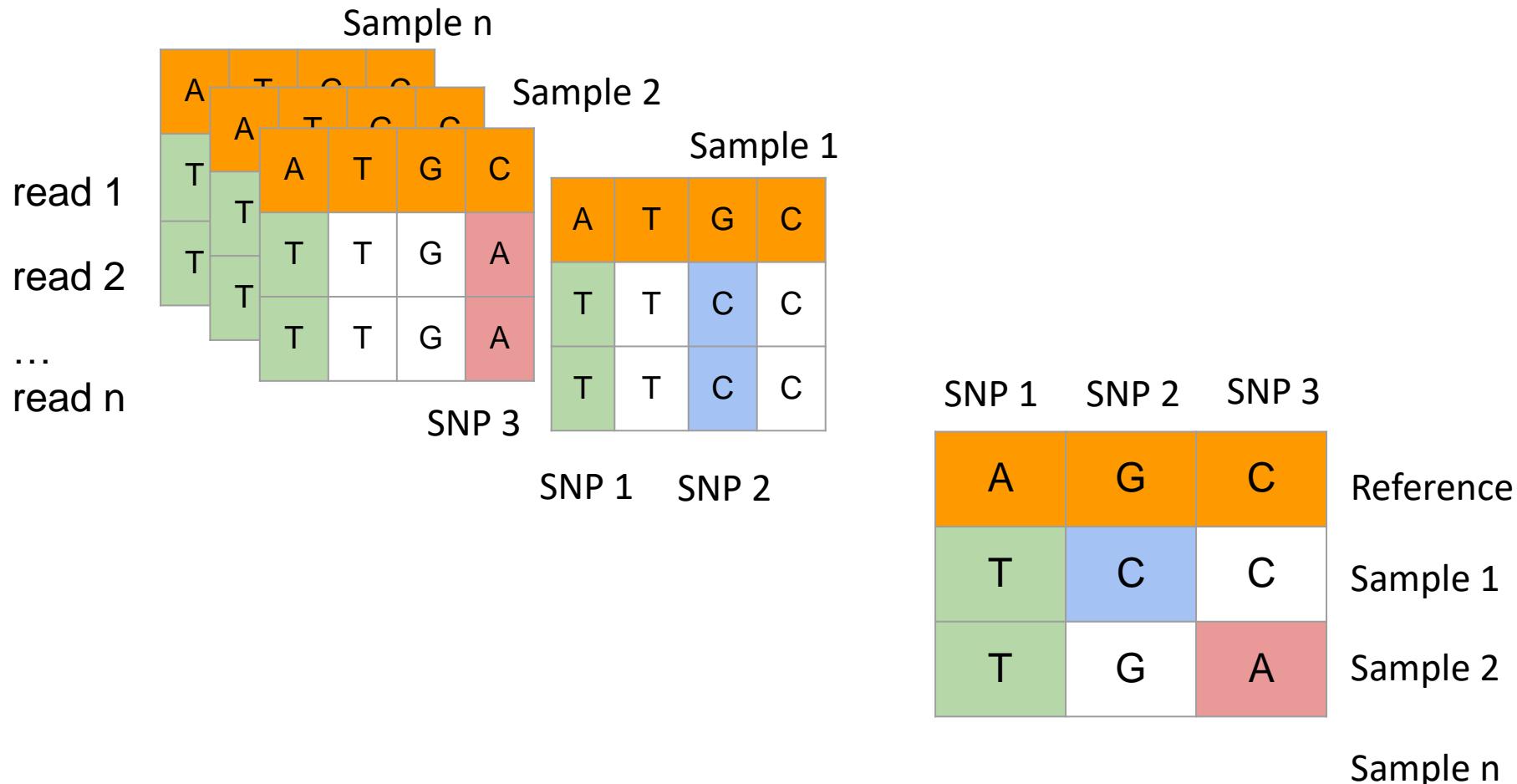
- Live-SET

High-quality SNPs para crear filogenia para investigación de brotes

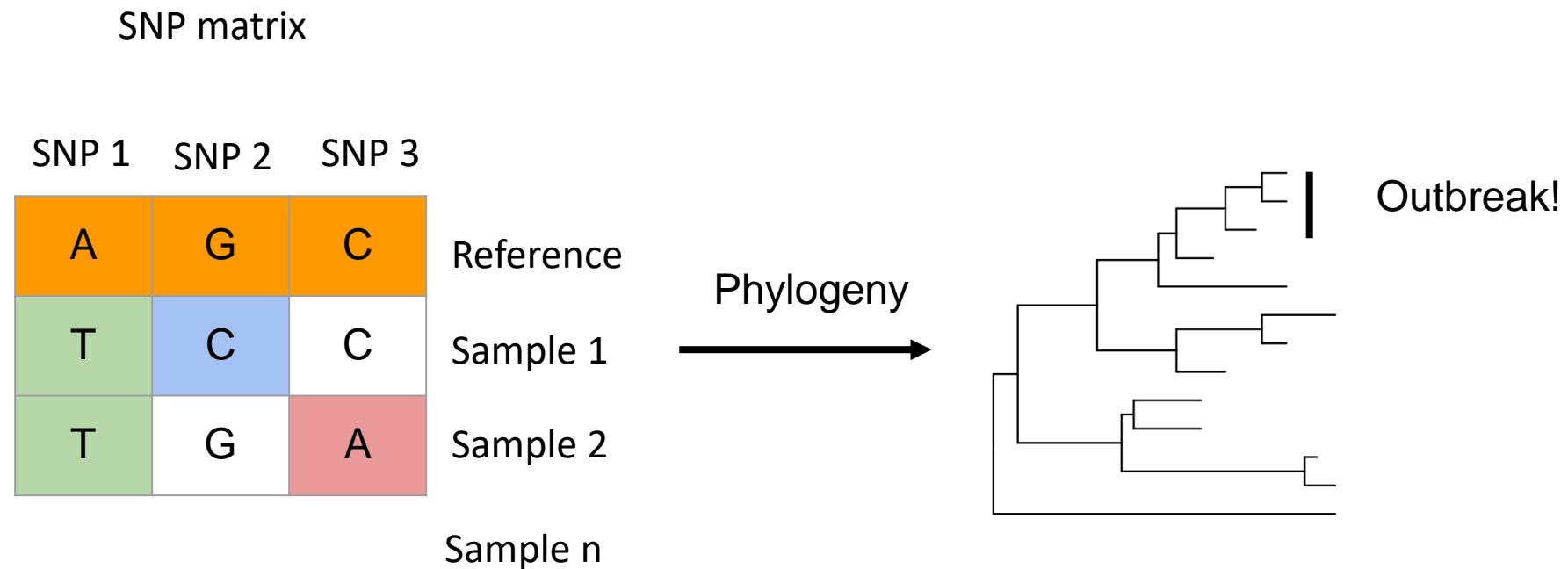
