



CENTRO NACIONAL DE ALIMENTACIÓN

Agencia Española de Seguridad Alimentaria

JORNADAS DE REFERENCIA

Majadahonda, 10 de Junio de 2026

Iria Uhía Castro

Sonia Aguayo Balsas

Unidad de Secuenciación (MAII) del CNA –AESAN OA



25 años cuidando de tu alimentación



¿Por qué secuenciación masiva en alimentos?



Seguridad alimentaria

Detectar patógenos emergentes y disponer de trazabilidad rápida y asistencia como LNR en la investigación de brotes de transmisión alimentaria

Reglamento de Ejecución (UE) 2025/179: la WGS debe realizarse en aislados de patógenos alimentarios asociados a brotes. Los Laboratorios Nacionales de Referencia y/u Oficiales deberán estar acreditados según EN ISO/IEC 17025:2017.

European Accreditation ha publicado la guía EA 4/24-G:2025 para la acreditación de laboratorios que realizan pruebas genéticas con NGS.



Resolución y alcance

Identificación a nivel de cepa

Análisis de resistencias, patogeneicidad, virulencia.

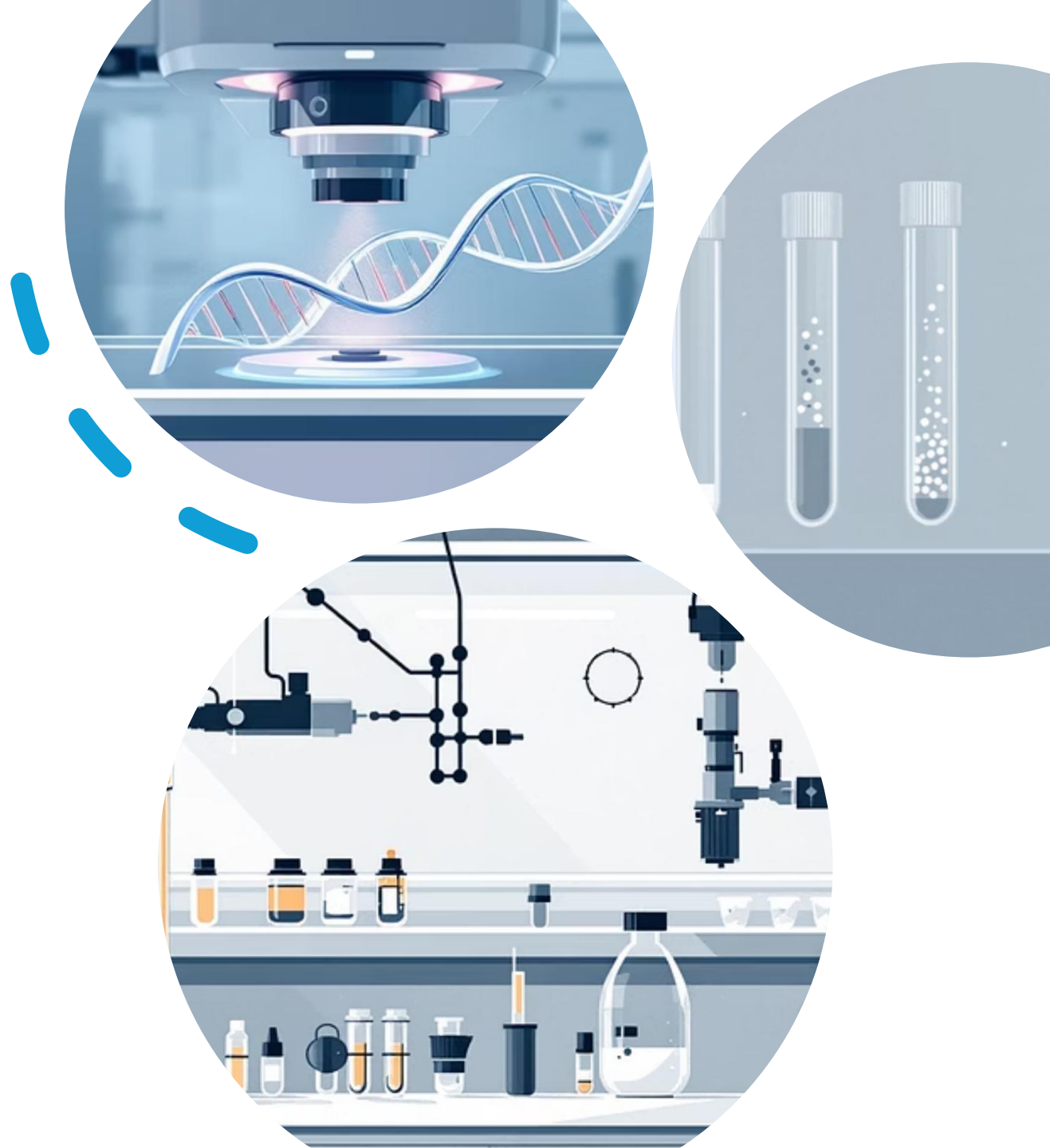


Eficiencia

Multiplexación y mayor rendimiento frente a técnicas convencionales.



- **Validación de la Secuenciación del Genoma Completo (WGS) para Acreditación**





Inter Biorisks-EURLs WG on NGS

Organization  Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail  and 5 more organizations

 Records

 Members

EURL *E. coli* (coordinator)
EURL *Listeria monocytogenes*
EURL CPS
EURL *Salmonella*
EURL *Campylobacter*
EURL Parasites
EURL AR
EURL Food borne viruses
Observers: SANTE G4, EFSA, ECDC

36 results found

Versions

☐ View all versions

Publication date

March 30, 2026 (1) Standard 

Guidelines for the validation of Whole Genome Sequencing for accreditation

Inter-EURLs Working Group on NGS

According to Commission Implementing Regulation (EU) 2025/179 (1), Whole Genome Sequencing (WGS) shall be performed on isolates of food-borne pathogens associated with food-borne outbreaks. For this, National Reference Laboratories and/or Official Laboratories shall be accredited for this method according to EN ISO/IEC 17025:2017 (1). To enable National Accreditation...

Part of [Inter Biorisks-EURLs WG on NGS](#)

Uploaded on March 30, 2026

 417  350

January 21, 2026 (2) Other 

Supporting document for preparing high quality DNA for Whole Genome Sequencing

Michelacci, Valeria

In the framework of the activities of the Inter EURLs Working Group on Next Generation Sequencing, the need for guidelines for checking the quality of extracted DNA to be used for sequencing was highlighted. This document is meant to provide guidance in this crucial preliminary step of the laboratory work to produce NGS data. It was agreed to start considering the three pathogens...

Part of [Inter Biorisks-EURLs WG on NGS](#)

Uploaded on January 22, 2026

2 more versions exist for this record

 1,511  576

December 9, 2025 (1) Standard 

Guidelines for Dry-Lab Quality Control Metrics and Acceptance Thresholds in Short-Read WGS of zoonotic and indicator bacteria

Sayeb, Maroua ; Asséré, Adrien 

Foreword The working group (WG) has been established by the European Commission with the aim to promote the use of NGS across the EURLs' networks, build NGS capacity within the European Union (EU) and ensure liaison with the work of the EURLs and the work of EFSA and ECDC on the NGS mandate sent by the Commission. The WG includes all the EURLs operating...

Part of [Inter Biorisks-EURLs WG on NGS](#)

Uploaded on December 15, 2025

 486  310



ISOLATING GENOMIC DNA FROM *LISTERIA MONOCYTOGENES*

LIBRARY PREPARATION USING THE NEXTERA XT KIT FOR ILLUMINA MISEQ SEQUENCING



Guidance document for WGS-laboratory procedures

Caccio, Simone Mario¹

Show affiliations

Contributors

Research group (13):

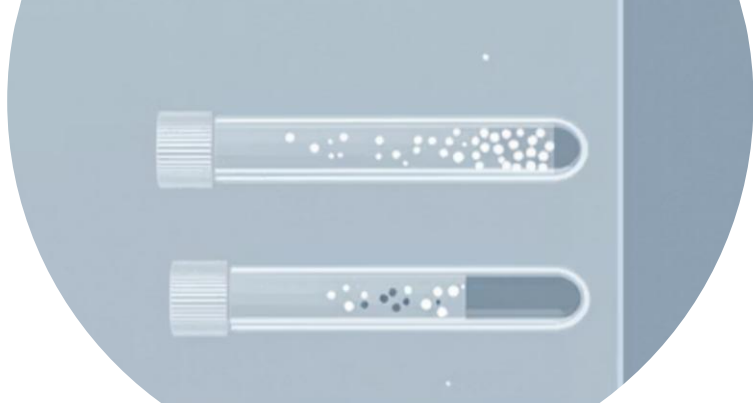
Michelacci, Valeria¹ ; Asséré, Adrien² ; Cavauiolo, Marina² ; Mooijman, Kirsten³ ; Morabito, Stefano¹ ;
Karlsense Pedersen, Susanne⁴ ; Sayeb, Maroua² ; Segerman, Bo⁵ ; Simonsson, Magnus⁶ ; Skarin, Hanna⁵ ;
Tozzoli, Rosangela¹ ; van Hoek, Angela³ ; Hendriksen, Rene Sjøgren⁴

Show affiliations

The use of whole genome sequencing (WGS) for diagnostics and surveillance of pathogenic microorganisms is becoming more and more widespread. Common applications include the identification of pathogens (bacteria, viruses, and parasites), investigation of antimicrobial resistance (AMR) determinants, virulence gene detection and investigations of outbreaks.

The availability of high quality DNA is a prerequisite for the successful application of WGS methodologies. In this regard, efficient DNA extraction is strongly influenced by the nature of the pathogen and by the possibility to use in vitro culture systems prior to it. In the case of uncultivable pathogens (e.g., most parasites), enrichment procedures are needed to purify the target organisms from the outnumbering number of unrelated organisms, including the host.

This guidance document provides a list of the available Standard Operating Procedures (SOPs) or Laboratory Operating Procedures (LOPs) for nucleic acid extraction, to validate the performances and suitability for the NRLs network. In addition, information was sought on procedures that can potentially reduce the time from culture to sequencing, using different library preparation and sequencing kits.



