

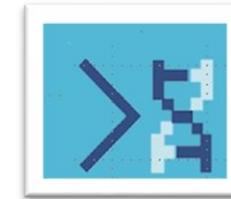
Bioinformática aplicada a la Microbiología Clínica



Isabel Cuesta y Sara Monzón

UNIDAD DE BIOINFORMÁTICA (BU-ISCIII)

Unidades Centrales Científico Técnicas – SGSAFI-ISCIII



30 Marzo 2023

Master Bioinformática aplicada a la Medicina Personalizada y la Salud

Index

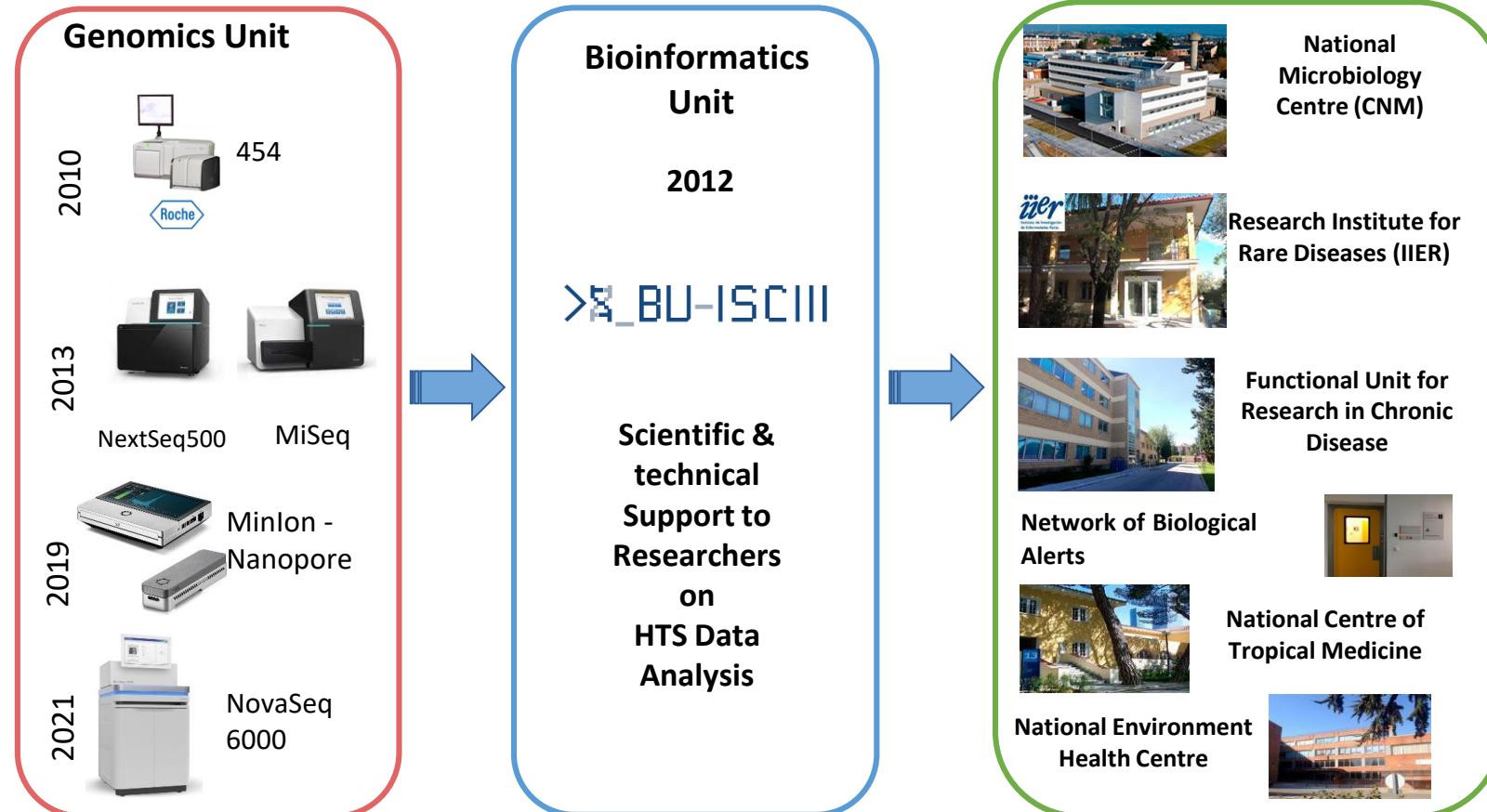
- BU-ISCIII
- High-throughput sequencing (HTS) applications in Microbiology
- Concepts: HTS - library strategies and Outbreak investigation
- Bioinformatics analysis in microbial genomics
- Bacterial and Viral Genome Sequencing
- Viralrecon: SARS-CoV-2 genome reconstruction software

Index

- **BU-ISCIII**

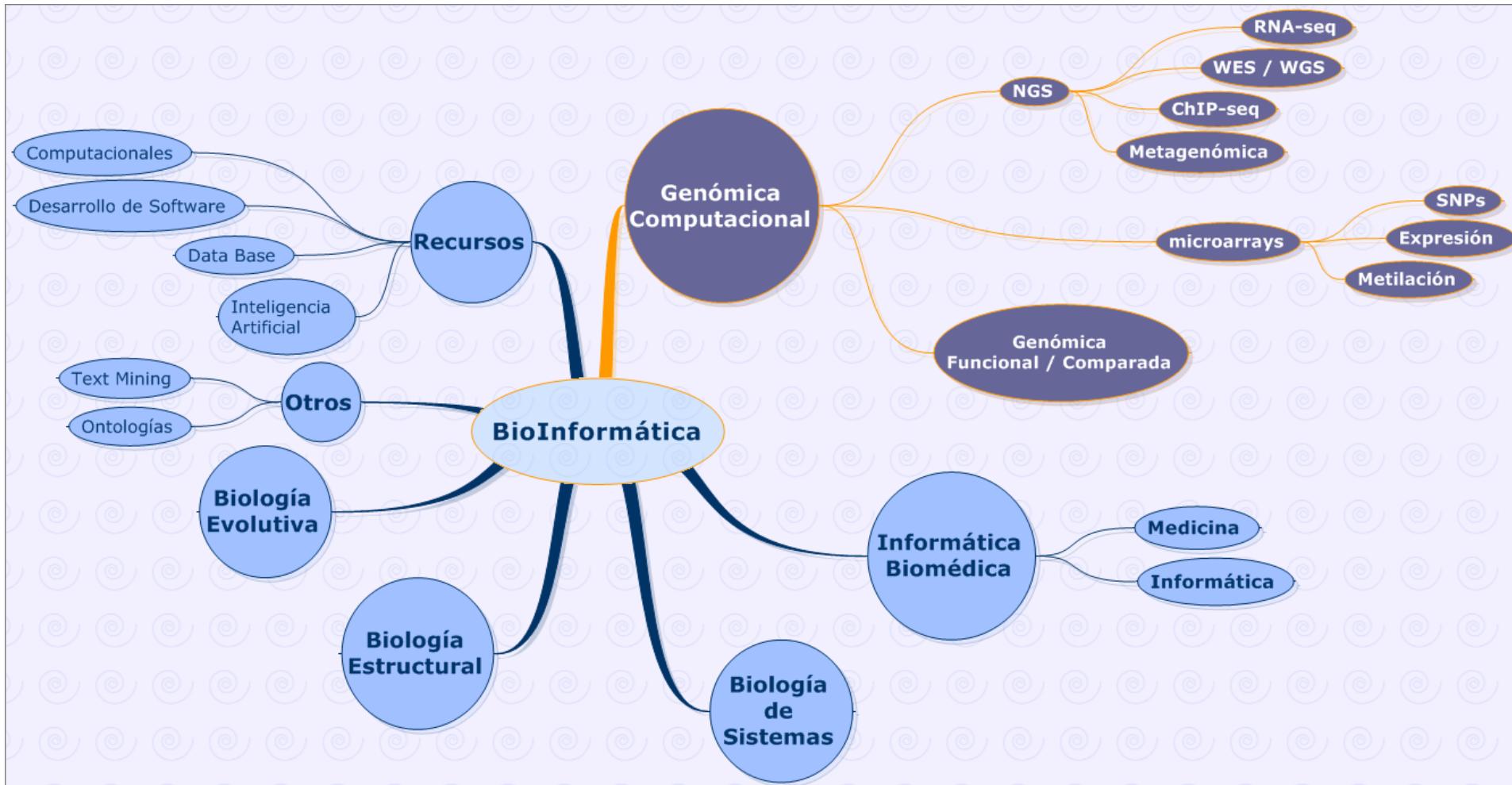
- High-throughput sequencing (HTS) applications in Microbiology
- Concepts: HTS – library strategies and Outbreak investigation
- Bioinformatics analysis in microbial genomics
- Bacterial and Viral Genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

BU-ISCIII: As example of Clinical Bioinformatics Unit



11
AÑOS

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

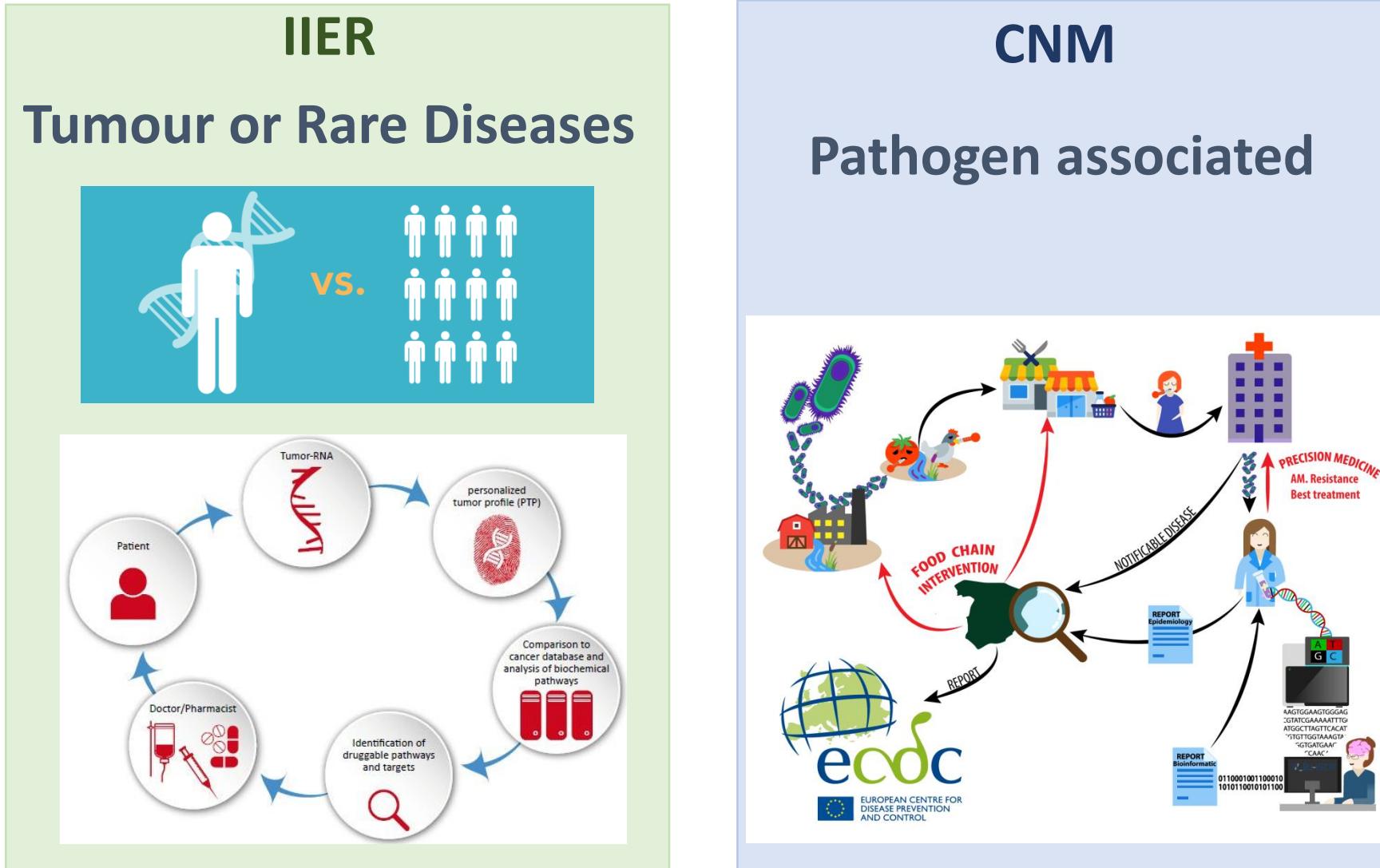
- INFRASTRUCTURE
 - HPC
 - Workstations
- REPOSITORY:
 - GitHub
 - Benchmarking
 - Validation

