



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

EVALUACIÓN DEL EFECTO DE LA  
TEMPERATURA DE  
INACTIVACIÓN, EXTRACCIÓN Y  
FRAGMENTACIÓN DEL ADN<sub>g</sub> DE  
*Mycobacterium tuberculosis* EN EL  
SECUENCIAMIENTO USANDO LA  
TECNOLOGÍA OXFORD  
NANOPORE

TESIS PARA OPTAR EL GRADO DE  
MAESTRA EN MICROBIOLOGÍA

CANDY LIZBETH LEÓN PALOMINO

LIMA - PERÚ

2025



**ASESOR**

DR. PATRICIA SHEEN CORTAVARRIA

**CO ASESOR**

DR. MIRKO JUAN ZIMIC PERALTA

**JURADO DE TESIS**

DRA. MONICA JEHNNY PAJUELO TRAVEZAÑO

PRESIDENTE

DR. HOLGER MAITA MALPARTIDA

VOCAL

DRA. NANCY CHILE ANDRADE

SECRETARIO (A)

## **DEDICATORIA.**

El *junzi* trabaja en la raíz: una vez plantada la raíz, nace el *dao*.

*The Analects of Confucius*

## **AGRADECIMIENTOS**

Esta tesis es el resultado de un esfuerzo colectivo y del apoyo de muchas personas. En primer lugar, agradezco a la Dra. Patricia Sheen, mi asesora, cuya orientación, exigencia y respaldo fueron determinantes para el avance de este proyecto y para mi formación profesional y personal. De igual modo, expreso mi más profundo agradecimiento al Dr. Mirko Zimic, quién con su apoyo y empuje me ayudaron y motivaron a la continuación de este trabajo, y al Dr. Louis Grandjean, cuyo empuje hizo posible su desarrollo. Muy agradecida con los tres doctores por haberme permitido y apoyado con el desarrollo del presente trabajo.

Un agradecimiento especial al equipo del Laboratorio de Bioinformática y Biología Molecular y del Laboratorio de Mycobacterium, en especial al Dr. Jorge Coronel y Walter por sus aportes para el desarrollo del proyecto, a Stefany, Jhojailith, Angela, María de Grecia, Adiana y María del Carmen por sus sugerencias y apoyo en el crecimiento dentro del laboratorio. Un agradecimiento especial a Diego Taquiri, Omar Romero y José Perez por su apoyo y soporte durante el desarrollo del trabajo. Un profundo agradecimiento a Alice Osmaston por su arduo apoyo, por su compañía y soporte y constancia durante el desarrollo de esta tesis.

Finalmente a mi familia (padres, hermanas, compañero y peluditos) por su compañía y apoyo constante en mi crecimiento.

## **FUENTES DE FINANCIAMIENTO**

Tesis Autofinanciada

## DECLARACIÓN DE ORIGINALIDAD

Los egresados:

N°	APELLIDOS Y NOMBRES
1.	LEÓN PALOMINO CANDY LIZBETH

*(Agregar filas adicionales si hay más autores)*

Pertencientes al programa de la **MAESTRÍA EN MICROBIOLOGÍA**, autores del trabajo titulado: **EVALUACIÓN DEL EFECTO DE LA TEMPERATURA DE INACTIVACIÓN, EXTRACCIÓN Y FRAGMENTACIÓN DEL ADNg DE Mycobacterium tuberculosis EN EL SECUENCIAMIENTO USANDO LA TECNOLOGÍA OXFORD NANOPORE**, el cual ha sido elaborado, sustentado y aprobado, según corresponda, para optar por el grado de **MAESTRA EN MICROBIOLOGÍA** bajo la modalidad de **TESIS**.

En calidad de docentes asesores de la Universidad Peruana Cayetano Heredia:

N°	APELLIDOS Y NOMBRES DEL DOCENTE	FACULTAD	NIVEL DE ASESORÍA
1.	SHEEN CORTAVARRIA PATRICIA	FACI	MAESTRÍA
2.	ZIMIC PERALTA MIRKO JUAN	FACI	MAESTRÍA

Declaramos que el contenido del presente documento es original y que las citas y referencias a otros autores cumplen con las normas académicas establecidas. En ese sentido, hacemos constar que:

- El documento presenta un porcentaje de similitud de **5%**, según el reporte emitido por el software **Turnitin®** (identificador de entrega: **2855794468**; fecha de entrega: **12-01-2026**).
- Tras una revisión detallada del reporte y del contenido del trabajo en cuestión, no se han identificado indicios de plagio.
- Se certifica que el documento respeta los principios de integridad académica y cumple con los requisitos institucionales de originalidad.