3. DNA quality check

Despite no specific quality criteria are generally required for input DNA by library preparation kits, assessing the quality of the DNA sample is advisable after extraction, to ensure its pureness and inspect fragmentation and to avoid problems in downstream protocols needed for WGS.

As a quality check, agarose gel or capillary electrophoresis can be performed, to inspect if fragmentation appears with smeared bands at low molecular weight and to visualize absence of evident rRNA bands.

Spectrophotometric analysis of the DNA can also be performed to assess the ratio between the absorbance at 260 nm and 280 nm. The value of the 260/280 absorbance ratio should within the range 1.75-2.05 which implies pure DNA (2).

4. DNA concentration

The concentration of the DNA samples must be measured to fulfil minimum criteria of the library preparation kits used for WGS. The DNA extraction kits available on the market are generally able to extract adequate quantity of DNA of the target organisms for the majority of the library preparation kits. The amount of input DNA can significantly vary between different library preparation kits and sequencing services. In detail, some library preparation protocols require only 1 ng of DNA, but others may require at least 100 ng (3). Additionally, when sequencing is outsourced to an external service provider, the shipment of up to 600 ng of DNA may be required.

The concentration should be measured for each DNA sample using a fluorometric system (e.g. the Qubit fluorometer) and the corresponding reagents kit, avoiding spectrophotometric methods measuring absorbance, which provide useful information in the quality check step but are less accurate for DNA

5. Bibliography

1. Commission Implementing Regulation (EU) 2025/179 of 31 January 2025 on the collection and transmission of molecular analytical data within the frame of epidemiological investigations of food-borne outbreaks in accordance with Directive 2003/99/EC of the European Parliament and of the Council, OJ L, 2025/179, 3.2.2025.
2. ISO 23418:2022 "Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of bacteria — General requirements and guidance". Technical Committee: ISO/TC 34/SC 9 Microbiology; Publication date: June 2022
3. Inter EURLs Working Group on NGS "Guidance document for WGS-laboratory procedures"
<https://doi.org/10.5281/zenodo.10406850>



Inter Biorisks-EURLs WG on NGS

Published January 21, 2026 | Version 2

Other

Open

Supporting document for preparing high quality DNA for Whole Genome Sequencing

Michelacci, Valeria (Contact person)¹

Show affiliations

Contributors

Research group (13):

Asséré, Adrien¹ ; Caccio, Simone Mario² ; Cavauiolo, Marina¹ ; Mooijman, Kirsten³ ; Morabito, Stefano² ; Karlsmose Pedersen, Susanne⁴ ; Sayeb, Maroua¹ ; Segerman, Bo⁵ ; Simonsson, Magnus⁶ ; Skarin, Hanna⁵ ; Tozzoli, Rosangela² ; van Hoek, Angela³ ; Hendriksen, Rene Sjøgren⁴

Show affiliations

In the framework of the activities of the Inter EURLs Working Group on Next Generation Sequencing, the need for **guidelines for checking the quality of extracted DNA to be used for sequencing** was highlighted. This document is meant to provide guidance in this crucial preliminary step of the laboratory work to produce NGS data. It was agreed to start considering the three pathogens *Listeria monocytogenes*, *Salmonella* and Shiga toxin-producing *E. coli* (STEC), on which the Joint EFSA-ECDC molecular typing data collection is currently focused, with the aim of expanding the topic to other pathogens in a second step.



Published March 30, 2026 | Version 1

Standard

Open

Guidelines for the validation of Whole Genome Sequencing for accreditation

Inter-EURLs Working Group on NGS

According to Commission Implementing Regulation (EU) 2025/179 (1), Whole Genome Sequencing (WGS) shall be performed on isolates of food-borne pathogens associated with food-borne outbreaks. For this, National Reference Laboratories and/or Official Laboratories shall be accredited for this method according to EN ISO/IEC 17025:2017 (1).

To enable National Accreditation Bodies to assess laboratories in a harmonised way, European Accreditation has published a Guideline for the Accreditation of Laboratories carrying out Genetic Testing using Next Generation Sequencing (NGS) Techniques (EA 4/24- G: 2025) (3).

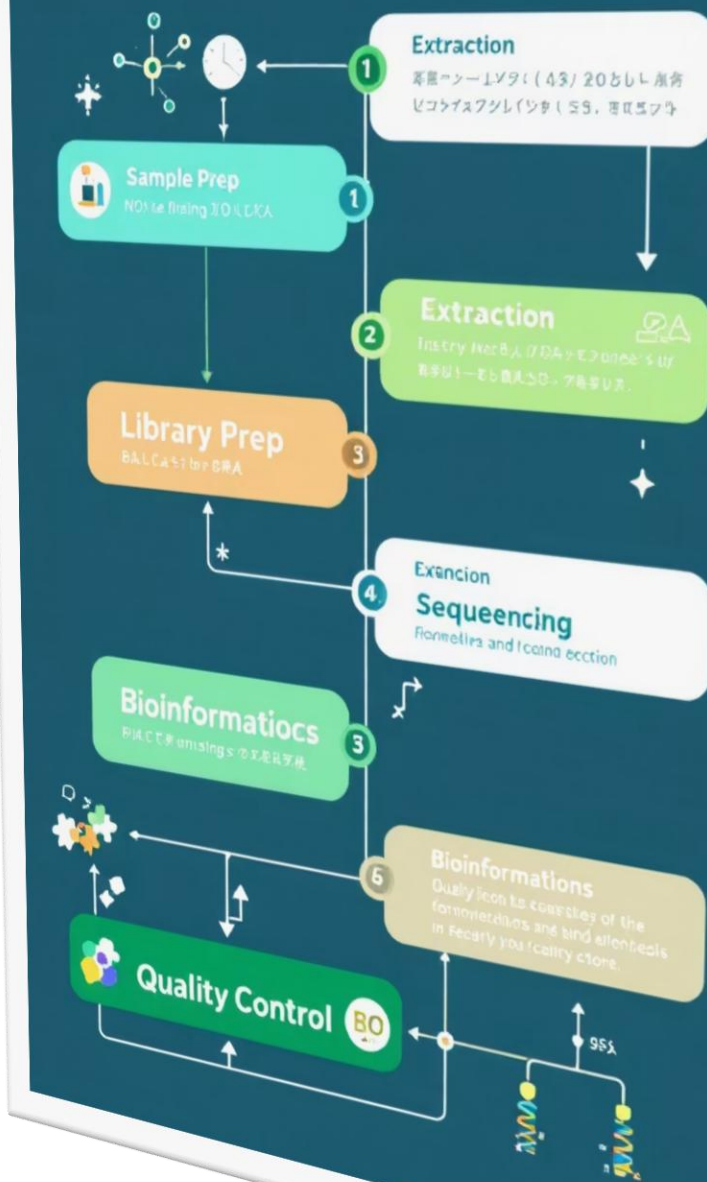
The present document produced by the Inter EURLs WG for NGS does not replace the procedures of the Quality Assurance Systems in use in the accredited laboratories, but it is meant to be used alongside, with the aim of giving reference to internationally recognized rules and procedures acknowledged by standards, where available, or other documents that can be used for the technical validation of the WGS workflow.

This document considers all steps of the workflow, including both wet lab and dry lab, but does not comprise the accreditation of the cluster analysis as this guidance document is meant to facilitate the accreditation of WGS workflow to respond to the Commission Implementing Regulation (EU) 2025/179 (produce data and send to EFSA) (1).

Este documento proporciona orientación para la acreditación de la WGS por partes (wet and dry lab) e indica los parámetros de validación bioinformática, incluyendo criterios de aceptación: *Campylobacter coli* y *C. jejuni*, *Escherichia coli* (STEC), *Listeria monocytogenes* y *Salmonella enterica*.

Sequencing lab Steps

for Queating Lab Diagram





EURL CPS
European Union Reference Laboratory for
Enteropathogenic *Escherichia coli* (EPEC)
<http://ecdc.europa.eu/en/eurl/cps>



EURL FV
European Union Reference Laboratory
Foodborne Viruses



EURL Lm
European Union Reference Laboratory for
Listeria monocytogenes
<http://ecdc.europa.eu/en/eurl/lm>



Especificaciones para la Acreditación del Flujo de Trabajo WGS

El flujo WGS comprende dos etapas principales: una parte de laboratorio «húmedo» (wet lab) para producir datos brutos de secuenciación, y una parte «seca» (dry lab) para obtener resultados interpretables mediante pipelines bioinformáticos. Cada etapa requiere procedimientos de validación distintos.

Etapas	Norma primaria	Norma secundaria	Referencia
Manipulación bacteriana	EN ISO 23418:2022	—	(4)
Extracción de ADN	EN ISO 23418:2022 (Tabla 1)	—	(4)
Preparación de librerías	EN ISO 23418:2022 (Tabla 1)	ISO 20397-2:2021	(4)(5)
Proceso de secuenciación	EN ISO 23418:2022	ISO 20397-2:2021	(4)(5)
Métricas de calidad de datos	ISO 20397-2:2021	—	(5)

Para información detallada, consultar la cláusula 10 (Tabla 4) de EN ISO 23418:2022.

Los EURL pueden proporcionar cepas para validación previa solicitud.

Validación del Wet Lab: Etapas y Parámetros

Los parámetros a determinar en el estudio de validación son: **repetibilidad** (replicados intra-ejecución), **reproducibilidad** (replicados inter-ejecución) y **acuerdo con otros métodos** (resultados correctos positivos y negativos sobre el total de muestras). Debe cubrirse un rango representativo de organismos diana.



1

Paso 1: Cultivo puro

Cultivos mismo día (operador 1) para repetibilidad; días distintos (operador 2) para reproducibilidad; cepas con características diferentes.

2

Paso 2: Extracción de ADN

Extracciones del mismo cultivo para repetibilidad; diferentes días y lotes de reactivos para reproducibilidad.

3

Paso 3: Secuenciación

Librerías preparadas mismo día y misma ejecución para repetibilidad; diferentes días, ejecuciones e instrumentos para reproducibilidad.