<https://github.com/lskatz/live-SET>

- WGS-Outbreaker

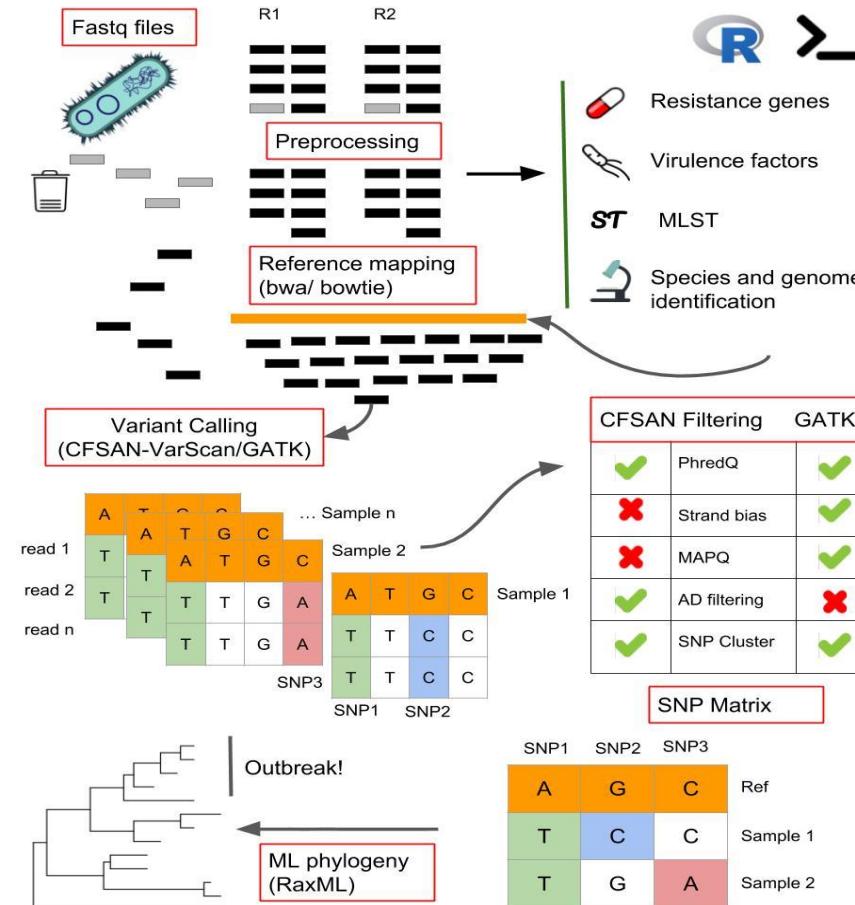
Generación de matriz de SNPs – BACTERIA -OUTBREAK ANALYSIS