DATA MANAGEMENT



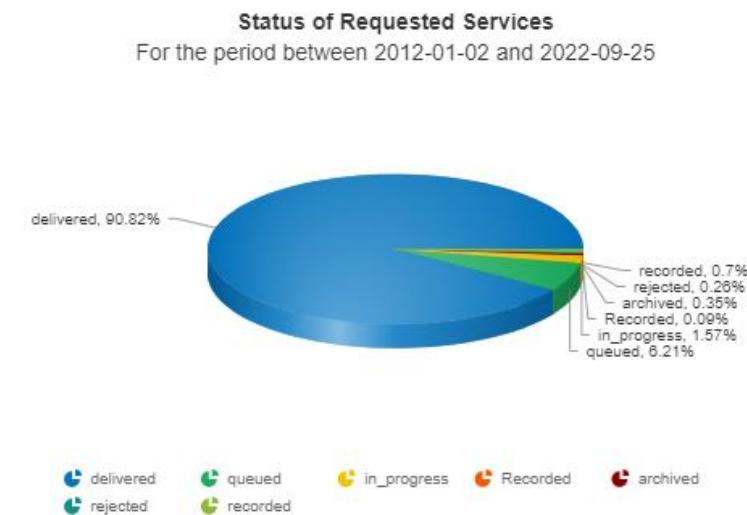
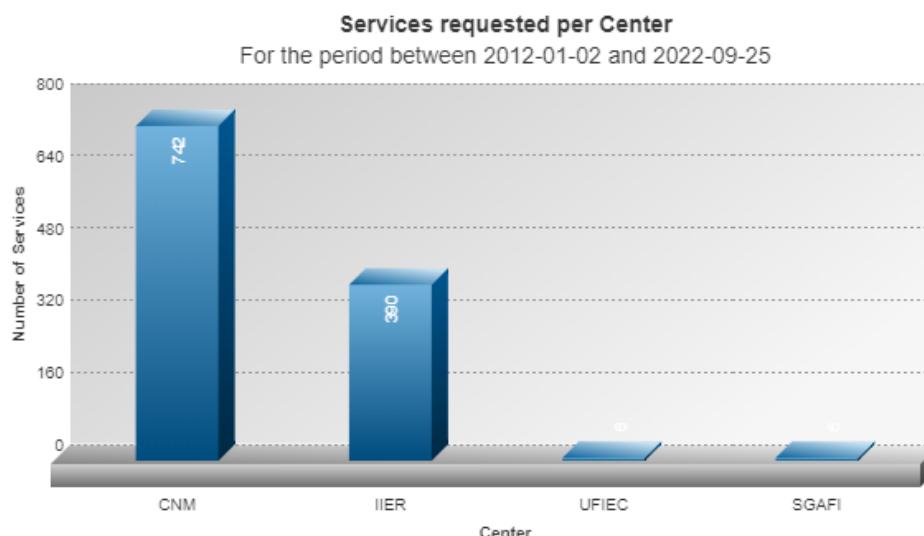
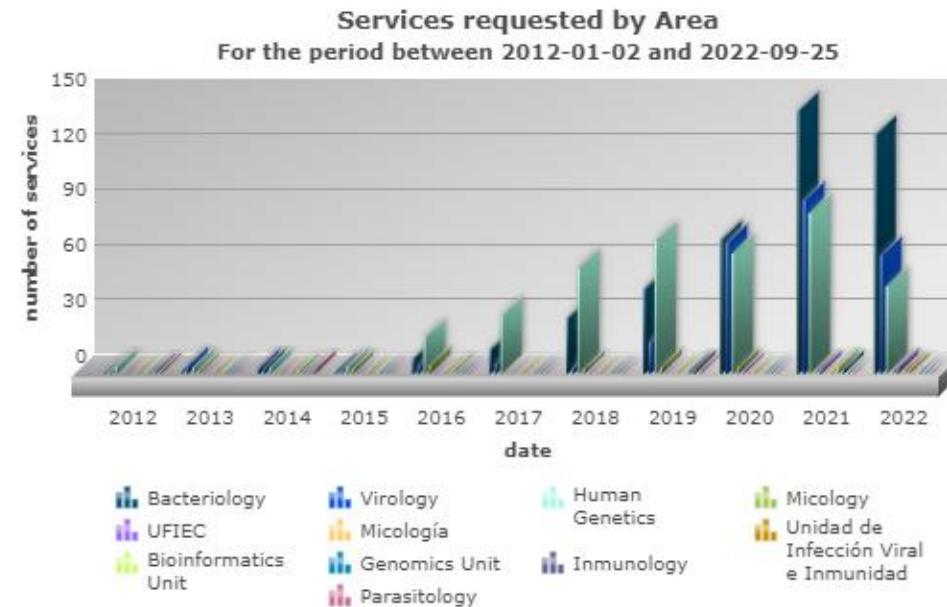
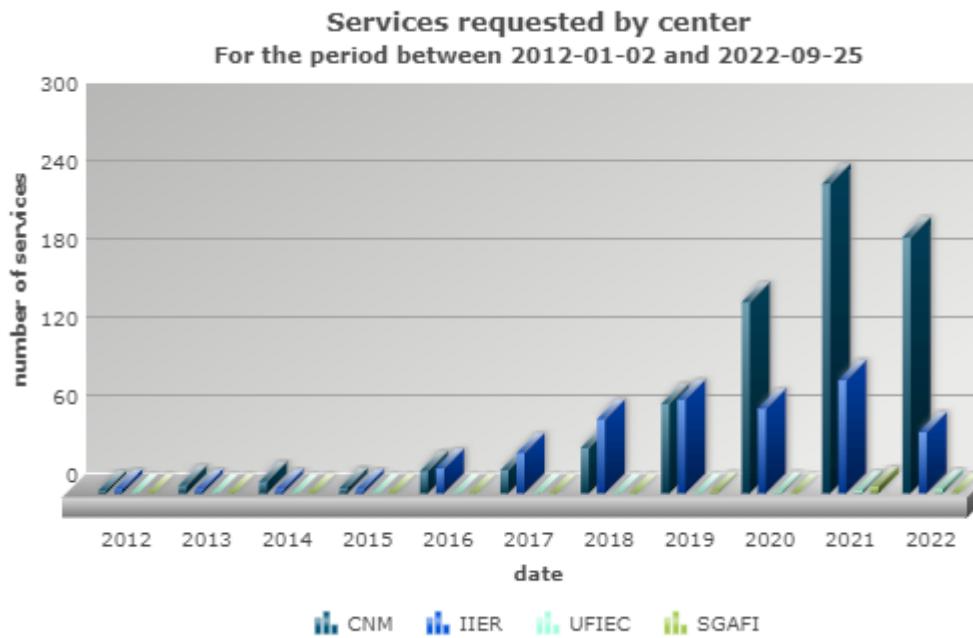
- INFRASTRUCTURE
- DATA WORKFLOW
- LIMS