Inter-EURLs Working Group on NGS (Next Generation Sequencing)



Criterios de Aceptación para la Producción de Datos WGS

Los umbrales de control de calidad (QC) para cada especie diana se detallan en el Deliverable 10 del GT Inter-EURL. Los parámetros QC evaluables incluyen: profundidad de lectura mínima, puntuación de calidad de bases, tamaño del genoma, contenido GC, nivel de contaminación, porcentaje de loci cgMLST ausentes e identificación de dianas relevantes.

Repetibilidad

100% de concordancia entre replicados intra-ejecución (misma carrera).

Reproducibilidad

100% de concordancia entre replicados inter-ejecución (distintas carreras).

Acuerdo con otros métodos

90–95% mínimo de resultados correctos (positivos y negativos) frente al método de referencia (Sanger, qPCR, dPCR o ensayos interlaboratoriales).

⚠ Los datos de cepas ocasionales que no cumplan estos criterios deben descartarse y repetirse. Si el problema afecta a más cepas, el laboratorio debe realizar un **Análisis de Causa Raíz**.

Validación del Análisis Bioinformático

Conforme a la norma EN ISO 23418:2022, los parámetros de validación de las pipelines bioinformáticas son:

1 Repetibilidad	2 Reproducibilidad	3 Acuerdo con otros métodos
Resultados idénticos al repetir el análisis sobre la misma infraestructura, con la misma versión de software y parámetros. Fórmula: (resultados en acuerdo / total de análisis repetidos) × 100%.	Resultados consistentes al analizar el mismo conjunto de datos en un entorno computacional diferente (distinto ordenador, clúster o nodo), con la misma versión y parámetros.	Resultados comparables al analizar el mismo conjunto de datos con una pipeline diferente desarrollada para la misma aplicación.

Criterios de Aceptación para la Validación de la Pipeline Bioinformática

Los siguientes criterios se aplican a todos los tipos de datos (datos de caracterización + producción de cadenas cgMLST):

100%	100%	90-95%
Repetibilidad	Reproducibilidad	Acuerdo con otros métodos

Para pipelines comerciales, los datos de referencia pueden provenir del desarrollador; para pipelines internas, los EURL pueden proporcionar datos y/o cepas. La sensibilidad y especificidad deben determinarse para los resultados de caracterización de cepas. Los criterios de aceptación son específicos para cada especie (ver Anexo 1). Aunque el análisis de agrupamiento no está en el alcance, el laboratorio puede validarlo usando datos con relaciones evolutivas conocidas (p. ej., cinco genomas relacionados + cinco no relacionados) o comparando con otras pipelines del mismo alcance (p. ej., pipeline WGS de EFSA).

Anexo 1: Criterios de Aceptación por Patógeno

Criterios de sensibilidad y especificidad sugeridos por los EURL respectivos para cada patógeno cubierto por este documento.

Patógeno / Determinación	Sensibilidad	Especificidad
<i>Campylobacter</i> — Especie y ST (MLST)	100%	100%
<i>E. coli</i> (STEC) — Genes <i>stx1</i> , <i>stx2</i> , <i>eae</i>	100%	100%
<i>E. coli</i> — Genes de virulencia adicionales	>95%	100%
<i>E. coli</i> — Antígeno O, H, ST y subtipo <i>stx</i> *	>95% / >70%	100% / >95%
<i>Listeria monocytogenes</i> — Especie, Serotipo y ST	100%	100%
<i>Salmonella enterica</i> — Antígenos O y H	>95%	100%
<i>Salmonella enterica</i> — ST (MLST)	100%	100%

❑ * Para antígeno O, H, ST y subtipo *stx* en *E. coli*, la sensibilidad puede calcularse globalmente, mientras que la especificidad se determina para cada característica individual (p. ej., serogrupos o antígenos H individuales). La validación de la serotipificación WGS de *Salmonella* debe seguir ISO 16140-6:2019.



UNE-EN ISO 23418:2023. Microbiología de la cadena alimentaria. Secuenciación del genoma completo para el tipado y caracterización genómica de bacterias. Requisitos generales y orientación. (ISO 23418:2022)

Guía general de todo el proceso de secuenciación desde la preparación del laboratorio, la extracción de ácidos nucleicos y la preparación de librerías hasta el análisis bioinformático y registro de metadatos.



El desafío de la validación: ¿Qué nos pide la norma ISO?

Requisitos documentales

Protocolo
Criterios de aceptación
Registros
Trazabilidad.

Desempeño analítico

Sensibilidad
Especificidad
Precisión
Concordancia con
otros métodos

Control de calidad

Controles internos
Materiales de referencia
Verificación continua

Gestión de riesgos

Contaminación cruzada y errores instrumentales.

Interpretación práctica: traducir requisitos ISO a pruebas experimentales y criterios cuantificables.



Materiales de referencia y controles: base de la validación



Estándares certificados

Uso de controles

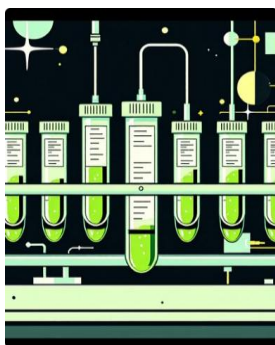
Cepas de referencia

para *Salmonella enterica*,
spp, *Listeria monocytogenes*,
E. Coli (STEC)

Campylobacter coli

Campylobacter jejuni

(EURL)



Controles negativos

Blanco de extracción

Control negativo del equipo

Tabla A.3. Recommended use of controls
ISO 23418:2022



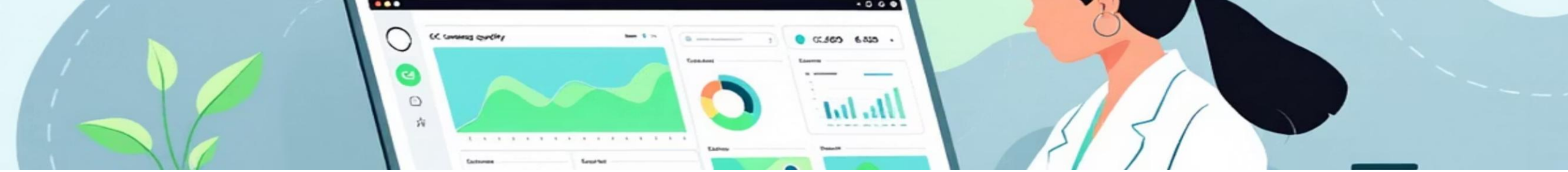
Registro riguroso de lotes, fechas y condiciones;
protocolos para almacenamiento.



Recomendaciones para la evaluación de la calidad de los datos de lectura corta de MiSeq antes del análisis bioinformático

(Tabla A.2. de la Norma UNE-EN ISO 23418).

Proceso	Problema potencial	Recomendaciones
Calidad de datos de secuencia	Datos de secuencia sin procesar de suficiente calidad, longitud de lectura y cobertura para el pronóstico previsto.	Las secuencias en formato FASTQ se pueden evaluar mediante la herramienta FastQC. Todas las secuencias deberían identificarse como de advertencia o compatibles con la calidad de la secuencia por base. La cobertura mínima estimada suele oscilar entre 20x y 60x.
Ejecutar parámetros de aceptación	Cobertura Q30 global	2 x 300 pb ≥ 70% 2 x 250 pb ≥ 75% 2 x 150 pb ≥ 80 %
	Lecturas que pasan el filtro	> 44 M
	Lectura del control negativo	< 10000
Parámetros de aceptación	Cobertura estimada (cobertura calculada en función del número de lecturas y el tamaño del genoma del organismo de objeto del estudio utilizando la ecuación de Lander/Waterman)	≥ 20X
	Media de Phred antes de recortar	≥ 30
	Contaminación	Se verifican las especies esperadas y la ausencia de especies/ cepas no esperadas (< 5% de lecturas identificadas como especie no objeto)
Criterios de la muestra	Puntuación GC	< 4% de desviación del contenido de GC esperado para las especies analizadas
	Puntuación de Phred Q30	≥ 150
	Lecturas por muestra	> 20 000



Conclusiones y preguntas

Estado actual

Metodo validado parcialmente_EFSA
System
Métricas clave cumplen criterios
preliminares;
Documentación en curso.

Impacto

Mayor capacidad de respuesta en
seguridad alimentaria y mejora en
trazabilidad de brotes.

Siguiente paso crítico

Completar validación por matrices
adicionales y preparar auditoría ISO.



ISO 20397-1:2022.

Biotechnology - Massively parallel sequencing. Part 1: Nucleic acid and library preparation.

Establece requisitos y guías para la evaluación de la calidad de ácidos nucleicos y para la preparación de librerías antes de la generación de datos por secuenciación masiva (MPS). Su objetivo es optimizar la calidad de los datos mediante control sistemático de las etapas previas a la secuenciación *per se*.



Alcance y referencias normativas

Alcance

Especifica requisitos generales y guía sobre evaluación de calidad de muestras de ácidos nucleicos y preparación/valoración de librerías para MPS.

Referencias

Incluye referencias normativas clave, especialmente ISO 20395:2019 para métodos de cuantificación de ácidos nucleicos (qPCR y dPCR).

Métodos para evaluar la integridad de la muestra

- ✓ ¿Cuál es el origen de la muestra de ADN (bacteria)? Relevante para anticipar los pasos de control de calidad
- ✓ ¿Qué método de aislamiento se ha utilizado para aislar la muestra de ADN? ¿dicho método elimina sales o detergentes?
- ✓ ¿Se ha utilizado un método fluorimétrico para cuantificar el ADN?
- ✓ ¿Dispone de suficiente cantidad de ADN para la preparación de la librería?
- ✓ ¿Se ha confirmado en un gel la calidad del ADN genómico?
- ✓ ¿Fue necesario añadir un paso de limpieza previo?
- ✓ ¿Cumple la muestra de ADN con los criterios específicos? ¿se ajustan las condiciones preanalíticas de la muestra al rango adecuado para el uso establecido?