Generación de matriz de SNPs – BACTERIA –OUTBREAK ANALYSIS



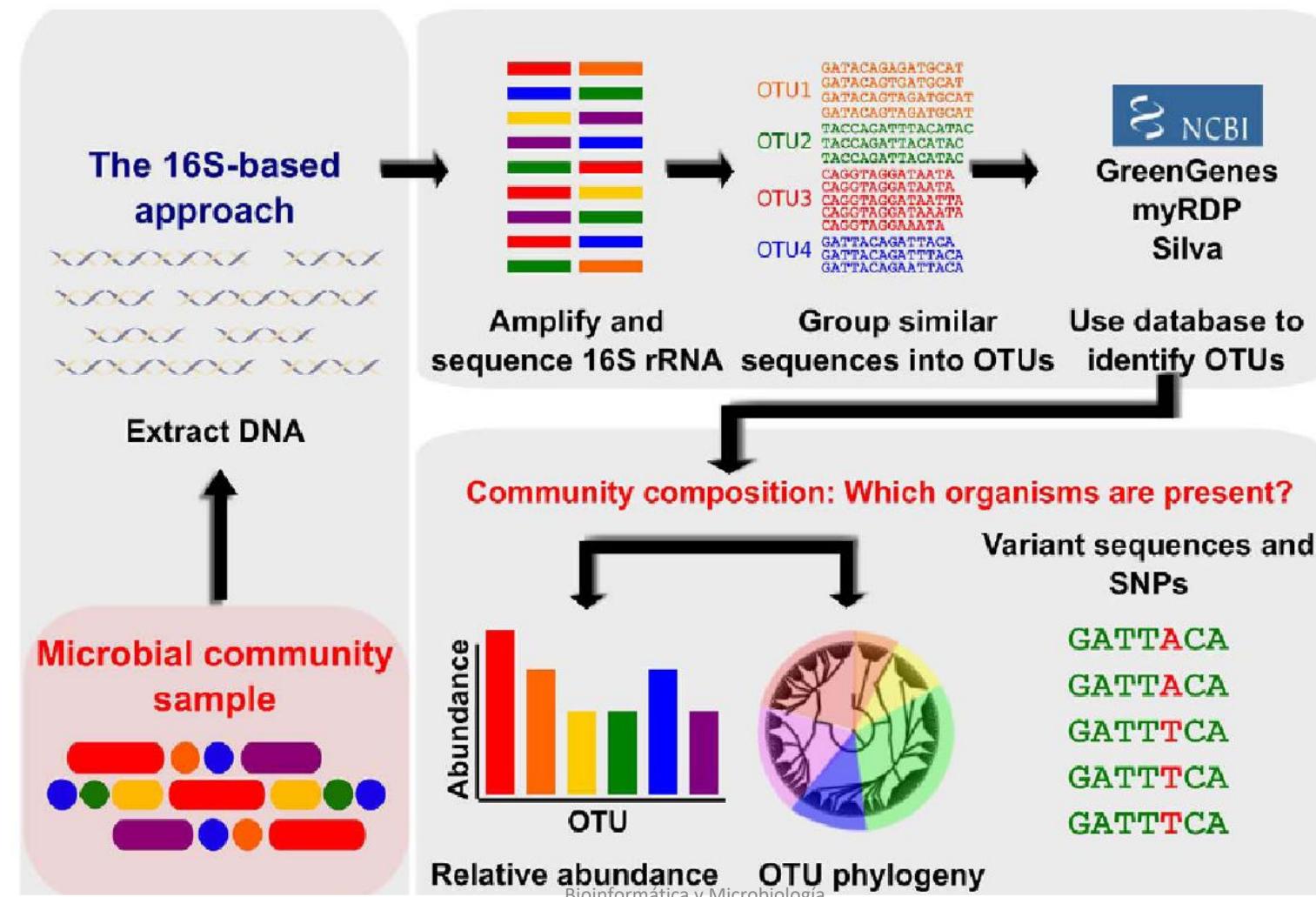
WGS-Outbreaker <https://github.com/BU-ISCIII/WGS-Outbreaker>