Lugar y fecha: **Lima, 12 de enero de 2026**



Firma del asesor  
N° DNI: 09541127  
ORCID: 0000-0002-7118-9301



Firma del Co-asesor  
N° DNI: 07620624  
ORCID: 0000-0002-7203-8847

# ÍNDICE

RESUMEN

ABSTRACT

1	INTRODUCCIÓN .....	1
1.1	Planteamiento del Problema .....	5
1.2	Pregunta de investigación.....	6
1.3	Justificación.....	7
2	OBJETIVOS .....	8
2.1	Objetivo general.....	8
2.2	Objetivos específicos.....	8
3	MARCO TEÓRICO .....	9
3.1	Tuberculosis y su impacto en la salud pública .....	9
3.2	Diagnóstico de <i>M. tuberculosis</i> .....	12
3.3	Importancia del WGS en TB .....	16
3.4	Tecnologías de secuenciamiento para <i>M. tuberculosis</i> .....	17
3.5	Secuenciamiento ONT, ventajas y limitaciones .....	21
3.6	Factores que afectan el secuenciamiento ONT .....	28
3.7	Avances en el secuenciamiento ONT de <i>Mycobacterium tuberculosis</i> .....	31
4	MATERIALES Y MÉTODOS.....	43
4.1	Diseño de estudio.....	43
4.2	Lugar de estudio .....	45
4.3	Material Biológico .....	45
4.4	Equipo de Protección Personal (EPP) e instalaciones .....	45
4.5	Preparación de muestras .....	46
4.6	Inactivación de la muestra .....	46
4.7	Extracción y purificación del ADN .....	47
4.8	Fragmentación del ADN .....	52
4.9	Preparación de librerías para secuenciar.....	52
4.10	Rendimiento del secuenciamiento .....	54
4.11	Análisis bioinformático.....	55
4.12	Flujograma de trabajo y análisis .....	56
4.13	Análisis estadístico .....	59
5	RESULTADOS .....	61

5.1	Efecto de la temperatura de inactivación (80 °C por 20 min y 95 °C por 1 hora) de Mtb en la cantidad, calidad e integridad del ADNg. ....	61
5.2	Efecto de cuatro métodos de extracción en la cantidad, calidad e integridad del ADNg de Mtb. ....	64
5.3	Purificación de las muestras para el secuenciamiento ONT del ADNg de Mtb. ..	67
5.4	Preparación de librerías para el secuenciamiento ONT.....	69
5.5	Efecto de la temperatura de inactivación (80°C por 20 min) en las métricas del secuenciamiento ONT del ADNg de Mtb. ....	71
5.6	Efecto de cuatro métodos de extracción de ADNg de Mtb en las métricas del secuenciamiento ONT.....	75
5.7	Peak Size de fragmentos de ADNg y rendimiento de celdas de flujo tras la fragmentación controlada 10 kpb. ....	79
5.8	Evaluación del efecto de la fragmentación a 10 kpb del ADNg de Mtb en las métricas del secuenciamiento ONT.....	85
5.9	Evaluación del efecto de la fragmentación a 10 kpb del ADNg de Mtb en las métricas de la calidad del ensamblaje post secuenciamiento ONT con read dúplex..	89
5.10	Confirmación de la inhibición del crecimiento de Mtb ante dos condiciones de inactivación térmica: 80 °C/20 min y 95 °C/1 hora.....	93
6	DISCUSIÓN .....	94
7	LIMITACIONES.....	103
8	CONCLUSIONES .....	104
9	RECOMENDACIONES .....	105
10	REFERENCIAS BIBLIOGRÁFICAS .....	107
11	ANEXOS	