Clinical Bioinformatics - Precision Medicine



Number of services: 2012 – 2022

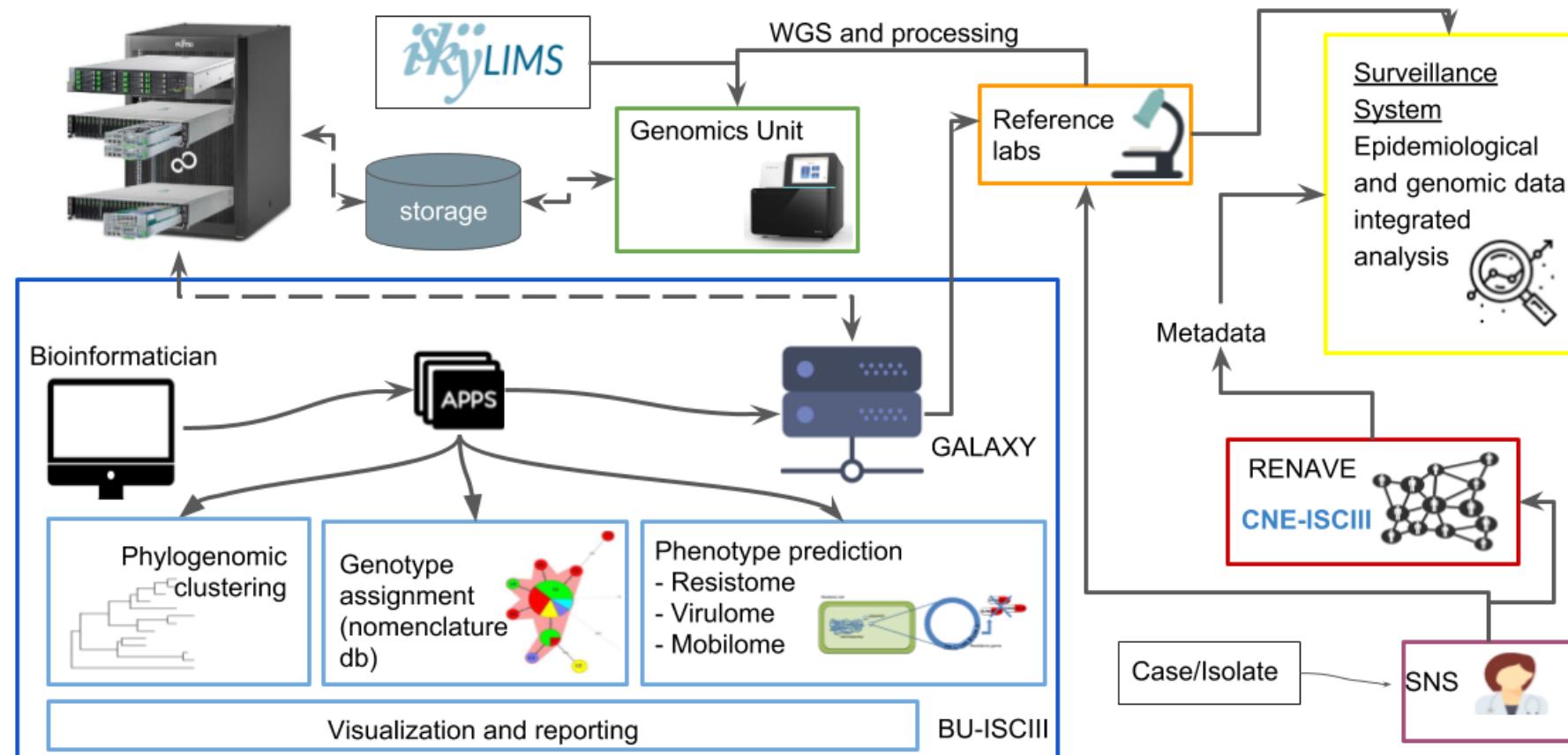
> BU-ISCIII



10
AÑOS

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



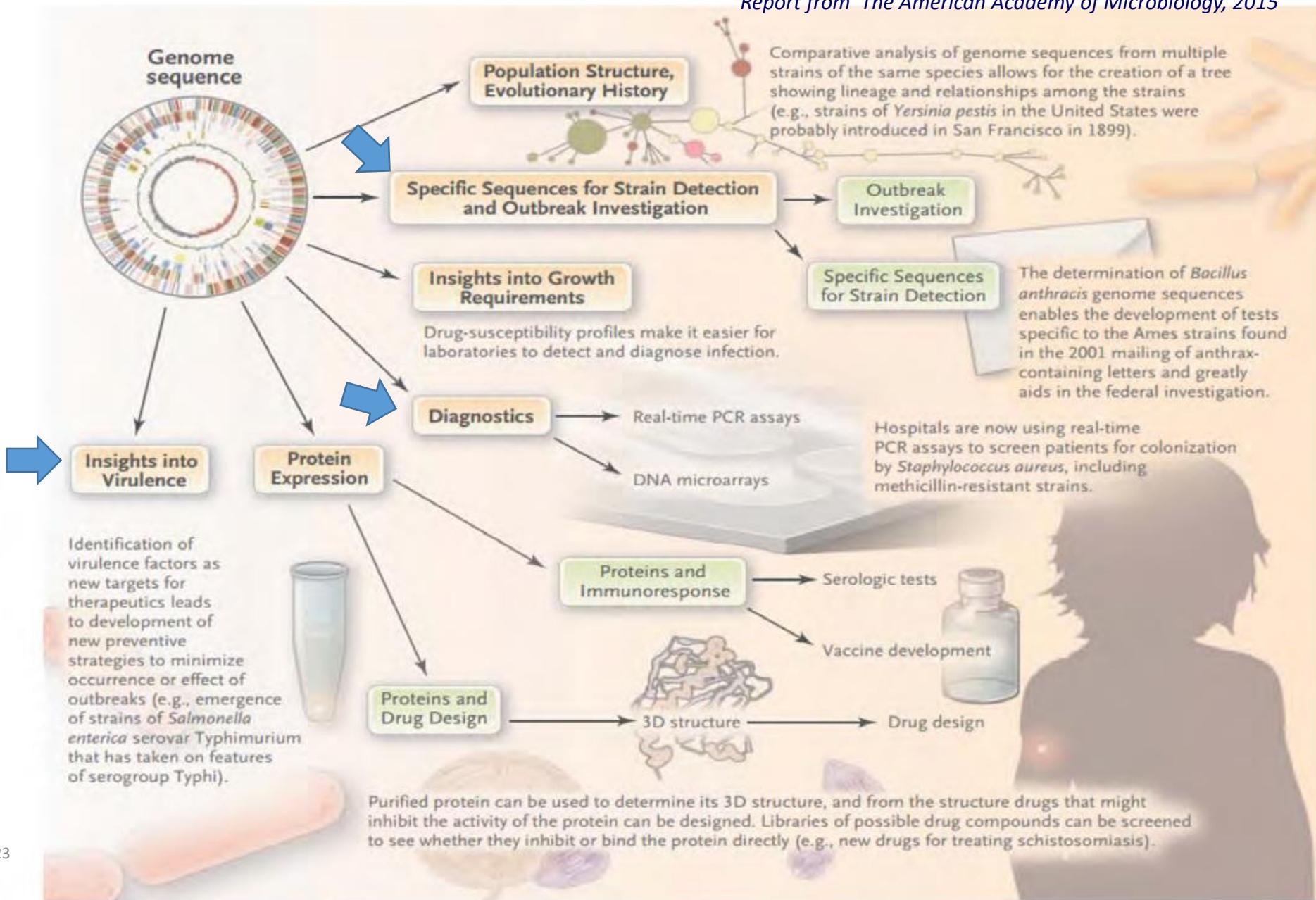
Index

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

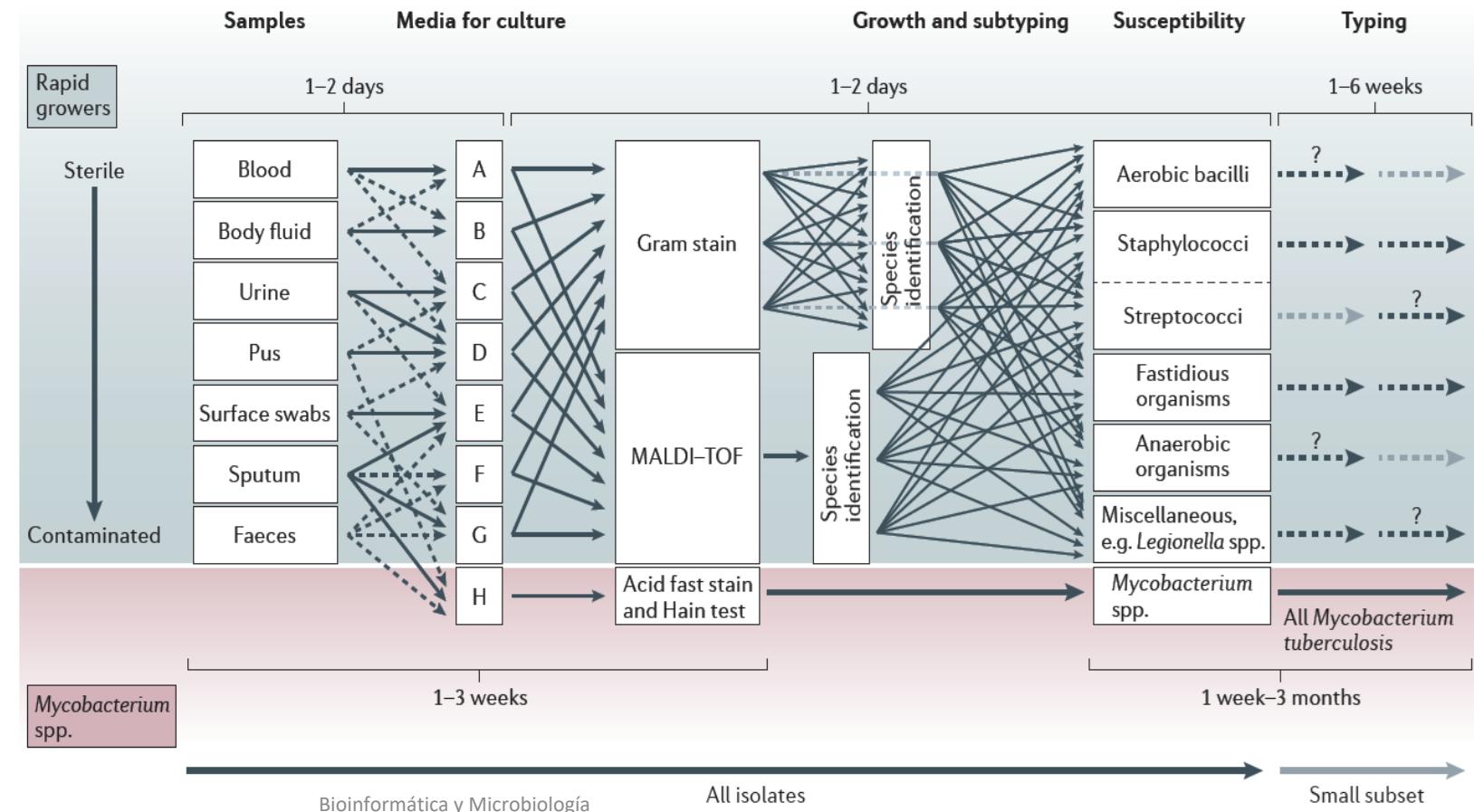
Microbial genomics

>X_BU-ISCIII

Report from The American Academy of Microbiology, 2015



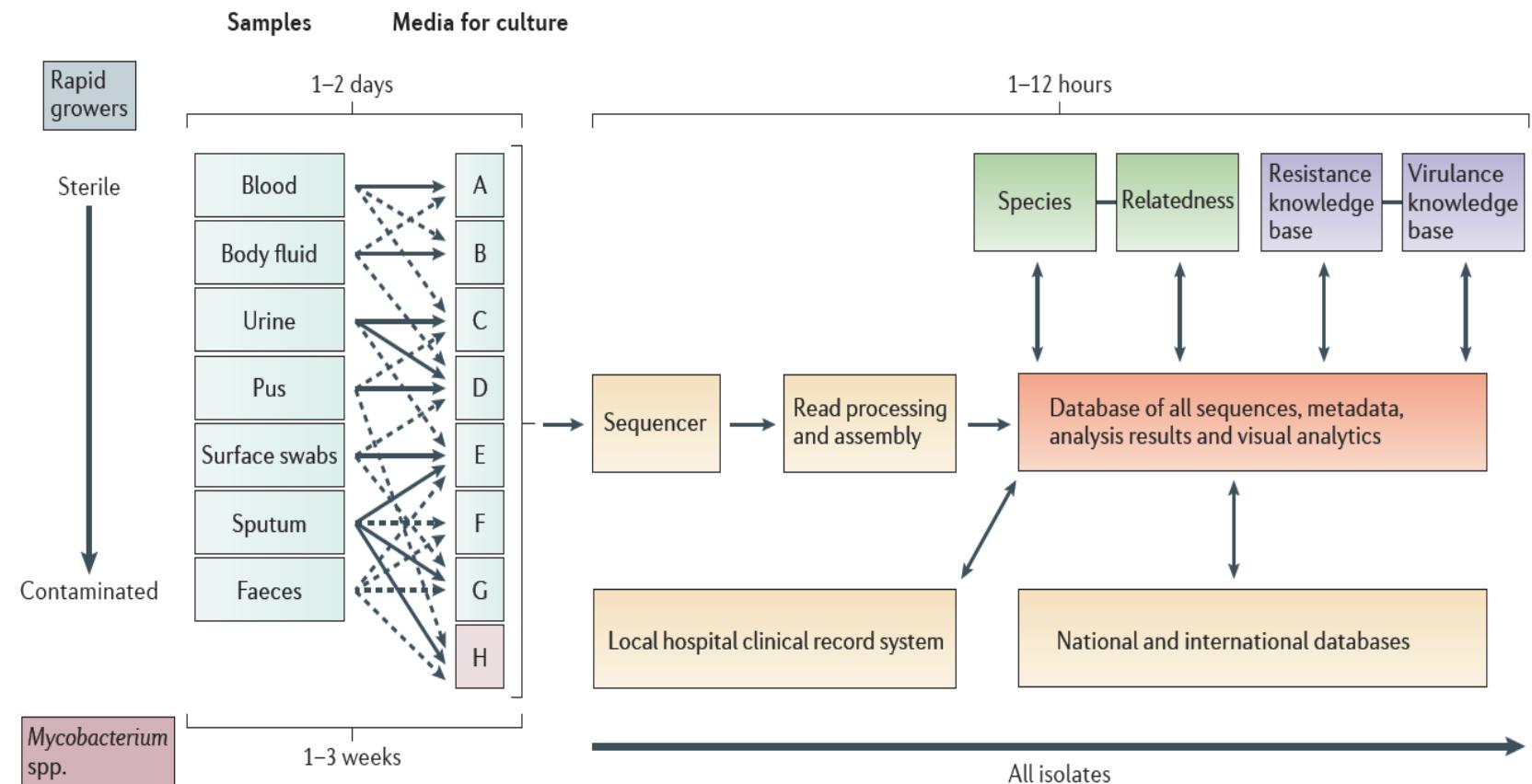
Classic techniques vs Whole Genome sequencing



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

ECDC roadmap and international commitment

Classic techniques vs Whole Genome sequencing



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

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

Causado por la toxo-infección de Escherichia coli enterohemorrágica (EHEC) (*Escherichia coli* O104:H4)

Muerte: 32 personas en Alemania, 1 Suecia y 1 Francia y 2263 infectados en 12 países de Europa.

Crisis Política y Económica Europa:
Alto impacto en la Economía Europea, mayor afectación en la Española