ISO 20397-1:2022. Biotechnology

Massively parallel sequencing. Part 1: Nucleic acid and library preparation.

✓ ¿Dispone de suficiente cantidad de ADN para la preparación de la librería?

Table B.1 — Examples of quality metrics criteria for selected MPS platforms^{a,b}

Platform names	Sample amount	Sample concentration	Fragmentation size range	Barcoding length range	Adaptor length	Control
illumina® ¹ NextSeq & NovoSeq	10 µg	200 ng/µl	100 bp to 500 bp	6 bp to 10 bp	10 bp to 15 bp	—
Thermo Fisher Ion S5™, ²	10 µg	200 ng/µl	100 bp to 500 bp	6 bp to 10 bp	10 bp to 15 bp	—
MGI ³ DNASEQ-G400	10 µg	200 ng/µl	100 bp to 500 bp	6 bp to 10 bp	10 bp to 15 bp	—
Oxford Nanopore ⁴ PromethION® ⁵	1 µg	200 ng/µl	1 kbp to 2 Mbp	6 bp to 10 bp	—	—
PacBio® Sequel II ^{6,7}	3 µg	200 ng/µl	5 kbp to 30 kbp	6 bp to 10 bp	—	—

El laboratorio debes implmentar un documento y un flujo de trabajo para la cuantificación del ADNg que asegure la precision y reproducibilidad de los resultados. Los requerimientos de cantidad y Calidad dependen de la tecnología de secuenciación empleada.

Se debe contar igualmente con un protocolo de control de la Calidad para definir claramente la Calidad del ADNg y de la composición de la librería. Implica cálculo de incertidumbre y de sensibilidad.

Permite un ajuste apropiado de la concentración del AND apta para ser analizada en el secuenciador. Ver Tabla B.1

Table B.2 — Optimal sample amounts and concentrations by application^a

MPS type	Applications	Sample amount	Sample concentration
Whole genome sequencing	Homozygous single nucleotide variants (SNVs) – single nucleotide changes in genes where the alleles are identical	1 µg to 10 µg	20 ng/µl to 200 ng/µl
	Heterozygous SNVs – single nucleotide changes in genes where the alleles are different from each other		
	Insertion or deletion mutations (INDELS) – mutations in the genome where nucleotides are inserted or removed		
	Copy number variation (CNV) – variance in the number of copies of a gene between individuals		
	Large structural variants		

ISO 20397-1:2022. Biotechnology

Massively parallel sequencing. Part 1: Nucleic acid and library preparation.



agencia española de seguridad alimentaria y nutrición



Preparación de Librerías de Ácidos Nucleicos:

El laboratorio debe establecer, implementar y documentar cada procedimiento para la preparación de librerías de ácidos nucleicos que garantice resultados precisos y reproducibles.

La calidad de la librería está determinada por los siguientes pasos en el procedimiento, que incluyen, entre otros:

a) fragmentación: confirmar la pureza de la librería preparada, así como la revisión de la distribución de los tamaños de los fragmentos generados

b) adición de secuencias universales: adaptadores / indexación _ secuencia única diferenciable de todos los demás índices en esa librería de secuenciación de ADN.

c) selección de tamaño de los fragmentos: método basado en microesferas y verificado por electrophoresis (Bioanalyzer, TapeStation)

d) amplificación: Los sesgos de la polimerasa pueden introducir errores; esto se puede minimizar aumentando la cobertura de la secuencia o de las réplicas técnicas.

e) purificación y limpieza: para eliminar exceso de adaptadores o de otros contaminantes (microesferas, centrifugado, enzimas, electroforesis)

f) cuantificación de la librería:

g) cualificación de la librería:

El procedimiento para la construcción de librerías depende de la tecnología de secuenciación y el equipo elegido

ISO 20397-1:2022. Biotechnology

Massively parallel sequencing. Part 1: Nucleic acid and library preparation.





ISO 20397-2:2021.

Biotecnología:

Evaluación de la calidad de
datos de secuenciación
masiva paralela

Define requisitos y recomendaciones para la evaluación de la calidad de datos generados por secuenciación masiva. Cubre etapas desde los datos en bruto, preprocesamiento, alineamiento y mapeo, hasta la llamada de variantes, validación y documentación. No aplica al ensamblaje de novo.



Introducción y alcance

El documento enfatiza la necesidad de monitorizar métricas de calidad en todas las etapas del flujo bioinformático —datos en bruto, alineamiento y llamada de variantes— y proporciona consideraciones y recomendaciones específicas por plataforma.

Alcance

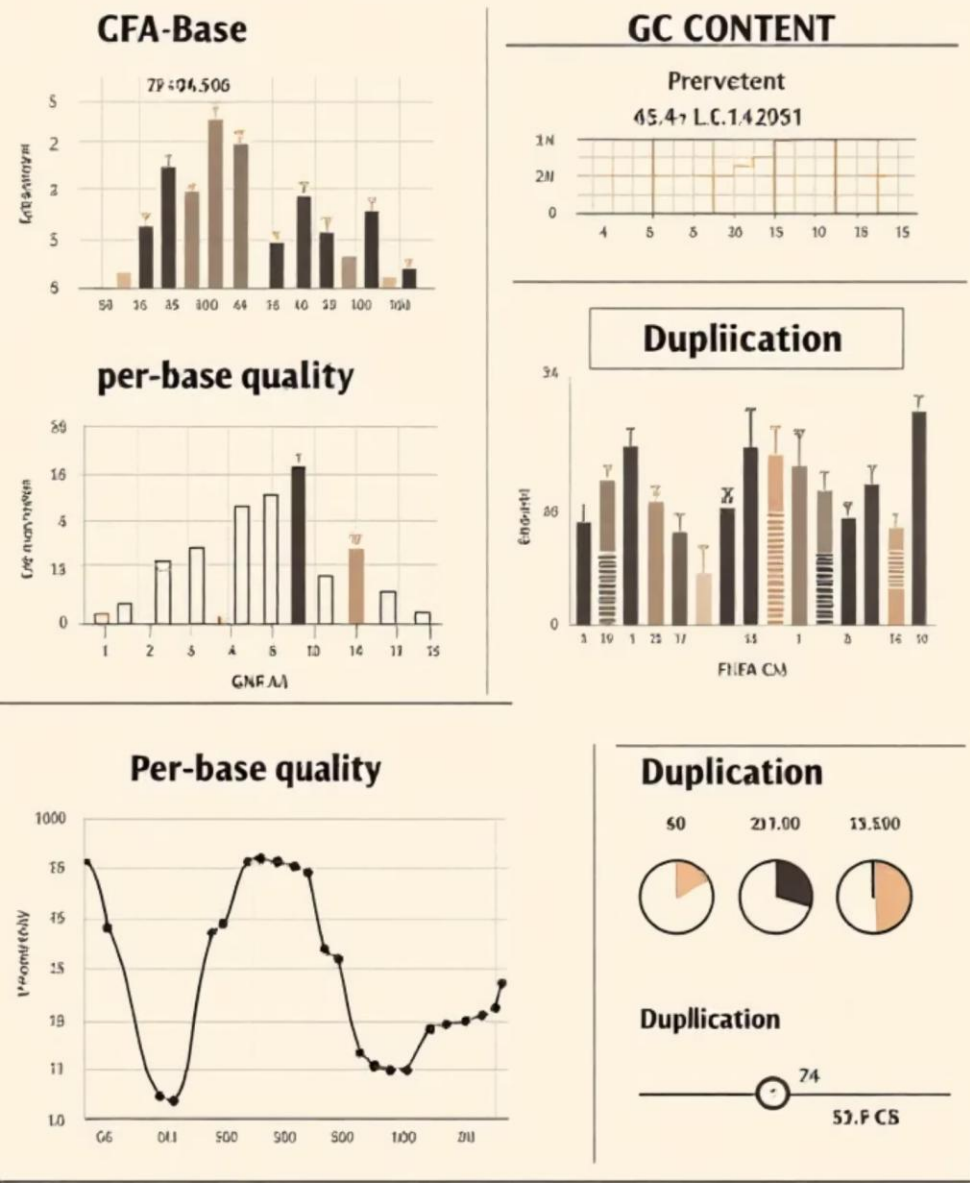
Requisitos generales y recomendaciones para evaluación y control de calidad de los datos; validación y documentación de procesos.

Exclusión

No incluye procedimientos relacionados con ensamblaje de novo.



Evaluación de la calidad de los datos en bruto — métricas y estadísticas básicas



Estadísticas básicas

Tipo de plataforma
tipo de lectura
kit
longitud
número de lecturas
GC
longitud total



Métricas de calidad

Per-base quality,
duplicación,
adaptadores,
error,
k-mer,
N-fragment,
homopolímeros.

Validación y Documentación — requisitos operativos y trazabilidad



Los laboratorios deben validar internamente la pipeline bioinformática, definir requisitos de rendimiento y parámetros de control utilizados para monitorizar continuamente el desempeño del ensayo. La validación debe establecer umbrales aceptables para métricas de calidad, métodos de preprocesamiento, sesgo GC, parámetros de mapeo, cobertura mínima y tasa máxima de duplicación. Cualquier cambio significativo en la pipeline requiere validación suplementaria.



Documentación obligatoria

Versiones de softwares, parámetros personalizados, versión de referencia, métricas QC y criterios de aceptación.

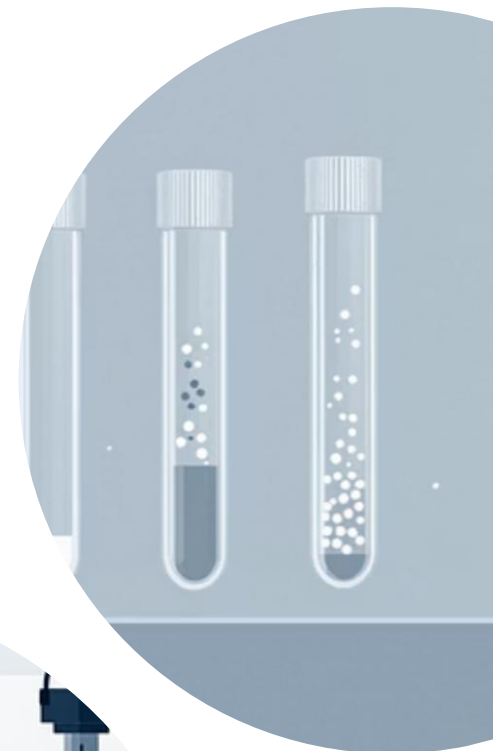


Integridad de archivos

Mecanismos para asegurar integridad y alertas frente a alteraciones no autorizadas.