## LISTA DE TABLAS

<b>Tabla 1.</b> Número de muestras por tratamiento térmico.....	48
<b>Tabla 2.</b> Secuenciamiento de las flow cells.....	54
<b>Tabla 3.</b> Rendimiento del secuenciamiento ONT .....	70
<b>Tabla 4.</b> Efecto de la temperatura de inactivación en las métricas del secuenciamiento ONT por cada método de extracción de ADNg .....	74
<b>Tabla 5.</b> Efecto de cuatro métodos de extracción de ADNg en las métricas del secuenciamiento ONT por temperatura de inactivación. ....	78
<b>Tabla 6.</b> Rendimiento del secuenciamiento ONT para las muestras No Fragmentadas vs Frgamentadas a 10 kpb.....	82
<b>Tabla 7.</b> Efecto de la fragmentación de ADNg a 10 kb en las métricas del secuenciamiento ONT. ....	84
<b>Tabla 8.</b> Efecto de la fragmentación de ADNg a 10 kpb en las métricas de precisión del secuenciamiento ONT por cada método de extracción. ....	88
<b>Tabla 9.</b> Métricas de la data dúplex del secuenciamiento ONT.....	91

## LISTA DE FIGURAS

<b>Figura 1.</b> Evolución de tecnologías de secuenciamiento. ....	17
<b>Figura 2.</b> Principio del secuenciamiento Nanopore .....	22
<b>Figura 3.</b> Pipeline de procesamiento y ensamblaje genómico de datos ONT (lecturas simplex y dúplex). ....	56
<b>Figura 4.</b> Flujograma del análisis de la evaluación del efecto de la temperatura de inactivación de Mtb (80°C por 20 min y 95°C por 1 hora) y el método de extracción de ADNg en la cantidad, calidad e integridad del ADNg. ....	57
<b>Figura 5.</b> Flujograma del análisis de la evaluación del efecto de la temperatura de inactivación (80°C por 20 min), extracción y fragmentación de ADNg de Mtb en el rendimiento del secuenciamiento ONT. ....	59
<b>Figura 6.</b> Efecto de la temperatura de inactivación de Mtb en la cantidad, calidad e integridad del ADNg por cada método de extracción. ....	63
<b>Figura 7.</b> Efecto de cuatro métodos de extracción en la cantidad, calidad e integridad del ADNg por cada temperatura de inactivación de Mtb. ....	66
<b>Figura 8.</b> Peak size de los fragmentos de ADN para cada método de extracción de ADNg y cada temperatura de inactivación de Mtb. ....	68
<b>Figura 9.</b> Efecto de la temperatura de inactivación 80°C/ 20 min de Mtb en el rendimiento del secuenciamiento ONT del ADNg de Mtb. ....	73
<b>Figura 10.</b> Efecto de cuatro métodos de extracción de ADNg de Mtb en el rendimiento del secuenciamiento ONT. ....	77
<b>Figura 11.</b> Peak size del ADNg para cada condición de fragmentación en tres métodos de extracción de ADNg de Mtb. ....	81

<b>Figura 12.</b> Efecto de la fragmentación a 10 kpb del ADNg de Mtb en el rendimiento del secuenciamiento ONT.....	87
<b>Figura 13.</b> Efecto de la fragmentación a 10 kpb del ADNg de Mtb sobre las métricas de la calidad del ensamblaje post secuenciamiento ONT con read dúplex. ....	92
<b>Figura 14.</b> Evaluación del crecimiento de Mtb inactivadas a 80°C/20 min y 95°C/1h tras 90 días de incubación a 37°C.....	93

## RESUMEN

La secuenciación de lectura larga por Oxford Nanopore Technologies aplicada a *Mycobacterium tuberculosis* ofrece ventajas para resolver regiones repetitivas y estructuralmente complejas, aunque su desempeño depende críticamente de decisiones pre-analíticas. Este estudio evaluó el impacto de (i) la inactivación térmica — 95 °C/1 h; 80 °C/20 min —, (ii) cuatro métodos de extracción de ADN genómico (ADNg) y (iii) la fragmentación del ADNg a ~10 kpb sobre métricas clave del secuenciamiento (N50, cobertura, profundidad y precisión) y sobre el ensamblaje con lecturas dúplex. Los resultados mostraron que la inactivación a 95 °C/1 h degrada el ADNg y reduce la proporción de lecturas largas. En contraste, la condición sin inactivación térmica combinada con el método de extracción de ADNg M3b con purificación reforzada mejoró el rendimiento del secuenciamiento. Además, la fragmentación controlada del ADN aumentó el rendimiento en gigabases y la profundidad genómica a 214.7 X, con un N50 menor de valor 7.33 kb, manteniendo la cobertura genómica completa. Con el método M3b fragmentado a ~10 kpb, el ensamblaje con lecturas dúplex alcanzó coberturas casi totales del genoma y precisiones comparables a los valores de referencia de ONT, con bajas tasas de errores e indels. En conclusión: (1) debe evitarse la inactivación a 95 °C/1 h si se busca maximizar lecturas largas; (2) los métodos de extracción como M3b con purificación adicional optimizan las razones 260/280 y 260/230 y el rendimiento del secuenciamiento; (3) el método M3b fragmentado a ~10 kpb  $\pm$  2 kpb es recomendable cuando se prioriza profundidad y rendimiento total, asumiendo un N50 menor y (4) las lecturas dúplex permiten ensamblajes de alta