Universitätsklinikum
Hamburg-Eppendorf

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

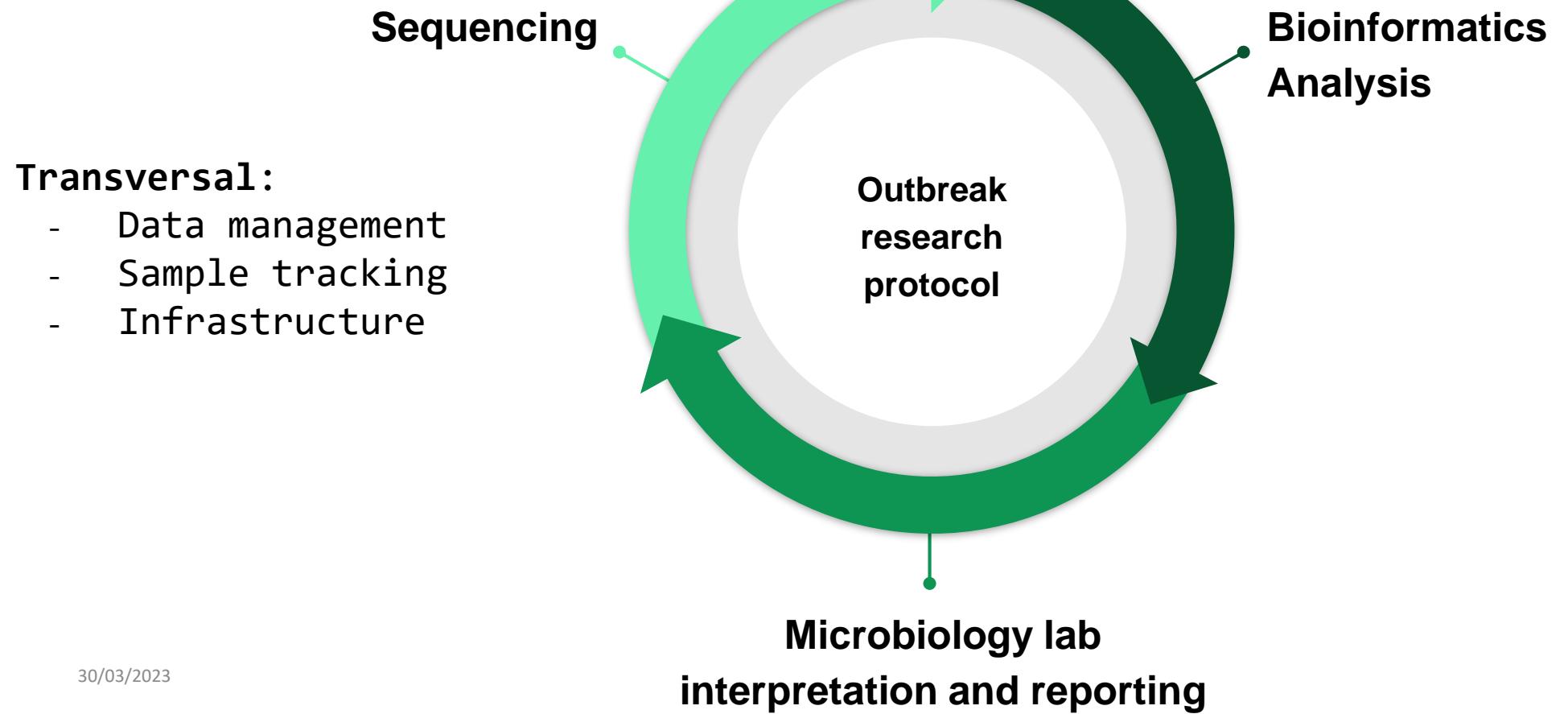
EFE 25.08.2019

- 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

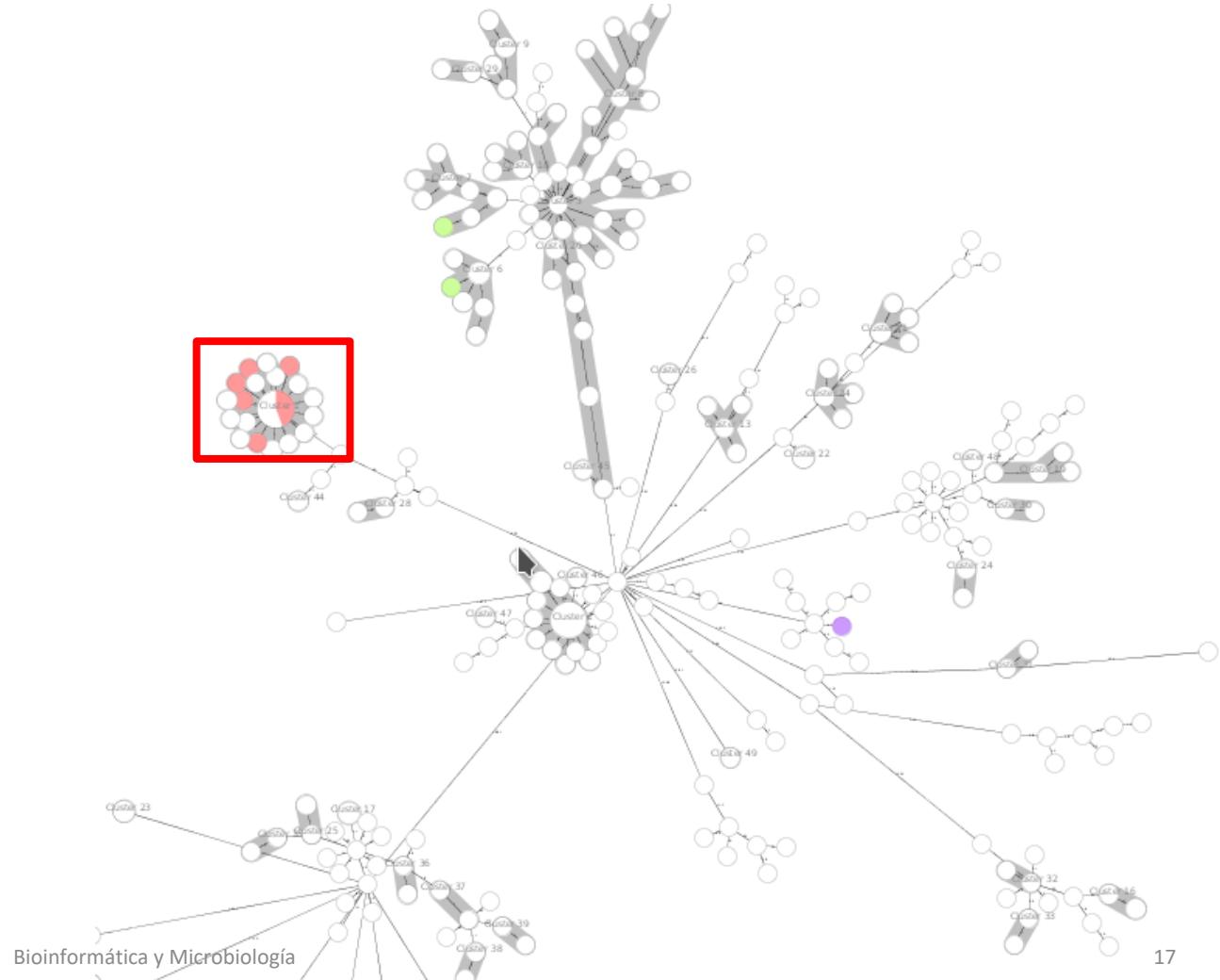


Andalusian Listeria Outbreak

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

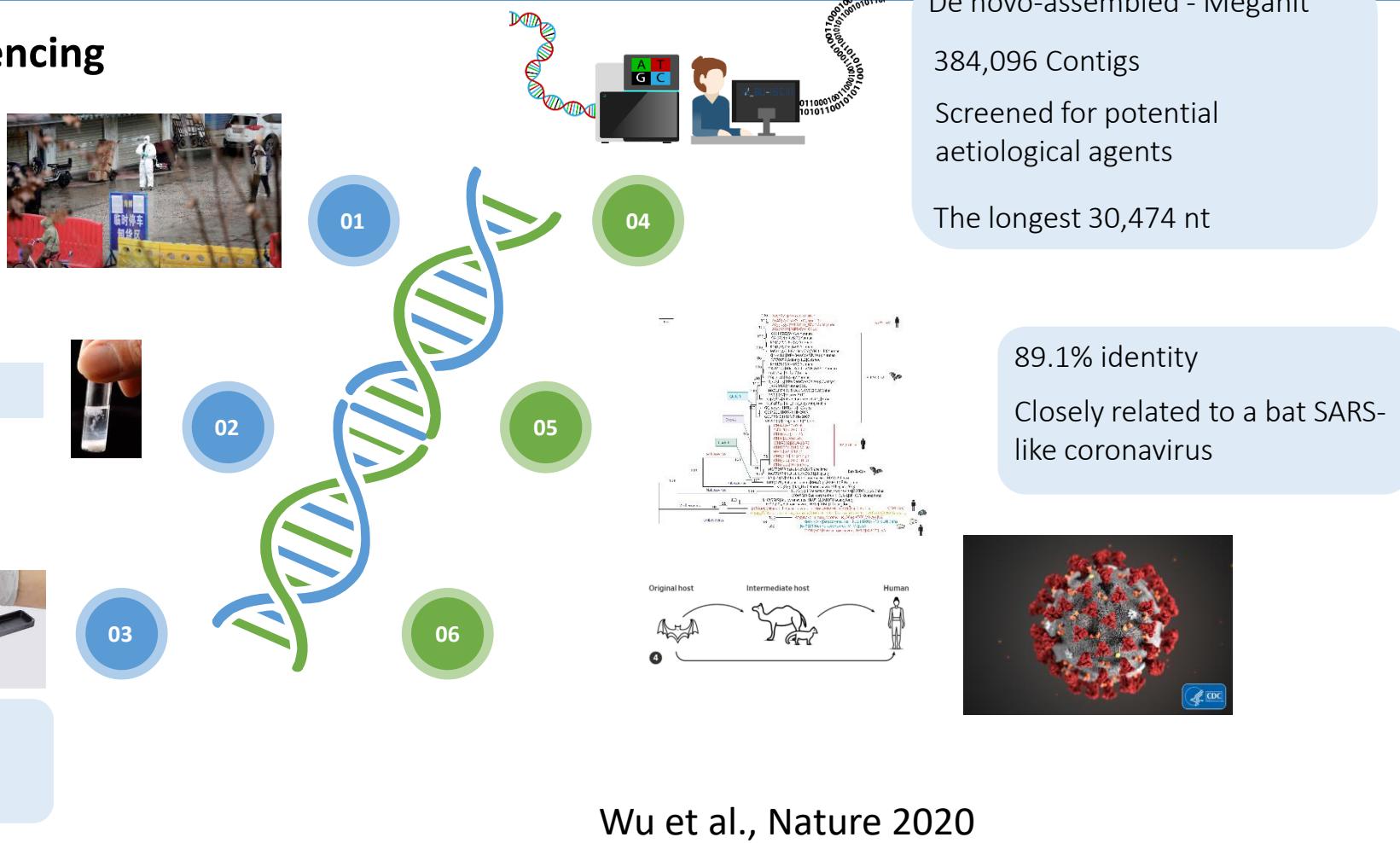
Results:

- 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

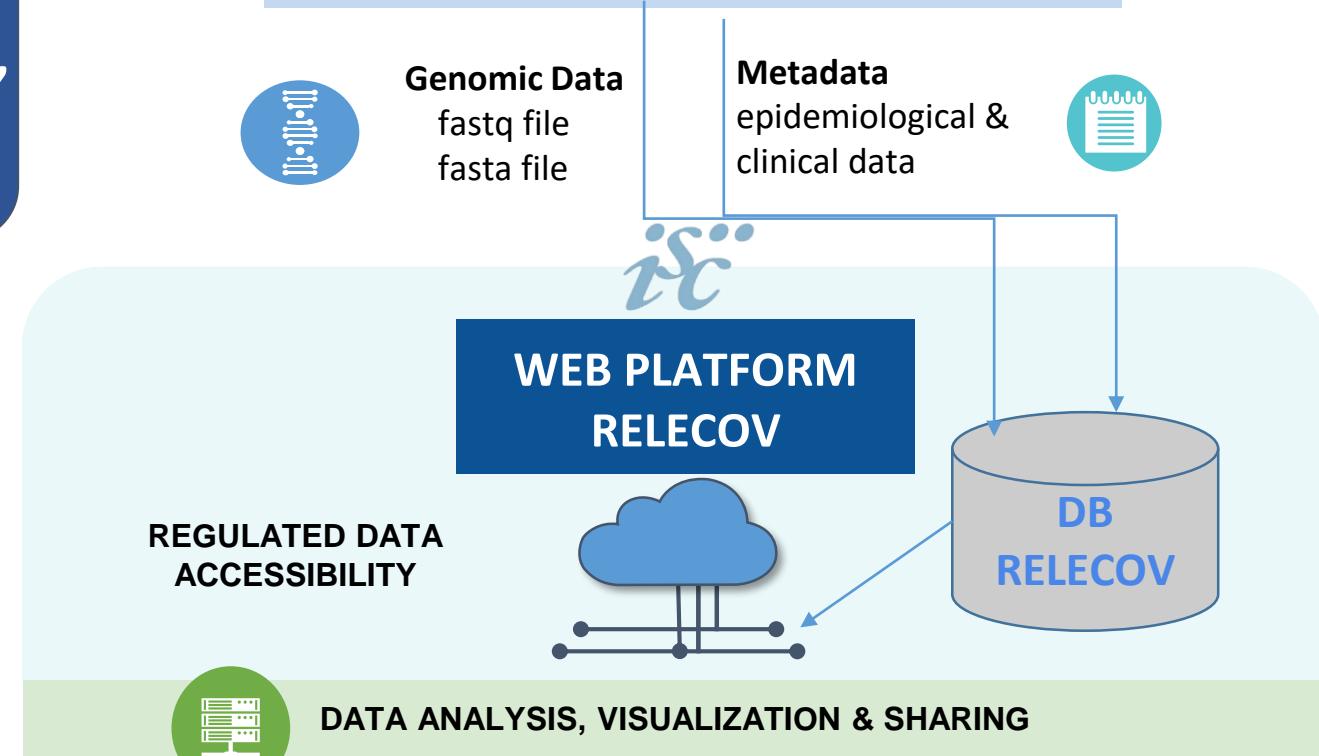


Meta-transcriptomic library

2x150 MiniSeq 56,565,928 sequences reads

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
- 30/03/2023

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

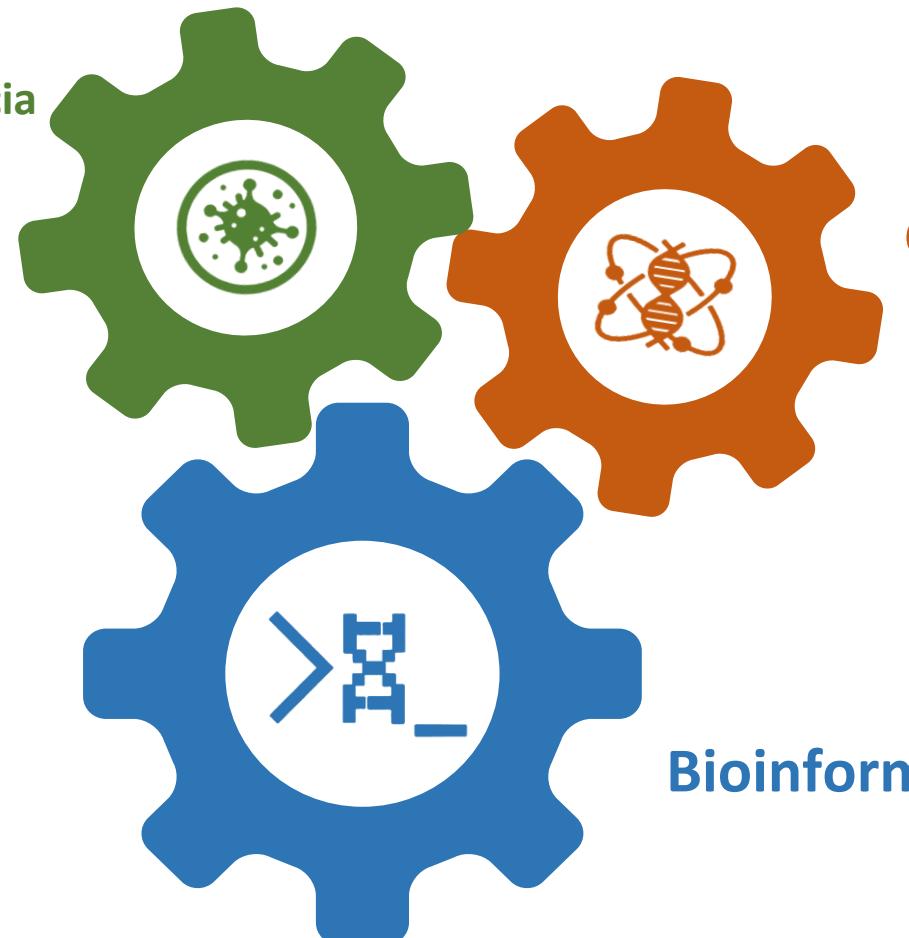
- Detection and variant surveillance.
- **Nextstrain Spain**
- **outbreak.info**
- **COG-UK Consortium**