- **Fin Parte 1**

De la teoría a la acción....



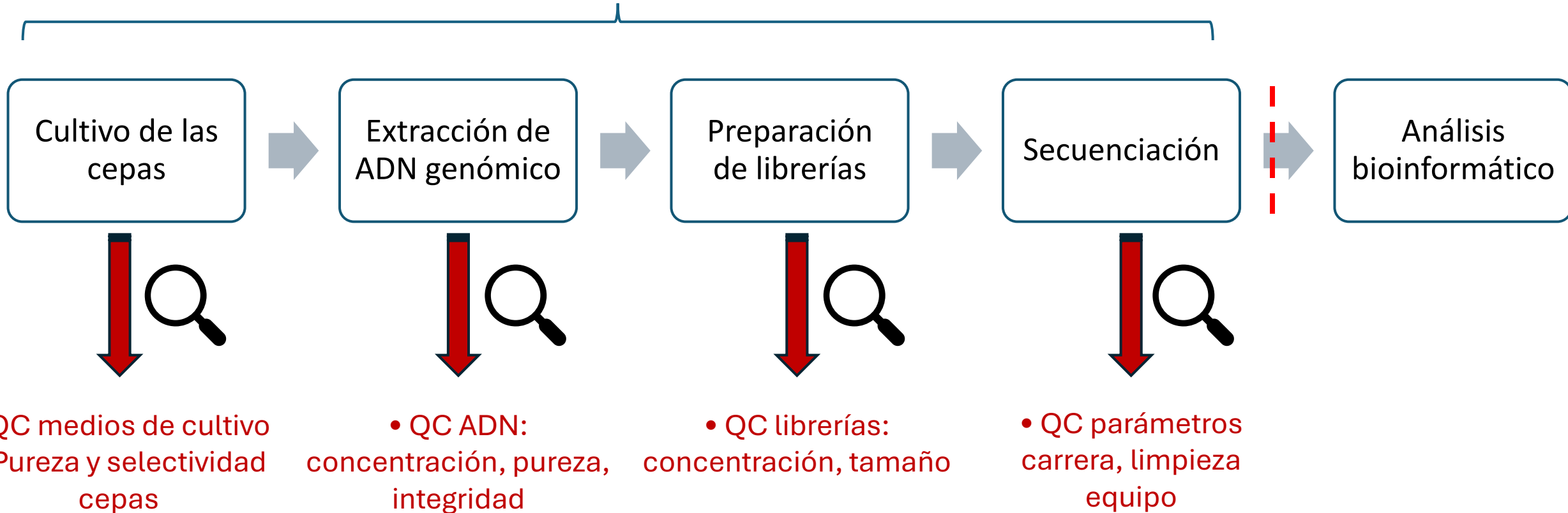


Validación del método de secuenciación masiva del genoma bacteriano: Parte húmeda

Iria Uhía Castro
Microbiología Alimentaria I-II
R.U. Secuenciación- CNA
Jornadas CNA, 9-10 junio 2026

Esquema de la secuenciación

Validación parte húmeda



Equipo de validación WGS-parte húmeda (CNA):



Raquel

Rosa

- ✓ Tres analistas realizan la técnica completa individualmente.
- ✓ Repetibilidad- cada analista.
- ✓ Reproducibilidad- entre analistas (x3).

Diseño de la validación WGS-parte húmeda (CNA):

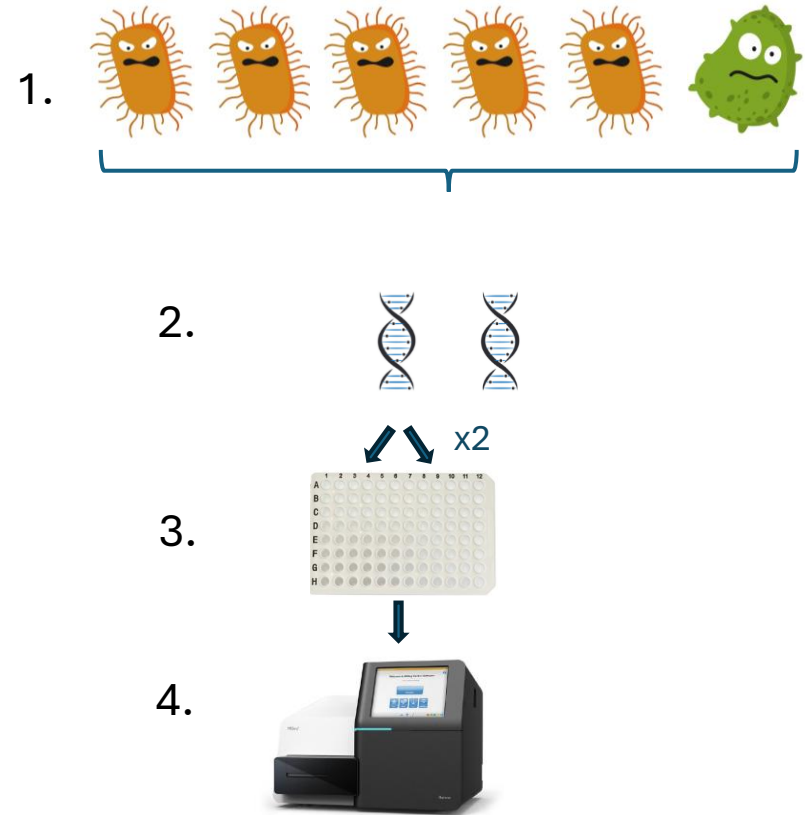
Cada una de las 3 analistas:

1. En total, **5 cepas** distintas del microorganismo a validar (preferiblemente de referencia/secuenciadas previamente por otra organización). Incluir ≥ 1 microorganismo no relacionado en cada carrera*.

2. Extracción de ADN de cada cepa por duplicado. Selección de uno de los duplicados para secuenciación.

3. Preparación de librerías de ADN de cada muestra por duplicado y *pool* de las librerías.

4. Secuenciación: Illumina MiSeq.



Cada analista realiza la validación en días distintos, y si es posible, con distintos lotes de reactivos.

Diseño de la validación WGS-parte húmeda (CNA):

ISO 23418:2022. “Concordancia con otros métodos”: Cepas relacionadas y no relacionadas

* Cepas de al menos 2 microorganismos con características diferentes. En el CNA:

- Gram +/-
- Genoma de distinto tamaño
- Distinto porcentaje de G+C

Microorganismo	Gram	Genoma (Mb)	% G+C
<i>E. coli</i>	-	≈ 5	≈ 50
<i>Campylobacter jejuni/coli</i>	-	≈ 1,7	≈ 30-32
<i>Salmonella enterica</i>	-	≈ 5	≈ 52
<i>Listeria monocytogenes</i>	+	≈ 3	≈ 38

Ejemplos:

- Para la validación de *E. coli*, incluir 1 cepa de *Listeria*
- Para la validación de *Campylobacter*, incluir 1 cepa de *E. coli* + 1 cepa de *Listeria*

Etapa 1. Cultivo

Procedimiento de cultivo (CNA):

- ✓ Crecimiento del microorganismo en agar selectivo + agar no selectivo.
- ✓ Pasar una colonia (varias en *Campylobacter*) del agar no selectivo a medio líquido e incubar.
- ✓ Utilizar 1 ml (x2) (10 ml x2 para *Campylobacter*) para la extracción de ADN y sembrar de nuevo en agar selectivo y no selectivo.
- ✓ Desechar el ADN si las placas aparecen contaminadas/no presentan las características fenotípicas adecuadas.



Control de medios (ISO 11133/ISOs específicas) y condiciones de incubación para cada microorganismo.

Etaapa 2. Extracción de ADN

✓ CNA: mismo kit para la extracción de ADN de todas las especies (Gram +/-): **Simplicidad y comodidad.**

Advertencias EURLs:

- Métodos basados en precipitación: posible pérdida de plásmidos.
- Evitar en lo posible el uso de EDTA.
- Evitar protocolos de extracción basados en hervido.
- **En general, cualquier protocolo comercial bien establecido permite obtener ADN de buena calidad para WGS – pero valorar cada caso.**



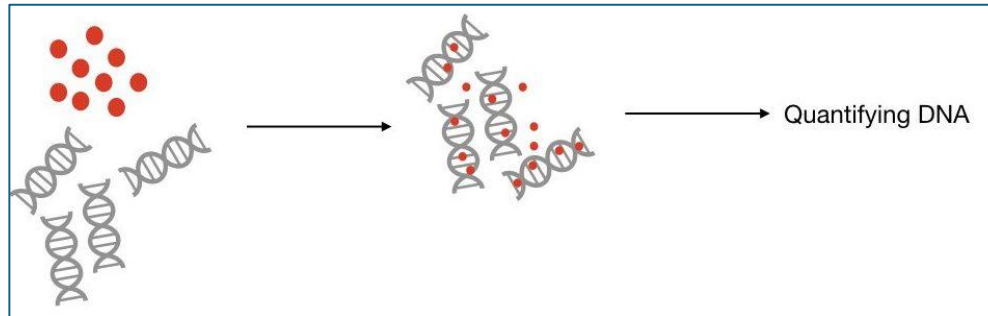
IMPORTANTE: el paso final de resuspensión del ADN debe realizarse preferiblemente en H₂O libre de nucleasas o tris HCl 10 mM pH 8.5. **No utilizar el tampón de elución de los kits** (posible presencia de concentración elevada de EDTA).



Etapas 2. Extracción de ADN

✓ Control de calidad del ADN:

1. Concentración: Medición en **fluorímetro**: específico para ADN de cadena doble y mucho más preciso a concentraciones bajas que la espectrofotometría.