precisión cuando la cobertura lo permite. Estas decisiones definen un flujo de trabajo que equilibra rendimiento y calidad, adecuado para vigilancia genómica y aplicaciones de investigación.

**Palabras clave:** Secuenciación de lectura larga, *Mycobacterium tuberculosis*, Inactivación térmica, Método de extracción de ADNg, Fragmentación del ADNg a 10 kpb

## ABSTRACT

Long-read sequencing by Oxford Nanopore Technologies applied to *Mycobacterium tuberculosis* offers advantages for resolving repetitive and structurally complex genomic regions, although its performance depends critically on pre-analytical decisions. This study evaluated the impact of (i) heat inactivation — 95 °C/1 h; 80 °C/20 min —, (ii) four genomic DNA (gDNA) extraction methods, and (iii) gDNA fragmentation to ~10 kbp on key sequencing metrics (N50, coverage, depth, and accuracy) and on duplex-read assembly. The results showed that inactivation at 95 °C/1 h degrades gDNA and reduces the proportion of long reads. In contrast, the non-heat-treated condition combined with the M3b gDNA extraction method with reinforced purification improved sequencing performance. In addition, controlled DNA fragmentation increased gigabase output and genomic depth to 214.7×, with an N50 of 7.33 kb while maintaining full genome coverage. With the fragmented M3b method to ~10 kbp, duplex-read assemblies achieved near-complete genome coverage and accuracies comparable to ONT reference values, with low error and indel rates. In conclusion: (1) heat inactivation at 95 °C/1 h should be avoided when maximizing long-read yield; (2) extraction methods such as M3b with additional purification optimize 260/280 and 260/230 ratios and sequencing performance; (3) the M3b method fragmented to ~10 kbp ± 2 kbp is recommended when depth and total output are prioritized, assuming a lower N50; and (4) duplex reads enable high-accuracy assemblies when coverage permits. These decisions define a workflow that balances yield and quality, suitable for genomic surveillance and research applications.

**Keywords:** Long-read sequencing, *Mycobacterium tuberculosis*, Heat inactivation  
,gDNA extraction method , gDNA fragmentation to ~10 kpb

## 1 INTRODUCCIÓN

La tuberculosis (TB), una enfermedad infecciosa causada por la bacteria *Mycobacterium tuberculosis* (Mtb), sigue constituyendo una de las amenazas más graves para la salud humana, siendo considerada una de las principales causas de muerte a nivel global (1,2). Según el “Informe Global sobre Tuberculosis 2024” de la Organización Mundial de la Salud (OMS), en 2023, se estima que 10,8 millones de personas enfermaron de TB en todo el mundo, lo que resultó en 1,25 millones de muertes (3). Los esfuerzos para controlar la TB se ven obstaculizados por el aumento de cepas de Mtb resistentes y multirresistentes a los fármacos, así como por la falta de pruebas rápidas y precisas para clasificar sus perfiles de resistencia (4).

Perú es uno de los 30 países con mayor carga de TB multidrogorresistente (TB-MDR) (1,2). El diagnóstico tradicional de drogo – resistencia (DR) de Mtb se basa en pruebas fenotípicas, que pueden tardar hasta dos meses retrasando el tratamiento (5). Para agilizar este proceso, se han implementado pruebas moleculares rápidas como el Xpert MTB/RIF y el GenoType MTBDRplus que permiten la identificación temprana de resistencia a los fármacos de primera línea (6). No obstante, estos métodos no detectan la resistencia a fármacos de segunda línea y pasan por alto mutaciones fuera de las regiones “*hot spots*” seleccionadas, lo que puede incurrir en resultados falsos negativos en la detección DR.