International sharing

GISAID
ENA
European Nucleotide Archive

M2: Organization of circuits between hospitals and regional laboratories

Microbiología
Laboratorios de Referencia



Genómica

Bioinformática

Bioinformática y Microbiología

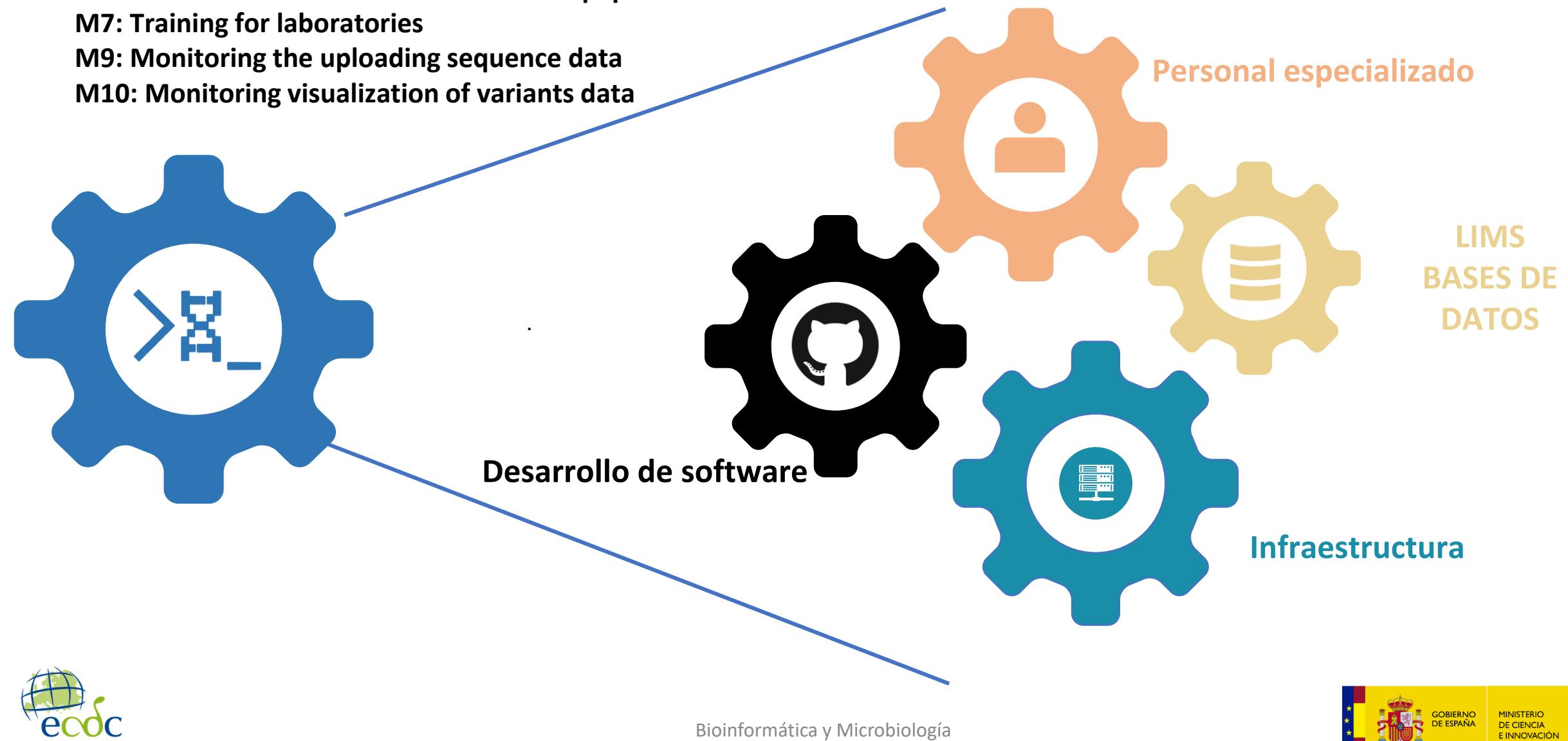
Milestones

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

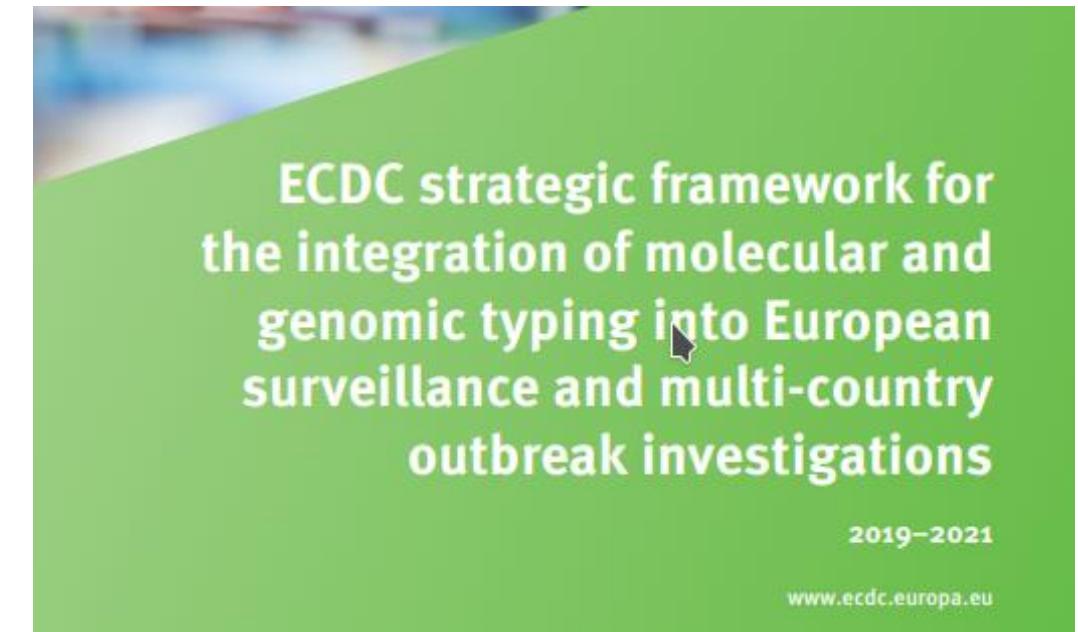
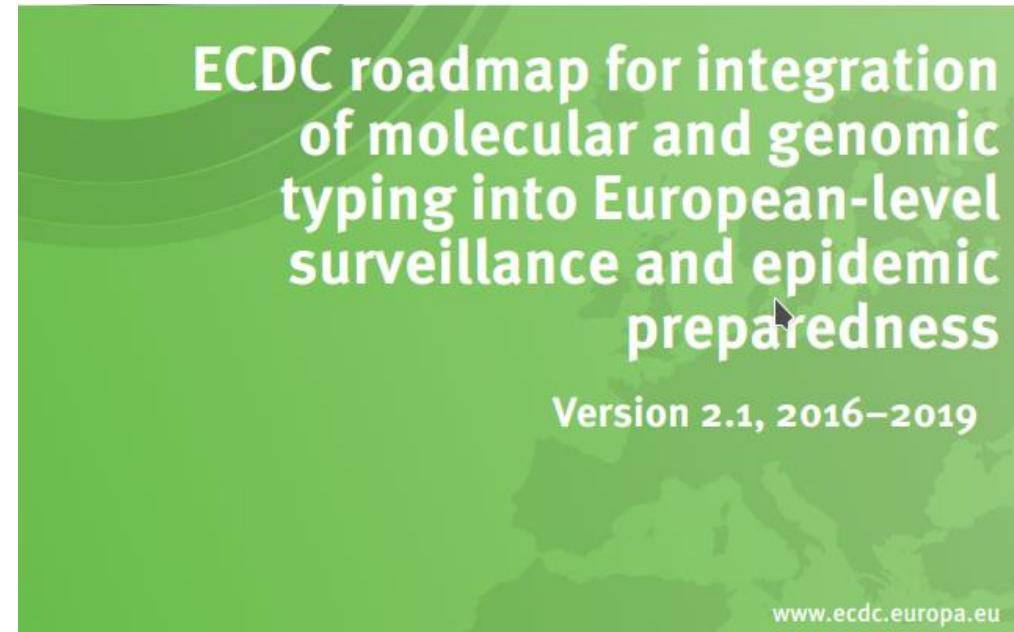


Formación

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



ECDC roadmap and international commitment



- **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- library strategies and Outbreak investigation**
- Bioinformatics analysis in microbial genomics
- Bacterial and Viral Genomics
- Viralrecon: SARS-CoV-2 genome reconstruction software