Genetic Education, Reviews / By Dr.Tushar Chauhan / 11/10/2018



⚠ Los valores de concentración obtenidos en espectrofotómetro/nanodrop **NO son fiables** para ADN que se vaya a procesar para su secuenciación (miden también ADN de cadena simple, ARN...).

⚠ Los protocolos de preparación de librerías de ADN pueden requerir una concentración mínima de ADNg de partida (por ejemplo, Illumina DNA Prep $\approx \geq 4$ ng/ μ l para evitar la normalización).

Etapa 2. Extracción de ADN

✓ Control de calidad del ADN:

2. Pureza: Sí se utiliza espectrofotometría.

- Ratio A_{260}/A_{280} (contaminación proteica). Recomendado (ISO 23418): 1,75 – 2,05
- Ratio A_{260}/A_{230} (contaminación orgánicos). Recomendado (ISO 23418): 2,0-2,2



Algunos reactivos de conservación del ADN pueden alterar el ratio A_{260}/A_{230} sin que esto afecte a los pasos subsiguientes de procesamiento (valorar cada caso).

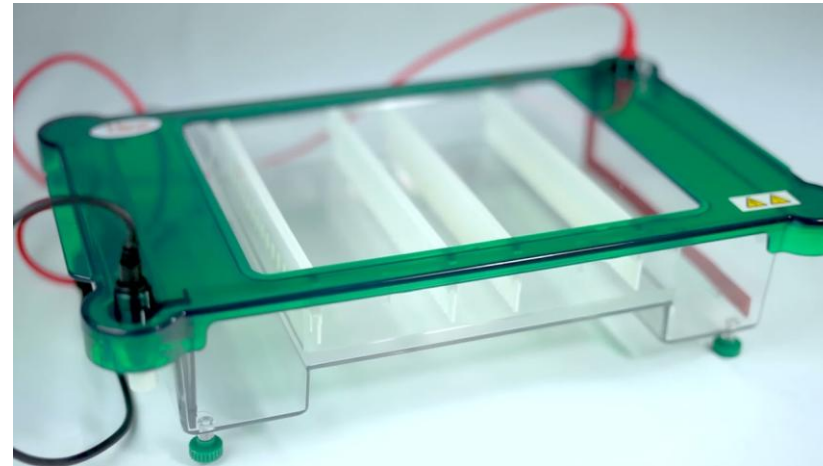
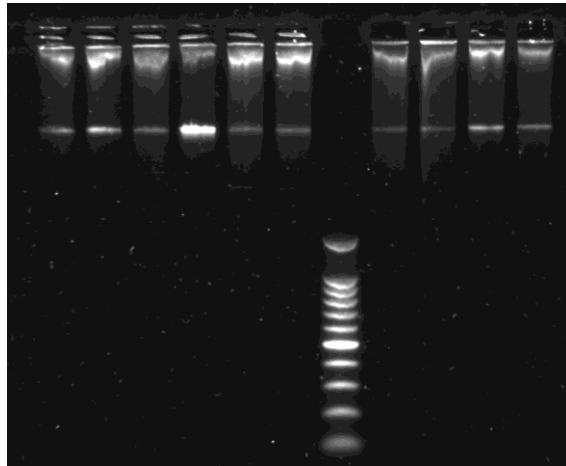


Concentraciones bajas de ADN hacen que estos ratios NO sean fiables (p. ej. por debajo de 10-20 ng/μL de acuerdo con ThermoFisher o NEB).

Etapas 2. Extracción de ADN

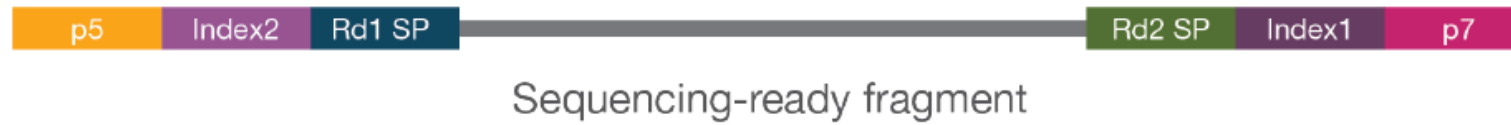
✓ Control de calidad del ADN:

3. Visualización de la integridad del ADNg mediante electroforesis en geles de agarosa, electroforesis capilar... **(opcional)**.



Etapa 3. Preparación de librerías

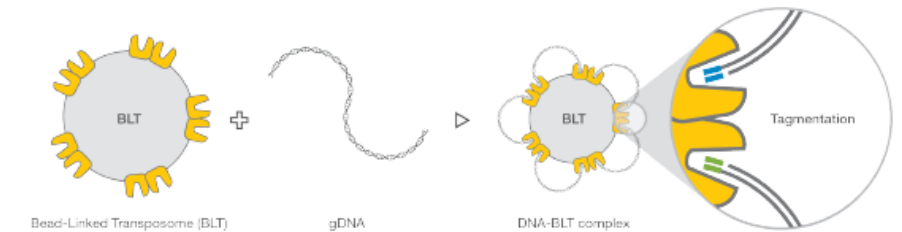
Librería de ADN para WGS: ADN (genómico) fragmentado y con secuencias adaptadoras en los extremos que permiten la unión a la plataforma de secuenciación, la lectura del fragmento y su posterior identificación.



✓ En el CNA: kit **Illumina DNA Prep**.

- Construye las librerías por **tagmentación**: Utiliza transposasas para fragmentar y etiquetar el ADN simultáneamente.

- Incluye un paso de amplificación mediante **PCR**, para incrementar la señal.



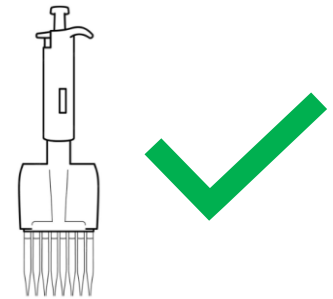
Recomendable que los pasos pre-PCR y post-PCR se lleven a cabo en estancias diferenciadas (ojo acreditación).

Etapla 3. Preparación de librerías

✓ Muestras por duplicado, incluyendo ADN de ≥ 1 cepa no relacionada y **controles negativos sin ADN** que pasen todas las etapas del protocolo.

✓ **Control positivo:** cepas de referencia para WGS/secuenciadas previamente en otras instituciones como EURLs, etc.

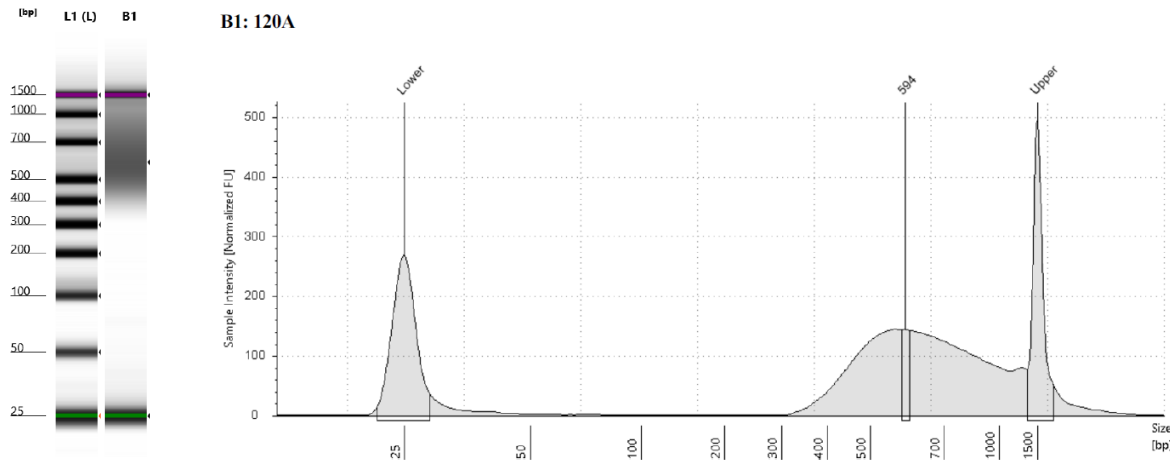
✓ Siempre que el protocolo lo permita, se recomienda el uso de **pipetas multicanal**.



✓ El tiempo de realización del protocolo completo es muy dependiente del número de muestras.

Etapa 3. Preparación de librerías

✓Al final de la preparación, calcular la **concentración de ADN (fluorímetro)** y el **tamaño de cada librería (electroforesis capilar; alrededor de 400-600 pb)**. Estos datos son necesarios para su posterior normalización.



Para el ajuste de las librerías a concentración 2 nM/4 nM (para secuenciación en plataforma MiSeq) se necesita una concentración mínima de cada librería de alrededor de 2 ng/μl (*aunque existen protocolos de ajuste para concentraciones más bajas*).

Etapa 3. Preparación de librerías

✓ **Normalización de las librerías y *pool*:** Convertir la concentración de cada librería de ng/μl a nM, ajustar a la concentración requerida de carga en el secuenciador (2 nM/ 4 nM en Miseq) y preparar un *pool* equimolar juntando todas las librerías. Herramientas **Nanomolar Conversion** y **Pooling Calculator** (Illumina).

illumina

Nanomolar Conversion

Number of Libraries:

Entry Method:

- ☐ Manual
- ☒ Different, import values
- ☐ Different, import values and names

Library Size (bp):

Description (Optional):

2.27
2.17
2.56
1.86

	Library Concentration (ng/μl)	Library Concentration (nM)
Library 1	2.27	7.01
Library 2	2.17	6.70
Library 3	2.56	7.90
Library 4	1.86	5.74
Library 5	2.78	8.58
Library 6	2.35	7.25
Library 7	7.53	23.25
Library 8	6.19	19.11
Library 9	3.56	10.99
Library 10	2.70	8.34

illumina

Pooling Calculator

Library Plexity:

Do the libraries have the same concentration?