Para una estrategia eficaz de control epidemiológico de la TB es esencial conocer los genomas completos de Mtb. En este contexto, el secuenciamiento del genoma

completo (WGS) emerge como una herramienta fundamental al facilitar un análisis exhaustivo del genoma de Mtb (7). Este enfoque no sólo ofrece perfiles más completos DR (7,8) que permite una predicción rápida y precisa de la DR fenotípica (8,9). Sino también permite la identificación de brotes, seguimiento de cadenas de transmisión, detección de nuevas cepas, y evaluación de la diversidad genética (10). Por tanto, la integración de WGS en el marco de gestión y control de la TB anuncia nuevas vías para mejorar las modalidades de diagnóstico y tratamiento.

El flujo de trabajo estándar para el análisis de WGS en cepas de Mtb ha considerado la secuenciación de tecnologías de lectura corta de alta precisión como las plataformas Illumina (10). Un desafío clave con la secuenciación de lectura cortas de Illumina en Mtb es la exclusión de alrededor del 10% del genoma debido a errores de mapeo en regiones complejas como las familias de genes PE (prolina-glutamato) y PPE (prolina-prolina-glutamato), exclusivas del complejo Mtb y caracterizadas por secuencias ricas en GC, con múltiples motivos repetitivos, elevada similitud entre sí y elementos genéticos móviles (10). Esta arquitectura dificulta la asignación única de lecturas de 150–300 pb, lo que genera fallas de alineamiento, pérdida de cobertura y una representación incompleta de estos loci. Así mismo, desde una perspectiva logística y financiera este tipo de secuenciamiento implica un alto costo de reactivos, mantenimiento de cadena de frío en el envío de estos y la necesidad de equipos y ambientes especializados. Todo esto representa una barrera significativa en entornos con recursos limitados (5,11).

En este sentido la incorporación de nuevas opciones de tecnologías de secuenciamiento como la tecnología de lectura larga de Oxford Nanopore ofrece

muchas ventajas, entre ellas: la longitud de lectura, el bajo costo de reactivos, el envío de reactivos sin necesidad de cadena de frío, la portabilidad de equipos y resultados rápidos (11). Estas ventajas prometen hacer este análisis de secuenciación más factible de realizar en todas las instalaciones y más robusto, permitiendo una cobertura más completa del genoma (10,12,13).

La tecnología de secuenciación Oxford Nanopore (Oxford Nanopore Technology, *ONT*) permite generar lecturas largas, facilitando la resolución de regiones genómicas complejas y la mejora del ensamblaje del genoma de *Mtb* (13,14). Sin embargo, su aplicación requiere optimizar varios procesos, incluida la extracción de ADN (15). El éxito de la secuenciación *ONT* en *Mtb* depende de obtener ADNg de alto peso molecular (HMW), en cantidades de 400 - 1000 ng, y con índices de pureza de 260/280 ~ 1.8 y 260/230 entre 2.0 - 2.2 (15,16). Actualmente, no existe un flujo de trabajo que garantice datos de alta calidad como resultado del secuenciamiento de ADNg de *Mtb* usando *ONT*. Los estudios realizados hasta la fecha se han centrado principalmente en la identificación y caracterización genómica de *Mtb*, con énfasis en la detección de variantes genéticas y predicción de resistencia a fármacos. Sin embargo, no hay estudios que se hayan centrado en la optimización del secuenciamiento *ONT* de *Mtb* y que garanticen una data detallada, rápida y precisa para los análisis genómico.

Para establecer un flujo de trabajo que optimice el secuenciamiento de ADNg de *Mtb* usando *ONT* es necesario evaluar tres puntos clave en relación a su extracción del ADNg. El primero es encontrar el ajuste preciso de la temperatura de inactivación, buscando equilibrar la bioseguridad necesaria para manipular *Mtb* y

la integridad del ADNg (17), el segundo es evaluar el riesgo significativo que implica el uso de fenol en la extracción del ADN durante la corrida de secuenciación – ya que pequeñas cantidades pueden dañar los poros de secuenciación, lo que reduce el desempeño de la ejecución (18), y el tercero es evaluar la condición de fragmentación del ADNg - cuidando de mantener aún fragmentos largos y la integridad de la muestra - para mitigar los desafíos planteados por el alto contenido de GC del genoma de Mtb y su propensión a la formación de estructuras secundarias que podrían bloquear los poros (19,20).