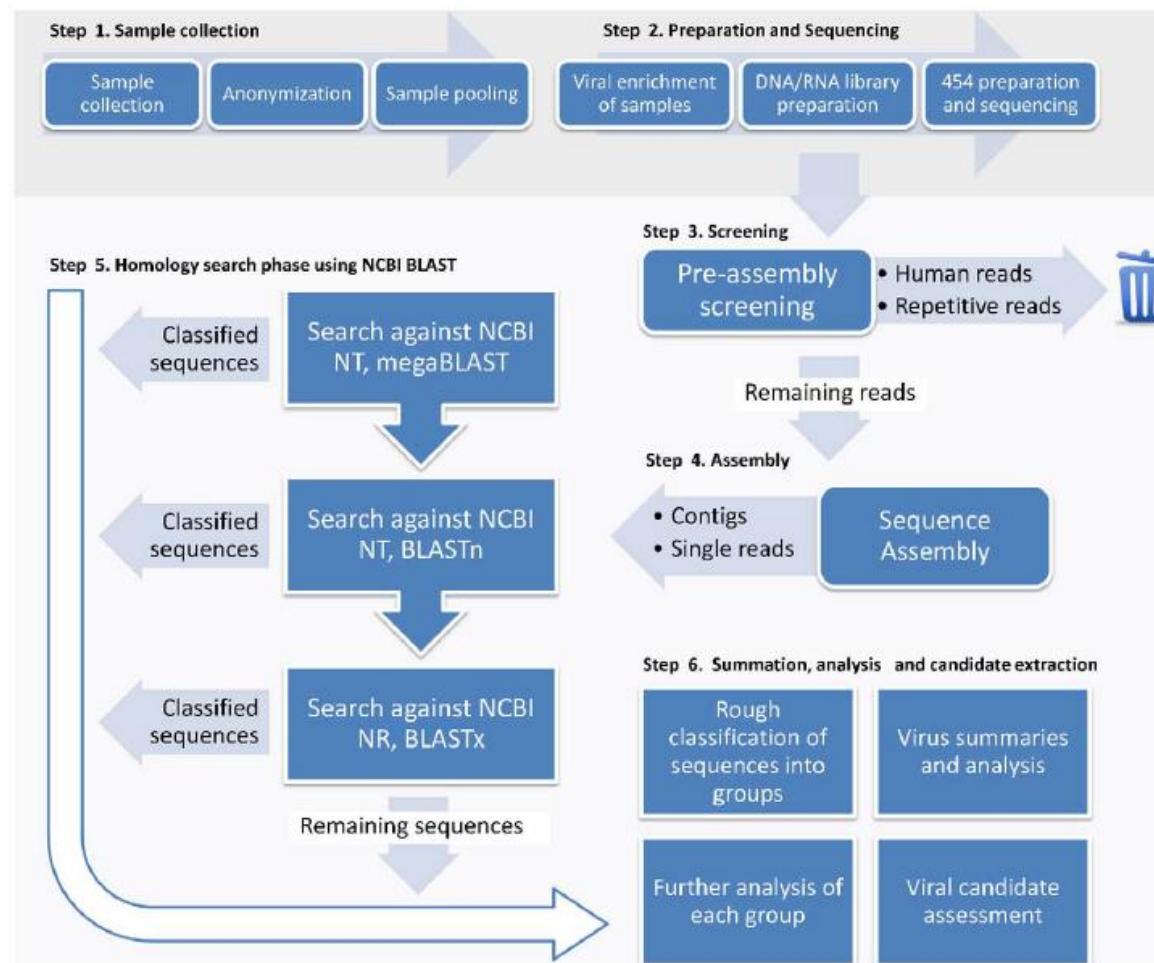
Metataxonomics vs Metagenomics (16S vs Shotgun)

	Metagenetics	Metagenomics
Amplified sequence	Marker regions	Whole genome
Computing time	Usually short	Usually long
Taxonomic composition	Yes	Yes
New pathogen detection	No	Yes
Genome coverage information	No	Yes

Metataxonomics – Target Metagenomics



Metagenomics



Lysholm et al., Plos One 2012:7,2, e30875

Index

- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology. Examples
- Concepts: HTS and Outbreak investigation
- Bacterial and Viral Genome Sequencing
- Bioinformatics analysis in microbial genomics
- **Viralrecon: SARS-CoV-2 genome reconstruction software**