Estrategias basadas en preparación de librería



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

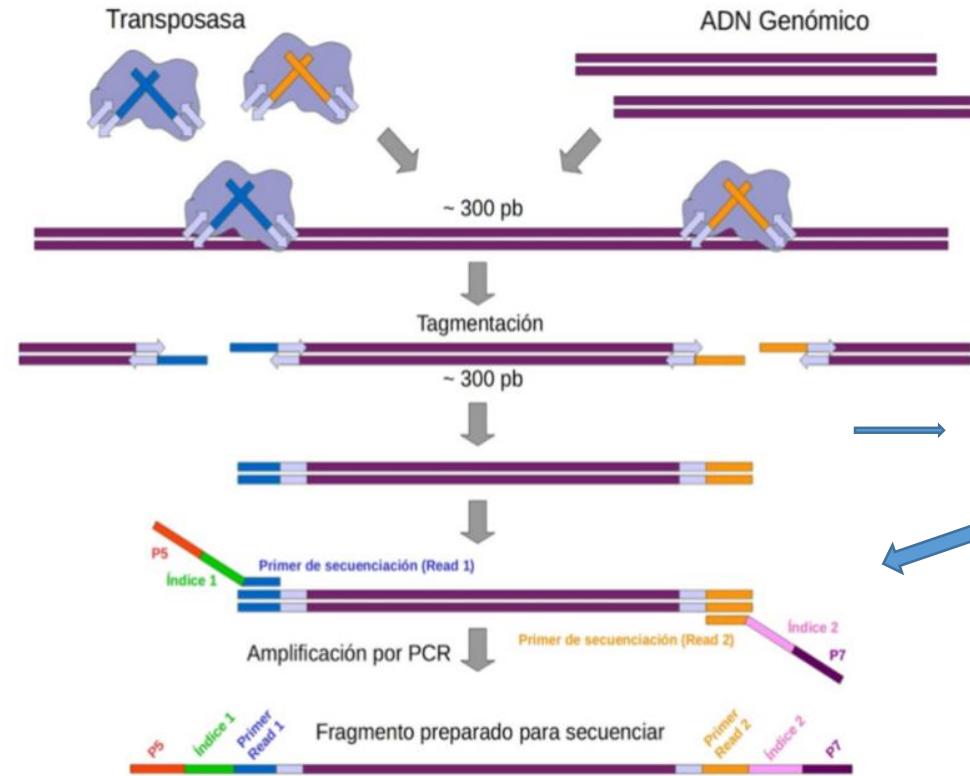
1. Metagenómica

IDENTIFICACIÓN MICROORGANISMOS

1. Metataxonomía

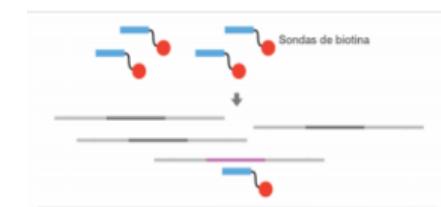
PREPARACIÓN LIBRERÍA

ENZIMÁTICA FÍSICA



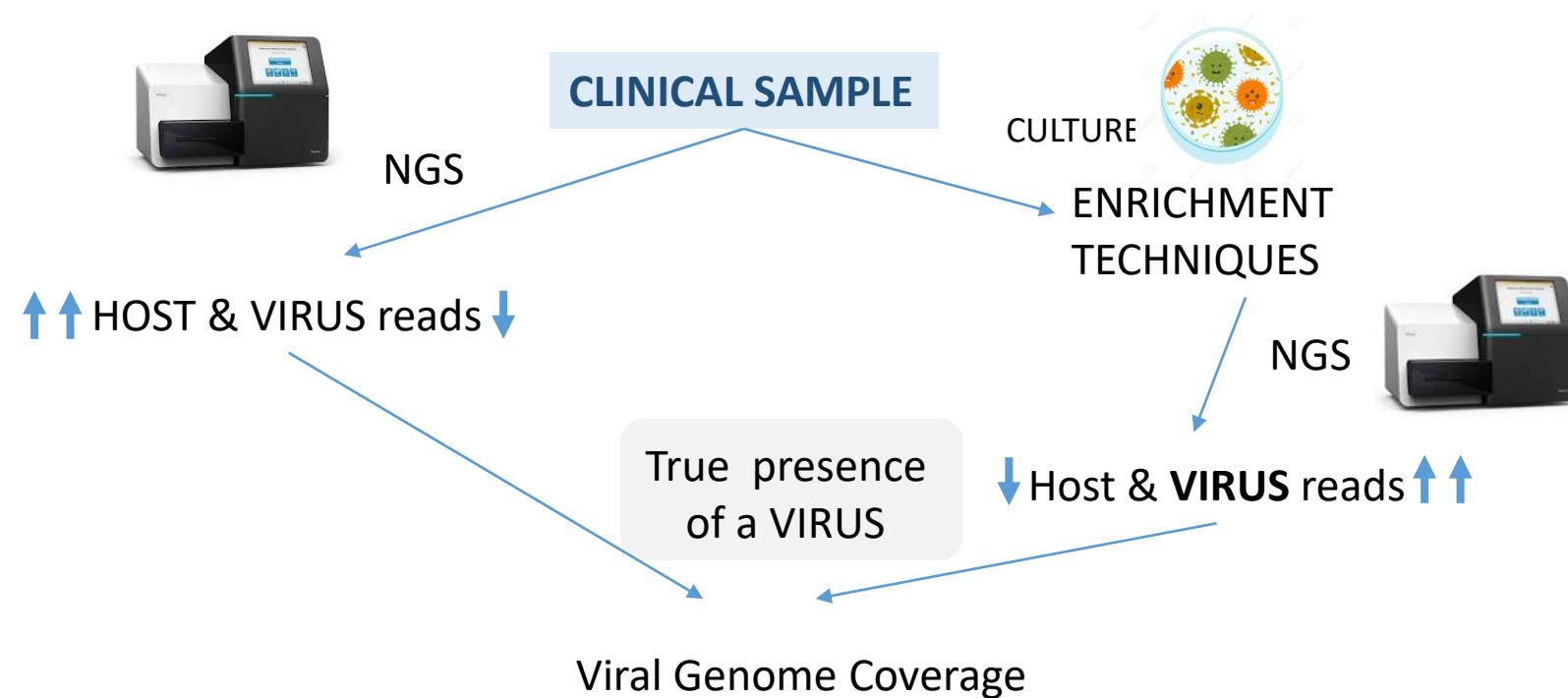
RNA -> cDNA

**ENRIQUECIMIENTO:
PCR
CAPTURA SONDAS**



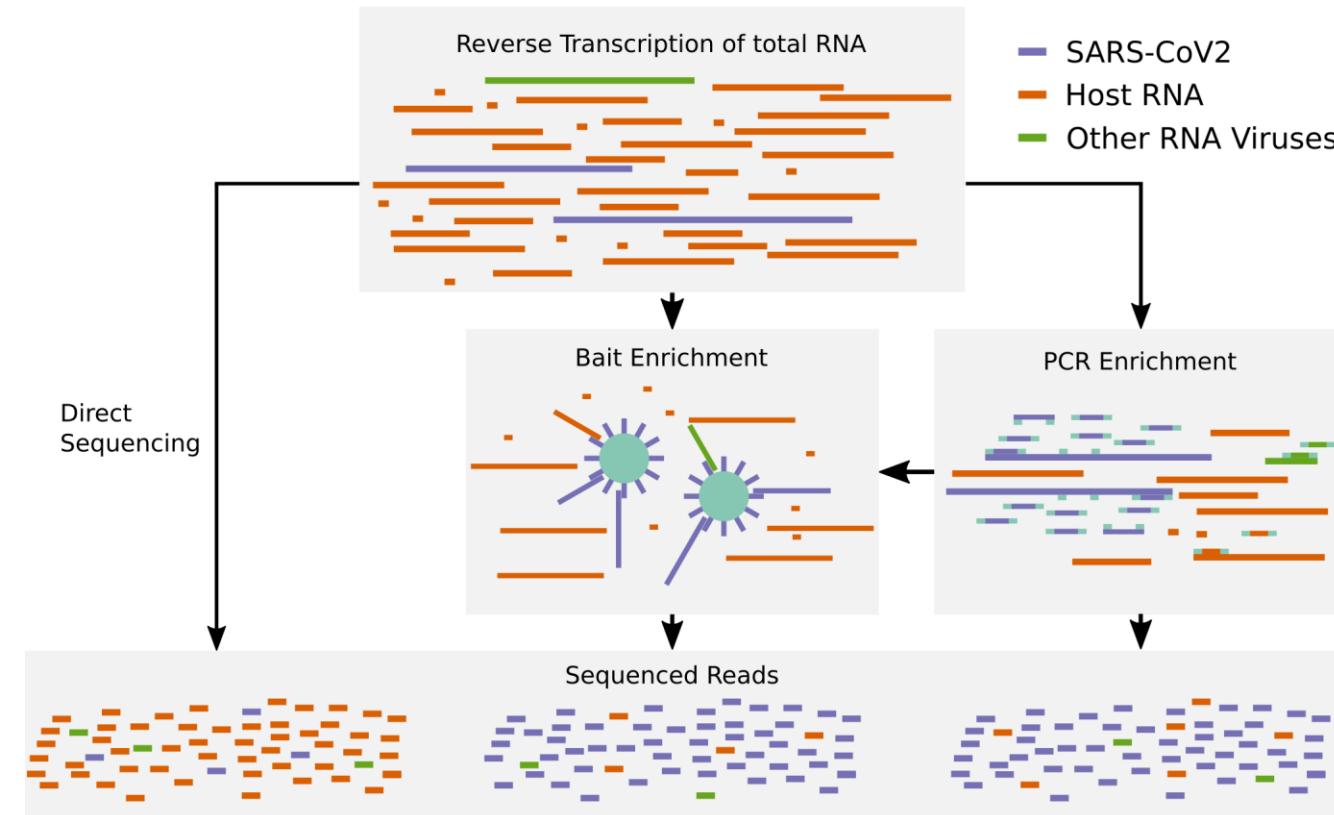
Guia Práctica Genómica https://www.uv.es/varnau/GM_Cap%C3%ADtulo_2.pdf

Viral Genome Sequencing

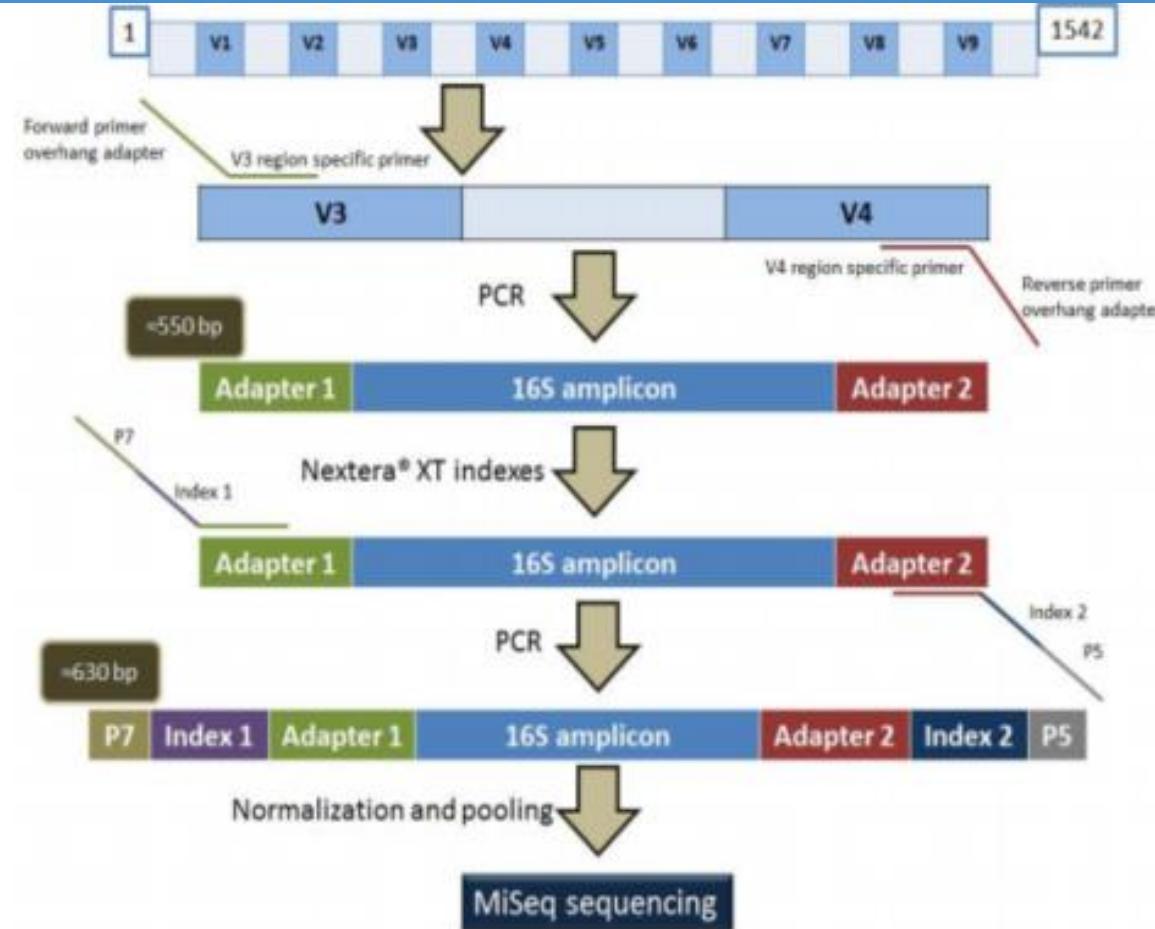


NGS needs a cutoff to determine the true presence of a pathogen versus carry-over or contamination between specimens or other non-specific reads.

Enrichment Techniques



PREPARACIÓN LIBRERÍA, rRNA 16S, caracterización microbiota



Index

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

Definitions of terms commonly used to classify genetic relatedness among bacterial strains

Simar et al., Current Opinion 2021

Table 2. Definitions of terms commonly used to classify genetic relatedness among bacterial strains

| Term | Definition |
|--------------------|---|
| Clade | A group of organisms that contains a single ancestor and its descendants; a monophyletic group [6] |
| Clone | A group of isolates that are genetically indistinguishable [though not necessarily identical] based on a particular molecular typing method and are presumed to be descendants of a common ancestor [7] |
| Sequence type (ST) | Organisms that possess identical allelic profiles of fragments of predetermined housekeeping genes [4] |
| Clonal group | All isolates that belong to a particular ST [8,9] |
| Clonal complex | A cluster of bacterial organisms that originate from a common ancestor and generally share at least 6/7 alleles of their associated ST with another member of the group [8] |
| Strain | Isolate(s) that are distinct from other isolates of the same genus and species based on phenotypic and/or genotypic features [9] |

Outbreak definition and Typing methods: DNA-based methods

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

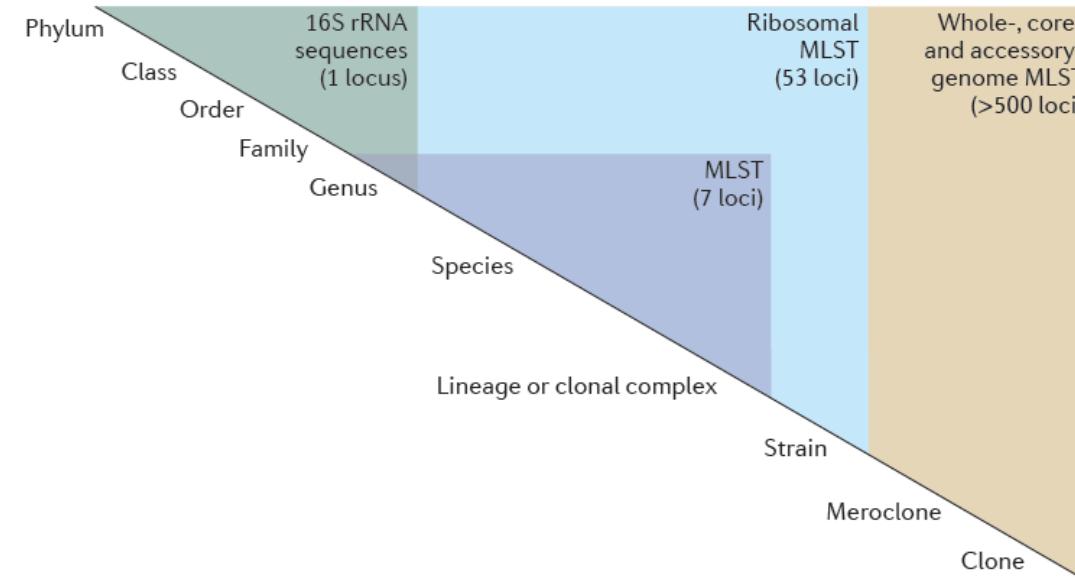
Bacterial identification and characterization at subspecies level is commonly known as **Microbial Typing**. Currently, these methodologies are fundamental tools in Clinical Microbiology and bacterial population genetics studies to track outbreaks and to study the dissemination and evolution of virulence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