- ☐ Same
- ☐ Different
- ☒ Different, import values
- ☐ Different, import values and names

Unit of Measure for Library:

- ☒ nM
- ☐ ng/μl

Pooled Library Concentration (nM):

Total Pooled Library Volume (μl):

Description (optional):

23.25
19.11
10.99
8.34

	Library Concentration (nM)	Library Volume (μl)	10 mM Tris-HCl, pH 8.5 (μl)	Pooling Volume (μl)
Total Pool Volume	200			
Library 1	7.01	11.4	28.6	20
Library 2	6.70	11.9	28.1	20
Library 3	7.90	10.1	29.9	20
Library 4	5.74	13.9	26.1	20
Library 5	8.58	9.3	30.7	20
Library 6	7.25	11	29	20
Library 7	23.25	3.4	36.6	20
Library 8	19.11	4.2	35.8	20
Library 9	10.99	7.3	32.7	20
Library 10	8.34	9.6	30.4	20

Etapa 3. Preparación de librerías



El paso de normalización previo al *pool* no es necesario en librerías preparadas con *Illumina DNA Prep* si se parte de muestras de entre 100-500 ng de ADN_g, pero es altamente recomendable para analistas no experimentados, y en el CNA lo hemos incluido en el protocolo de validación.

Etapa 3. Preparación de librerías

Las dos grandes preguntas:

1. ¿Cuántas muestras puedo incluir en un pool de librerías?



Depende...

- De la **profundidad de cobertura** (*depth of coverage*) buscada para las muestras.
- Del **tamaño de los genomas** a secuenciar.
- Del **kit de secuenciación** utilizado (Illumina).

$$C=N*L/G$$

N= N° lecturas

L= Longitud de lectura (Por ej. 150x2= 300)

G= Tamaño del genoma

- La profundidad de cobertura (**C**) en WGS es el número promedio de veces que se lee cada nucleótido (base) en una muestra.

- El número de lecturas (**N**) y su longitud (**L**) viene determinado por el kit de secuenciación utilizado.

Etapa 3. Preparación de librerías

1. ¿Cuántas muestras puedo incluir en un pool de librerías?

- Cada kit de secuenciación produce un número de lecturas máximo fijo, independientemente del número de muestras que carguemos.

Ejemplo: ¿Cuántas muestras de Salmonella puedo cargar con un kit v2 2 x 150 para que la cobertura sea 100x?

- Illumina Kit v2 2 x 150: 12-15 millones de lecturas. Longitud de lectura: 300.
- Genoma *Salmonella*: ≈5 Mb

Para una cobertura 100X: $100 = N \times 300 / 5.000.000$

$N = 1.666.667$ lecturas necesarias por muestra

15 millones de lecturas / 1.666.667 lecturas = 8 muestras de Salmonella

(El mismo kit permite cargar 26 muestras de Campylobacter (1,7 Mb) para una cobertura 100x)

$$C = N \times L / G$$

N= Nº lecturas

L= Longitud de lectura

G= Tamaño del genoma

Read length	Single reads ^c
MiSeq Reagent Kit v2	
2 × 25 bp	
2 × 150 bp	12-15M
2 × 250 bp	
MiSeq Reagent Kit v3	
2 × 75 bp	
2 × 300 bp	22-25M



Importante elegir un kit de secuenciación que se ajuste a nuestras necesidades analíticas.

Etapa 3. Preparación de librerías

Las dos grandes preguntas:

2. ¿Es necesario incluir los controles negativos en el pool de librerías?

- Si todo ha ido bien, los controles negativos no deberían contener ADN.
- Si los incluimos en el *pool* de librerías, diluirán al resto de muestras; habría que reajustar los cálculos...

→ Preguntamos a 3 EURLs diferentes. 1: no incluir. 2: sí (de distinta manera...).



- **Nuestra decisión para los ensayos de validación: incluir un volumen de cada control negativo sustituyendo a tampón de elución en algunas de las muestras.** De esta manera, no se produce una dilución extra.



Etapa 3. Preparación de librerías

2. ¿Es necesario incluir los controles negativos en el pool de librerías?

Ejemplo. Con estos valores obtenidos en el *Pooling calculator*:

Library Volume (µl)	10 mM Tris-HCl, pH 8.5 (µl)	Pooling Volume (µl)
11.4	28.6	20
11.9	28.1	20
10.1	29.9	20
13.9	26.1	20
9.3	30.7	20
11	29	20
3.4	36.6	20
4.2	35.8	20
7.3	32.7	20
9.6	30.4	20

Controles negativos añadidos sin que las muestras sufran una dilución adicional y sin reajustar los cálculos.



Elegimos un volumen similar al de las muestras (10 µl)

Sustituimos 10 µl de tampón por 10 µl de control negativo (réplica 1 y 2) en dos de las muestras

Etapa 4. Secuenciación

Previamente a la carga de la muestra en el cartucho:

✓ **Desnaturalización y dilución de las librerías normalizadas (*pool*):** Desnaturalizar con NaOH 0,2 N preparado en el momento desde un stock con pH comprobado. Dilución posterior según la concentración final de carga elegida:

- Kit MiSeq Reagent v3: 6–20 pM
 - Kit MiSeq Reagent v2: 6–10 pM
- **En el CNA: 12-13 pM v3 y 9-10 pM v2**

✓ **Dilución y desnaturalización de la librería control PhiX (Illumina):** Muy recomendable la adición de la librería PhiX como control del proceso de secuenciación.

- Kit MiSeq Reagent v3: Cargar la librería PhiX a 20 pM
 - Kit MiSeq Reagent v2: Cargar la librería PhiX a 12,5 pM
- **En el CNA (por ahora): misma concentración que el pool**



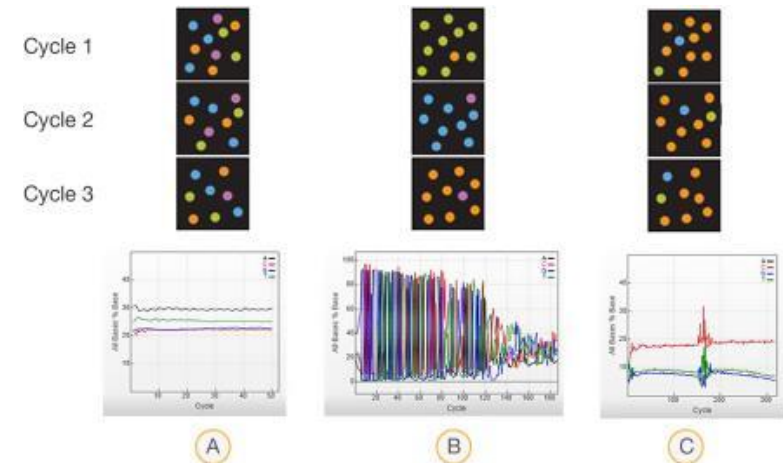
Etapa 4. Secuenciación

Previamente a la carga de la muestra en el cartucho:

✓ **Combinar nuestro pool de librerías con la librería control PhiX (Illumina):** Para la mayoría de librerías se añade un porcentaje bajo de PhiX (1 %) como control de secuenciación. Para librerías de baja diversidad o desequilibradas la concentración del control de PhiX debe ser de al menos un 5 %.

Las librerías de baja diversidad o desequilibradas (*low diversity/unbalanced libraries*) tienen una **proporción desigual de nucleótidos**. Las plataformas de secuenciación de Illumina requieren una representación equilibrada de las cuatro bases en cada ciclo para realizar una identificación precisa de las mismas.

Idea a mayores de PhiX: Incluir en la carrera una cepa de alta diversidad. Ejemplo: *Listeria* (38 % G+C) + *E. coli* (50 % G+C).



A. Equilibrada

B. Baja diversidad (% G+C $\neq$ 50)

C. Desequilibrada (una base a concentración mucho más baja)

Etaapa 4. Secuenciación

✓ Carga de la muestra en el cartucho del kit seleccionado.

Table 1: MiSeq System performance parameters

Read length	Total time ^a	Output	Quality scores ^b	Single reads ^c
MiSeq Reagent Kit v2				
2 × 25 bp	~5.5 hours	750-850 Mb	> 90% bases higher than Q30	12-15M
2 × 150 bp	~24 hours	4.5-5.1 Gb	> 80% bases higher than Q30	
2 × 250 bp	~39 hours	7.5-8.5 Gb	> 75% bases higher than Q30	
MiSeq Reagent Kit v3				
2 × 75 bp	~21 hours	3.3-3.8 Gb	> 85% bases higher than Q30	22-25M
2 × 300 bp	~56 hours	13.2-15 Gb	> 70% bases higher than Q30	

1

2

3

4

1. A mayor tamaño de lectura más cobertura del genoma, pero será necesario el recorte de las secuencias previo a su análisis (*trimming*).

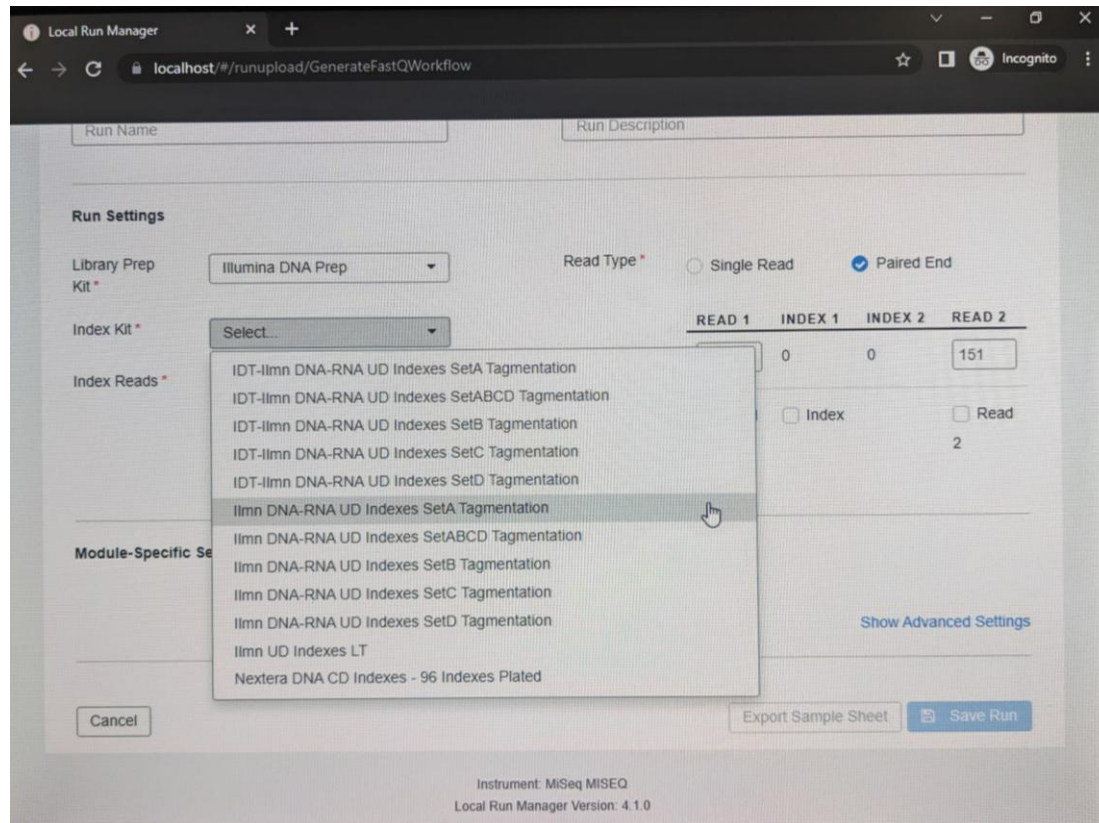
2. Tiempo de carrera.

3. Tamaño de los datos generados.

4. Métricas de calidad.

Etapa 4. Secuenciación

✓ Programar la carrera en el secuenciador Miseq: en el CNA usamos **Local Run Manager**



- ¡Cuidado! Seleccionar el **set de índices correcto**.
- Si se lanza una carrera señalando los índices incorrectos, se pueden reasignar *a posteriori* en *Local Run Manager*.

Edit Setup → cambiar los parámetros → **Requeue Analysis**

Etapa 4. Secuenciación

✓ Programar la carrera en el secuenciador Miseq: en el CNA usamos **Local Run Manager**

The screenshot shows the 'Module-Specific Settings' window in the Local Run Manager. The 'Adapter Trimming' toggle is set to 'On'. Below this, there is a 'Show Index Sequence' button. A table displays sample information with columns: SAMPLE ID, SAMPLE DESCRIPTION, INDEX WELL, INDEX 1 (I7), INDEX 2 (I5), and SAMPLE PROJECT. Row 1 shows sample 1 with index wells A02, UDP0009, and UDP0009. A dropdown menu for 'INDEX PLATE WELL' is open, showing options A01 through A08. At the bottom, there is a '+ 7 Rows' button, a 'Cancel' button, and 'Export Sample Sheet' and 'Save Run' buttons.

SAMPLE ID	SAMPLE DESCRIPTION	INDEX WELL	INDEX 1 (I7)	INDEX 2 (I5)	SAMPLE PROJECT
1		A02	UDP0009	UDP0009	
2					
3					
4					
5					
6					
7					
8					

- Incluir un **control de equipo**: una combinación de índices que no corresponde a ninguna muestra en la carrera actual, pero sí coincide con una de las combinaciones de índices utilizadas en la secuenciación anterior (*NGS EURL WG-Guidance document for WGS-Benchmarking*).

Si este control negativo genera lecturas, esto indica la posibilidad de contaminación con la librería de la secuenciación anterior.



En caso de contaminación (MiSeq) lavar con lejía al 0,01 % y todas las superficies de trabajo con lejía al 10 %.

Etaapa 4. Secuenciación

✓ **Carga del cartucho en el equipo.** Seguir paso a paso las instrucciones en la pantalla.



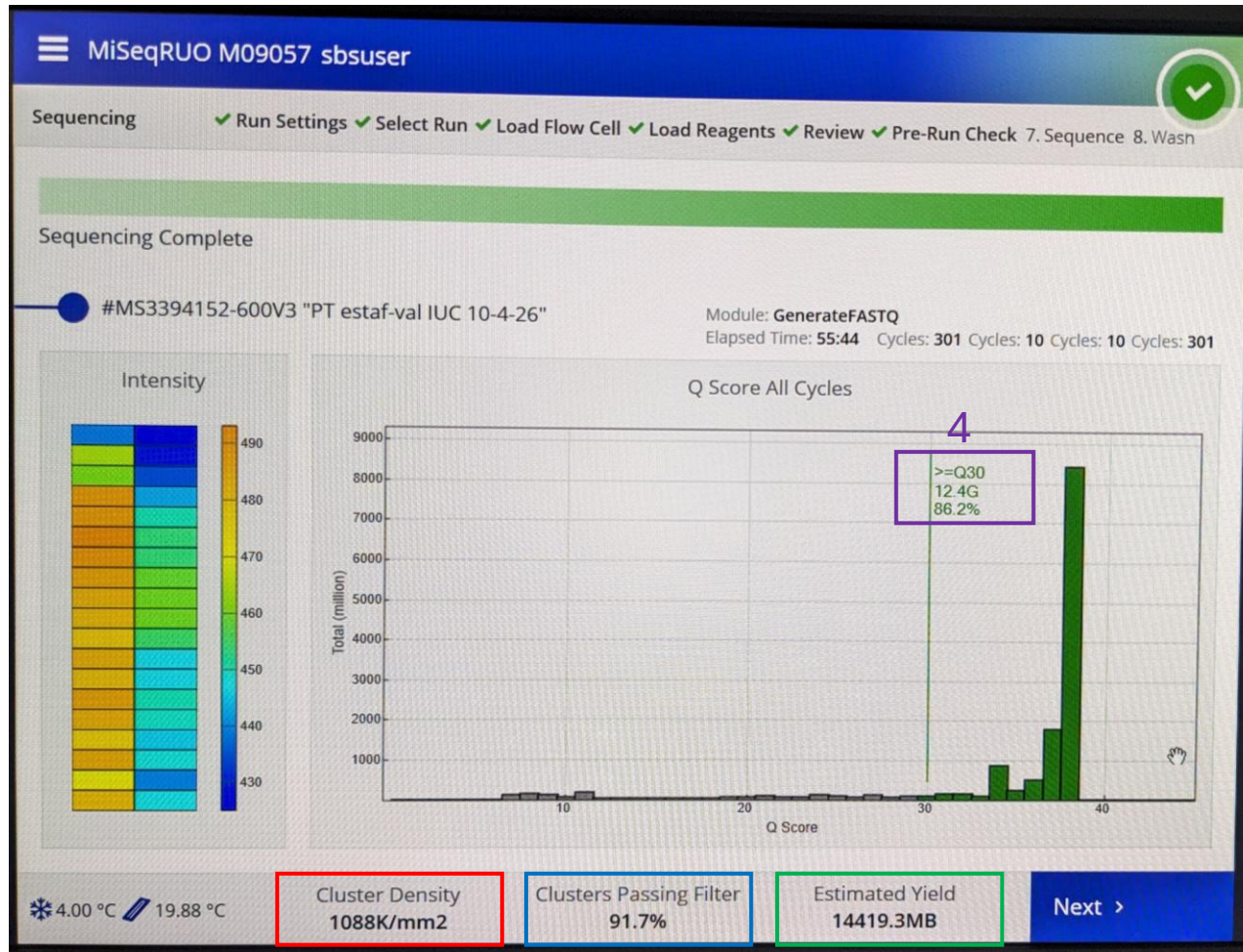
Presionar *Start* y...¡a esperar!



Mantener el equipo durante la carrera a la temperatura adecuada y **protegido de vibraciones.**

Etapa 4. Secuenciación

✓ **Fin de carrera:** comprobar las métricas de la carrera.



1

2

3

1. **Densidad de cluster óptima** (MiSeq): Dependiente del kit (v2: 1000-1200 K/mm²; V3: 1200-1400 K/mm²). Con librerías de baja diversidad, buscar una densidad un 30-40% menor (empíricamente).

2. **Clusters passing filter:** valor óptimo ≥80%.

3. **Estimated yield:** tamaño de los datos generados (según kit).

4. **% bases ≥ Q30** (según kit).

✓ **Secuencias (fastq) listas para su análisis, previo control de calidad en una plataforma bioinformática adecuada.**

Etapa 4. Secuenciación

✓ **QC de los archivos fastq:** Para esta validación, en el CNA usamos Galaxy Europe y/o Galaxy ARIES.

Galaxy Europe

291: Reads quality report ok

1.1 MB
formato html base de datos ? size 1.1 MB

[Preview](#) [Visualize](#) [Details](#) [Edit](#)

MultiQC
v1.11

General Stats

fastp

Filtered Reads

Insert Sizes

Sequence Quality

GC Content

N content

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2026-06-01, 09:30 based on data in:
/data/jwd06/main/000/103/760/103760322/working/multiqc_WDir

Welcome! Not sure where to start? [Watch a tutorial video](#) (6:06) [don't show again](#)

General Statistics

[Copy table](#) [Configure Columns](#) [Plot](#) Showing 2 rows and 4 columns.

Sample Name	% Duplication	GC content	% PF	% Adapter
SA01-A_S11_L001_001	1.5%	32.7%	97.8%	4.5%
SA01-B_S12_L001_001	1.1%	32.6%	99.0%	3.2%

Galaxy ARIES

FastQC Report

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

Basic Statistics

Measure	Value
Filename	173a_s1_l001_r1_001_fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	986079
Sequences flagged as poor quality	0
Sequence length	35-151
%GC	38

Per base sequence quality

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

Produced by **FastQC** (version 0.11.8)



Dudas/sugerencias:
iuhia@aesan.gob.es
rpeces@aesan.gob.es

Muchas gracias y... ¡buena suerte!

Unidad de secuenciación, Microbiología Alimentaria II

Raquel Cobo Gómez
Sonia González Crespos
Carmen González Sancho
Rosa Peces Pérez
Iria Uhía Castro
Sonia Aguayo Balsas