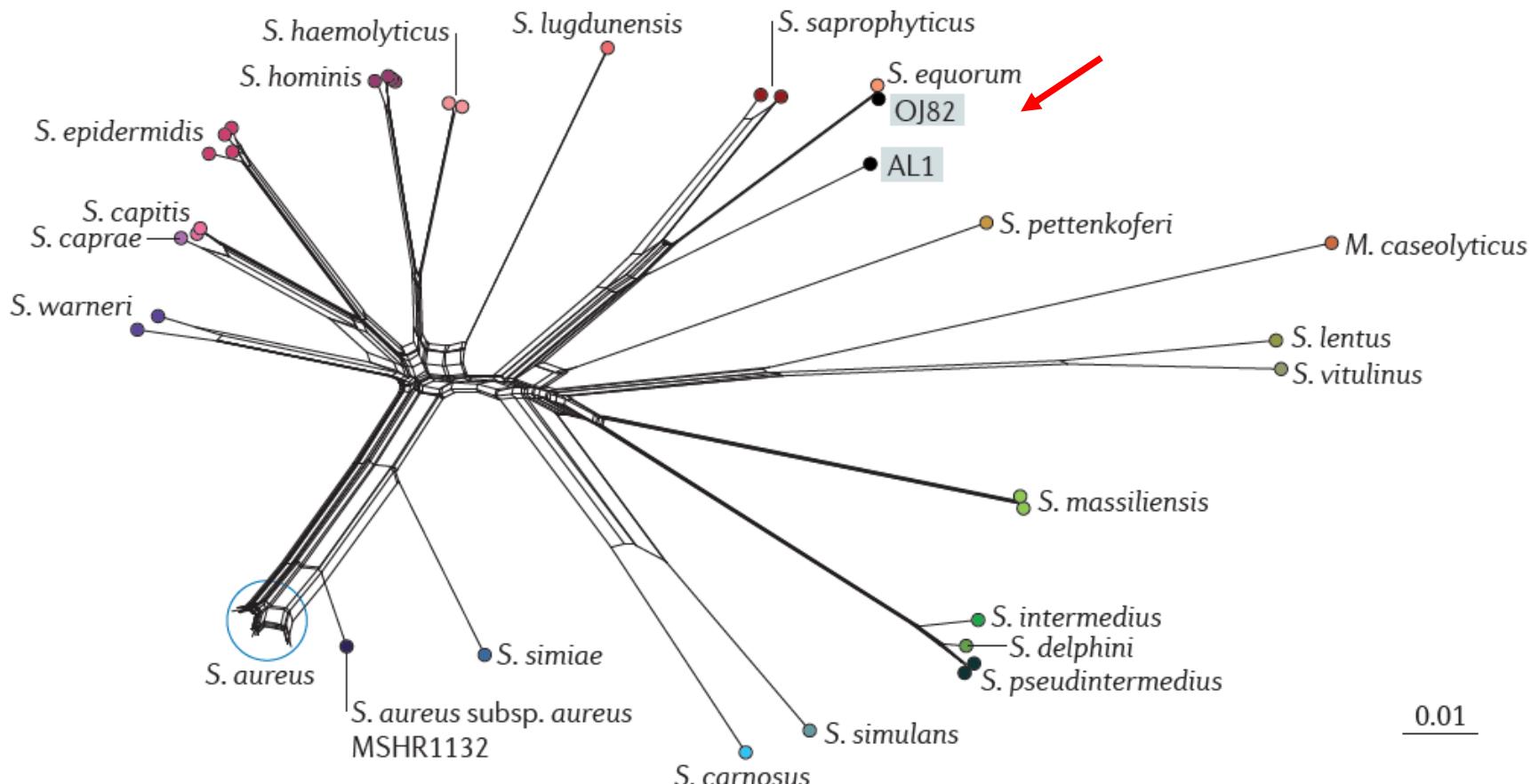
Different levels of sequence information can be associated with different taxonomic levels.

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

Ribosomal multilocus sequence typing-based analysis of *Staphylococcus* spp. whole-genome sequence data

Maiden et al., Nat Rev Microbiology 2013

Resolution of 52 staphylococcal isolates on the basis of nucleotide sequence diversity at 51 ribosomal multilocus sequence typing (rMLST) loci, permitting the determination of the species assignment of two recently described isolates, *Staphylococcus* sp. OJ82 and *Staphylococcus* sp. AL1. *Staphylococcus* sp. OJ82 probably corresponds to *Staphylococcus equorum*, whereas *Staphylococcus* sp. AL1 is related to, but distinct from, *S. equorum* and *Staphylococcus saprophyticus*.



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.

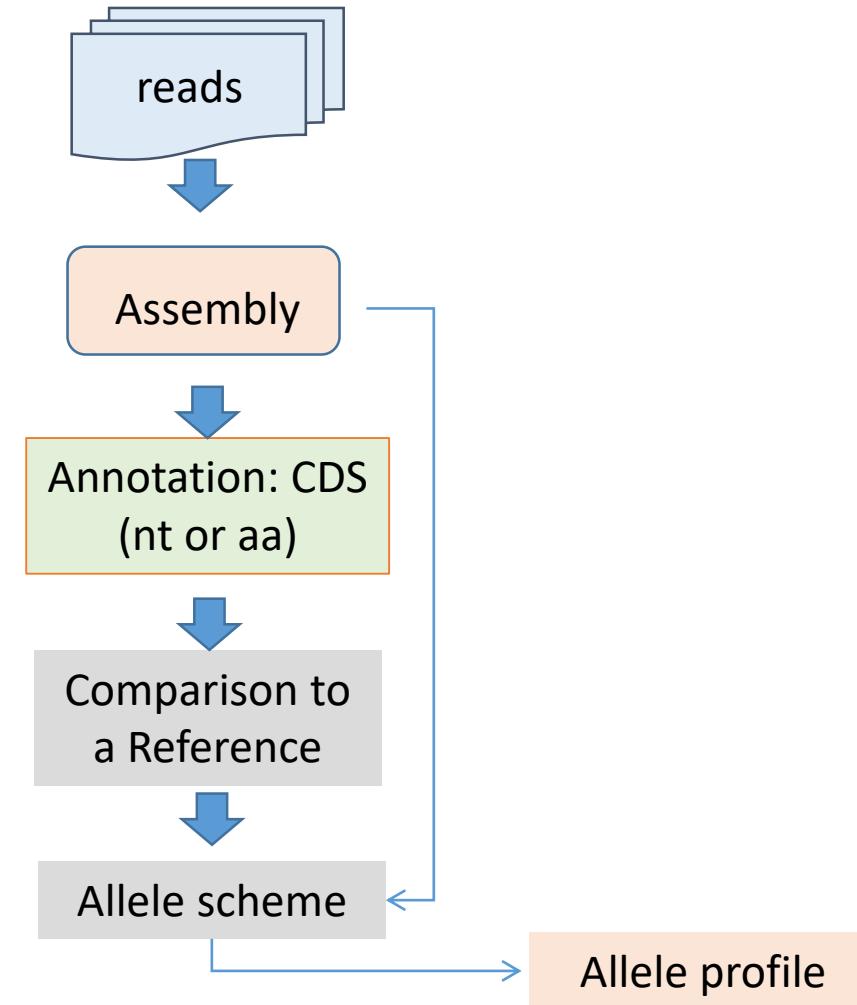
Pangenome: The term “pan-genome” refers to pan (from Greek πᾶν, whole) and genome (genome) referring to the inclusion of the core and the dispensable genome.

Phylogenetic approaches

Uelze et al., One Health Outlook 2020

| Method | Approach | Reference | Primary result | Secondary result |
|---------------------------------|---|---|---|---|
| cgMLST | Alignment to scheme of core genes | Set of allele sequences for set of core genes | Allele distance matrix | Minimum-spanning tree |
| wgMLST | Alignment to scheme of core and accessory genes | Set of allele sequences for set of core and accessory genes | Allele distance matrix | Minimum-spanning tree |
| SNP | Mapping to reference | Closely related reference genome | Core SNP alignment, SNP distance matrix | Neighbor-joining tree Maximum- likelihood tree |
| split K-mer based SNP detection | Pairwise K-mer comparison | No reference | Core SNP alignment, SNP distance matrix | Neighbor-joining tree |
| MinHash | Pairwise MinHash comparison and clustering | No reference | MinHash distances, clustering information | Neighbor-joining tree |

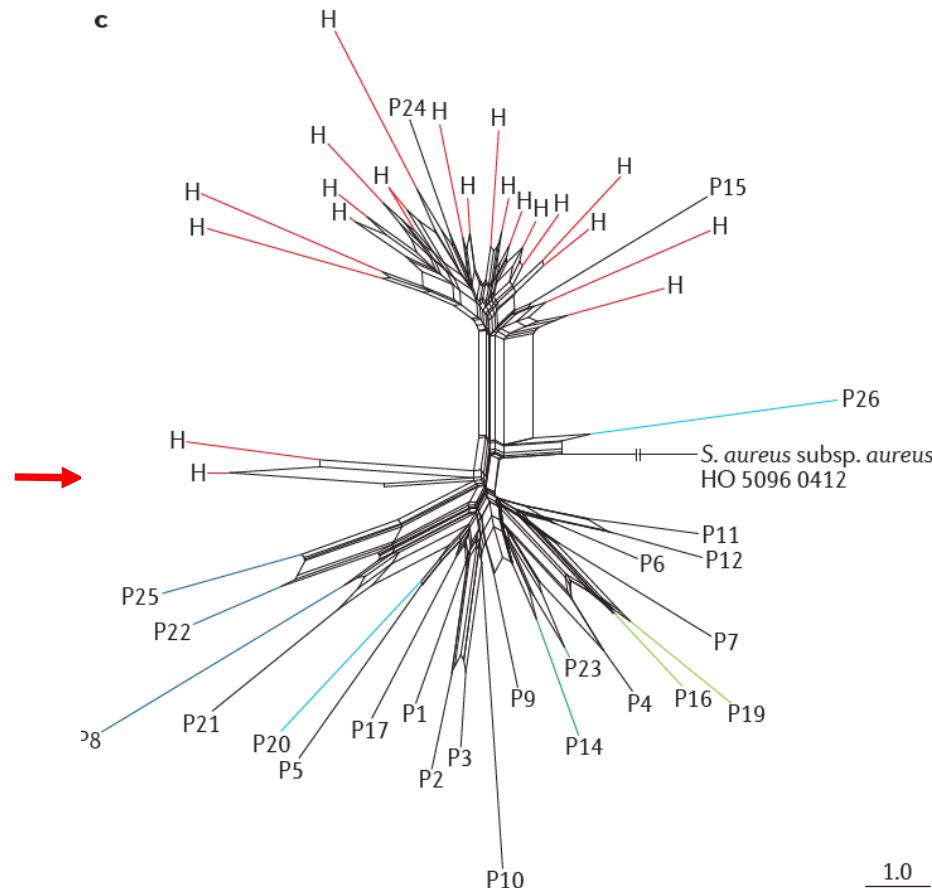
General analytical process for cgMLST / wgMLST



Gene-by-gene typing-based analysis of *Staphylococcus* spp. whole-genome sequence data

Maiden et al., Nat Rev Microbiology 2013

Resolution of multidrug-resistant *S. aureus* (MRSA) isolates from an outbreak in a special-care baby unit⁴¹, using a gene-by-gene comparison to a reference genome (*S. aureus* subsp. *aureus* HO 5096 0412). Twenty isolates obtained from a health care worker are indicated with the letter H and shown in red, whereas patient isolates are indicated with a letter P. Groups of isolates from patients who were members of the same family are shown in the same colour.



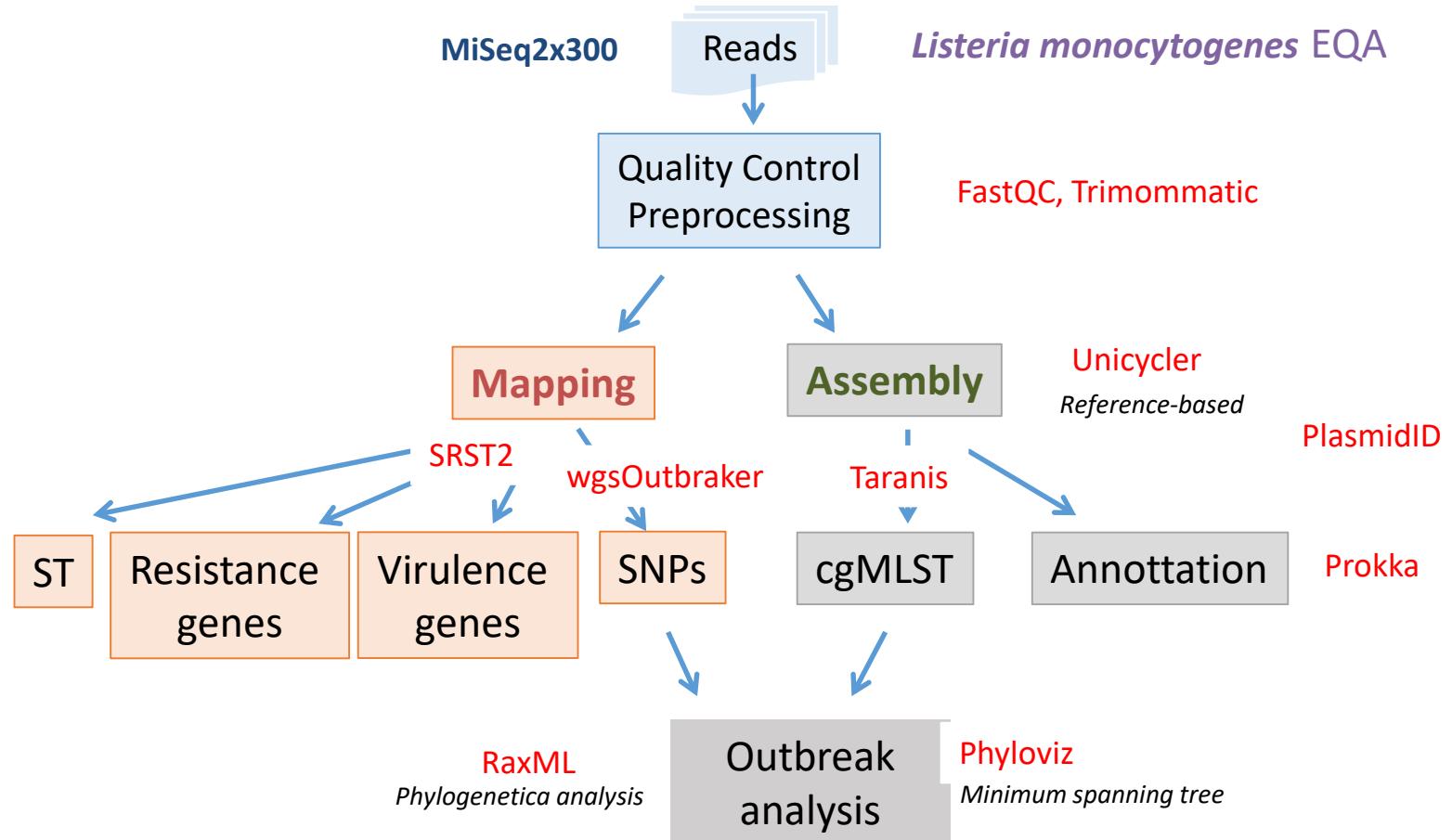
Index

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

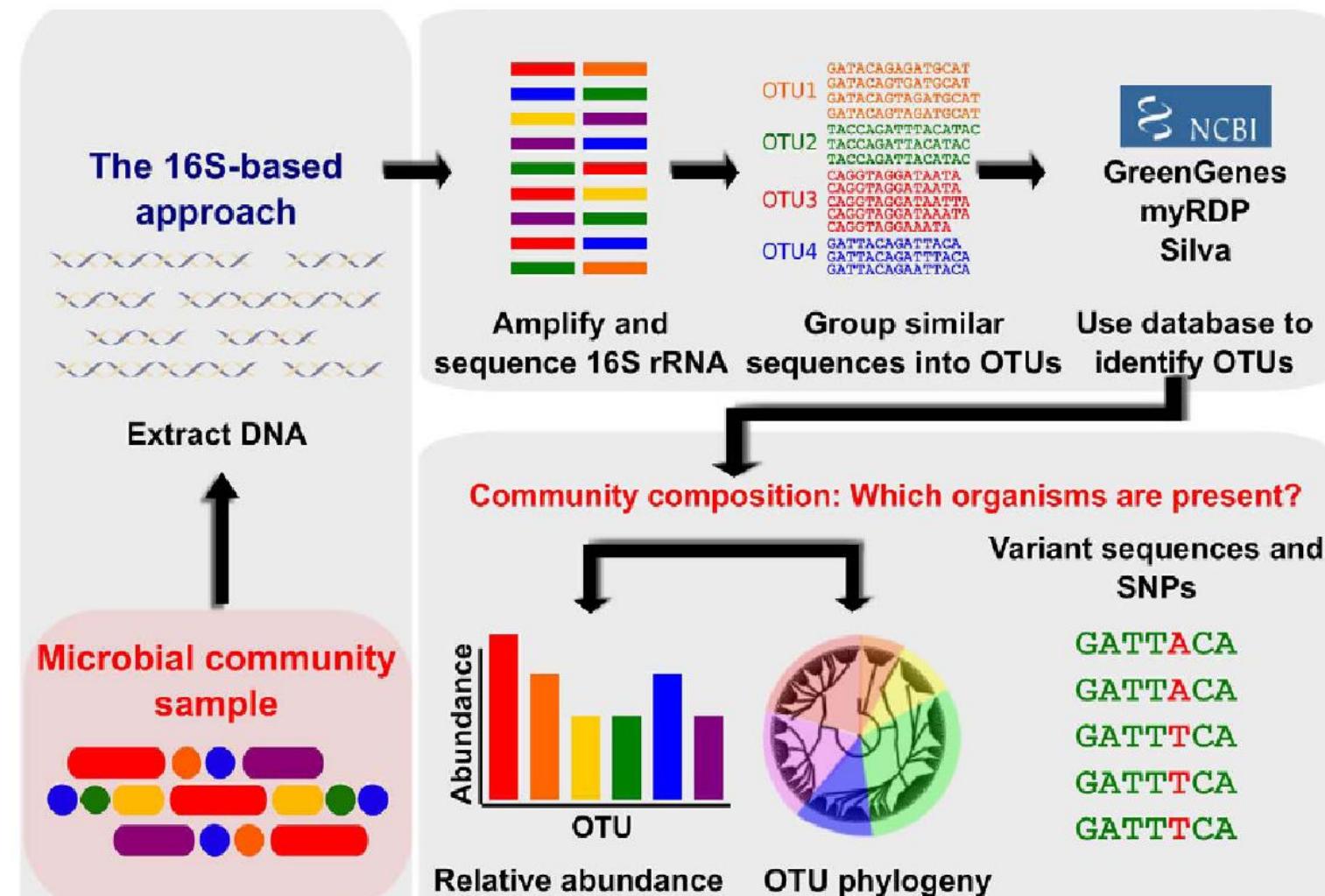
Workflow example



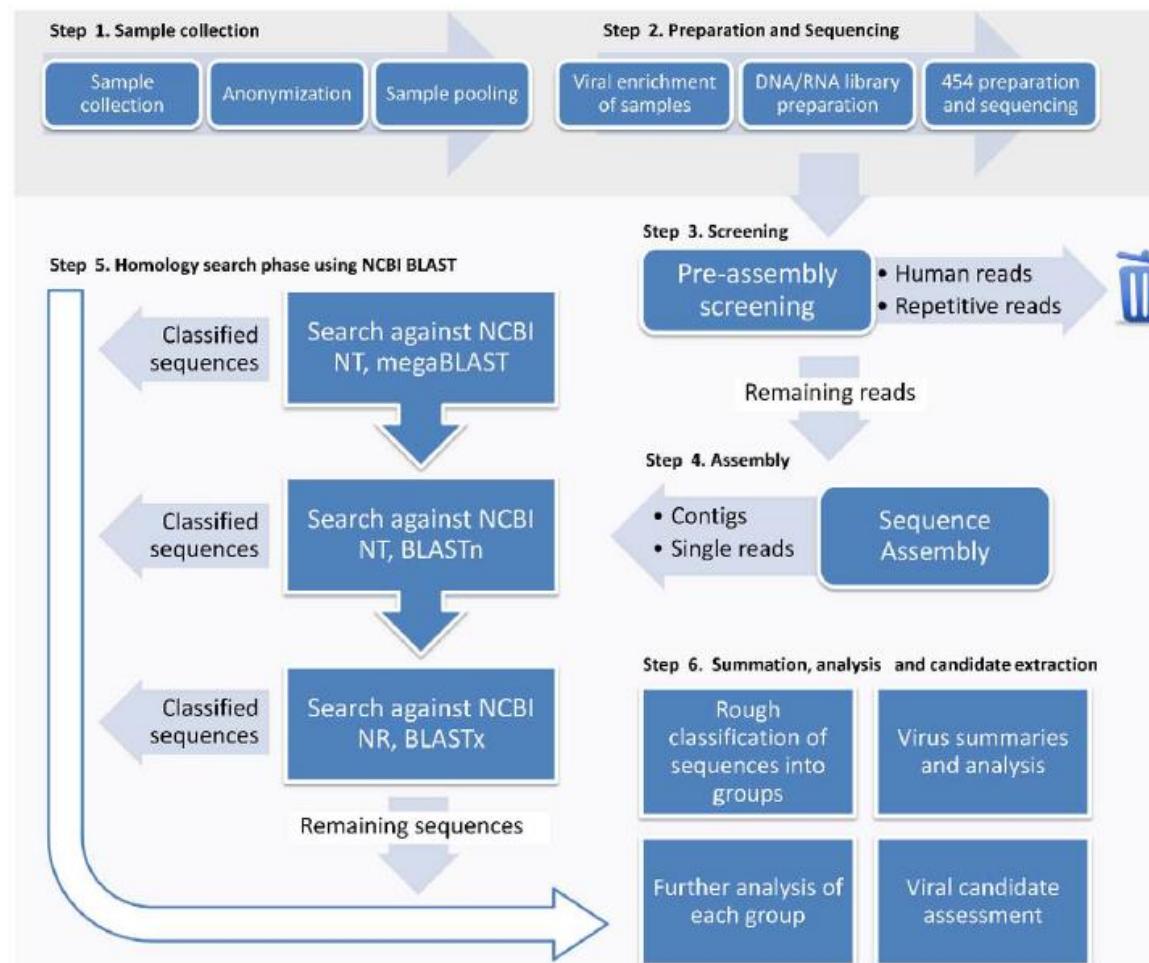
Metataxonomics vs Metagenomics (16S vs Shotgun)

| | 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

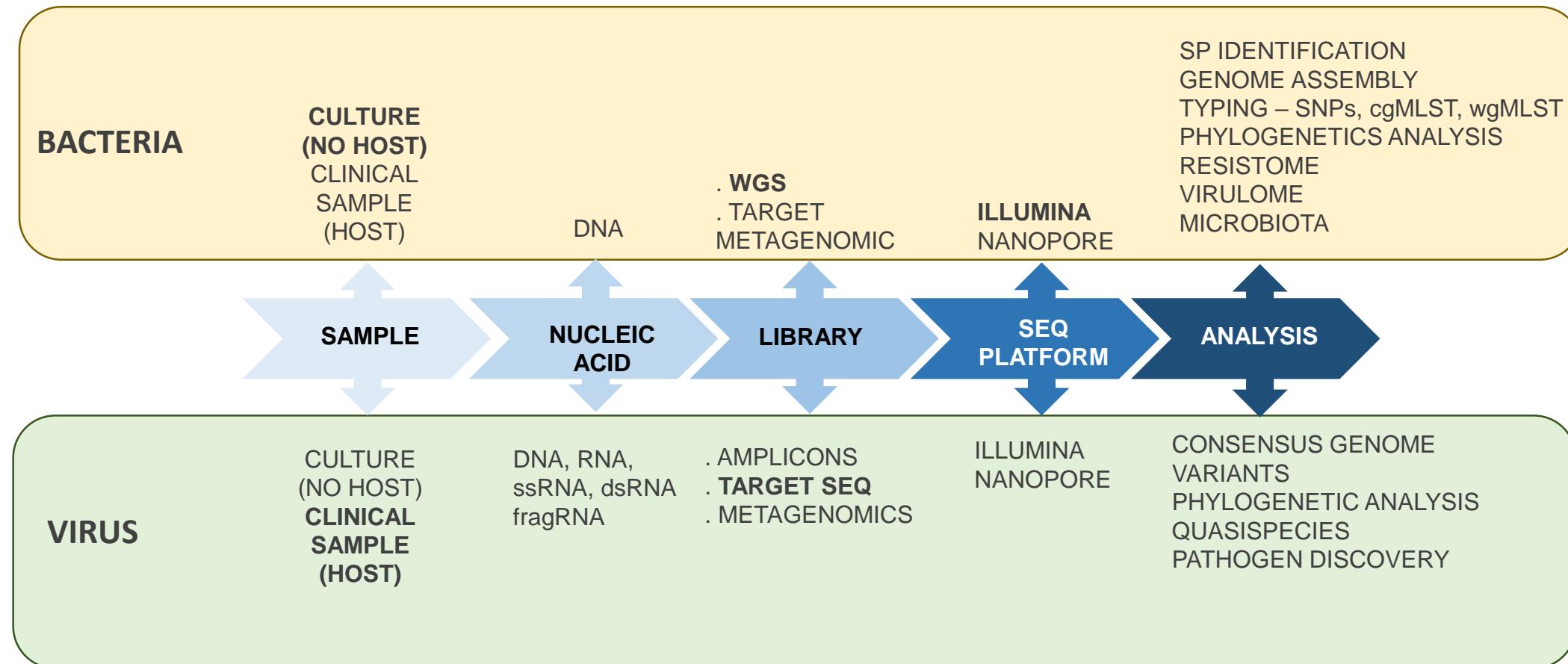


Lysholm et al., Plos One 2012:7,2, e30875

Index

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