La falta de estudios que aborden estos tres puntos clave subraya la necesidad de establecer un flujo de trabajo que optimice la secuenciación de Mtb mediante ONT; abordando las complejidades de la extracción y la secuenciación que permitan mejorar el rendimiento del secuenciamiento, al tiempo que garantiza la seguridad del laboratorio.

## 1.1 Planteamiento del Problema

El WGS de Mtb está actualmente dominado por la tecnología de secuenciación de lecturas cortas de la plataforma Illumina, siendo considerada el gold estándar para el secuenciamiento (21). Sin embargo, esta tecnología enfrenta limitaciones significativas como el alto costo de reactivos, mantenimiento de cadena de frío en el envío de estos y la necesidad de equipos y ambientes especializados. Así mismo, esta tecnología no logra caracterizar regiones genómicas complejas, como las familias de genes PE/PPE y las áreas repetitivas del genoma de Mtb (10). Estas limitaciones han restringido su uso en entornos con recursos limitados y han dificultado la obtención de ensamblajes completos y detallados del genoma (5,11).

En este contexto, la secuenciación de lectura larga de ONT ofrece múltiples ventajas para los laboratorios de referencia y diagnóstico como el bajo costo de inversión, kits de secuenciación libres de cadena de frío para su traslado, portabilidad del dispositivo de secuenciación, análisis en tiempo real, y la capacidad de generar lecturas extensas (11,15,22–24). Este último permite resolver regiones complejas del genoma de Mtb, facilitando un ensamblaje más completo (10,12,13). La combinación de su baja inversión de capital y la portabilidad de ONT hace a esta tecnología más accesible para países con recursos limitados y alta carga de TB, acercando el diagnóstico a los puntos de atención y fortaleciendo la vigilancia genómica (24). No obstante, no existe un protocolo estandarizado que asegure un buen rendimiento y desempeño del secuenciamiento de Mtb por ONT.

Para diseñar un flujo de trabajo optimizado de secuenciación de Mtb utilizando ONT es fundamental evaluar el efecto de la temperatura de inactivación, la extracción y fragmentación en el rendimiento del secuenciamiento. Esto con el fin de identificar el mejor método que proporcione la mayor cantidad, calidad e integridad de ADNg; así como el mayor rendimiento en el secueciamiento. Actualmente, no existen estudios que aborden de manera integral estos puntos clave. Esto limita la implementación de un flujo de trabajo optimizado que asegure datos de alta calidad y maximice el rendimiento en la secuenciación por ONT.

## **1.2 Pregunta de investigación**

El presente estudio plantea abordar los vacíos de conocimiento previamente expuestos para proponer un flujo de trabajo optimizado para secuenciar muestras de ADNg de Mtb usando la tecnología Oxford Nanopore. Por ello planteamos la siguiente pregunta de investigación: ¿La temperatura de inactivación, las técnicas de extracción y fragmentación del ADNg de *M. tuberculosis* afectan el rendimeinto del secuenciamiento usando la tecnología Oxford Nanopore?

### 1.3 Justificación

Este estudio busca explorar de manera integral el efecto de la inactivación térmica, el método de extracción, con y sin fenol, y la fragmentación del ADNg en el rendimiento de la secuenciación de Mtb con ONT. Estas 3 variables son determinantes para definir un flujo de trabajo optimizado que permita maximizar el rendimiento del secuenciamiento

Evaluar métodos de extracción de ADNg que emplean o no fenol-cloroformo es fundamental para obtener indicios experimentales sobre su impacto potencial en el rendimiento de la secuenciación por la tecnología ONT. Los métodos basados en fenol-cloroformo representan una alternativa más económica (\$2.6/muestra) (25) en comparación con los kits comerciales como ZymoBiomics DNA miniprep (\$7.6/muestra).

En base a lo expuesto, este estudio es necesario para explorar y llenar los vacíos de conocimiento sobre el impacto que tienen la temperatura de inactivación, los métodos de extracción y la fragmentación del ADNg en el rendimiento de la secuenciación ONT. Estos análisis exploratorios del efecto de estas 3 variables proporcionarán indicios clave para identificar las condiciones ideales del ADNg que aseguren un secuenciamiento eficiente y la generación de datos de alta calidad. La optimización de estas variables permitiría diseñar un flujo de trabajo eficiente y de fácil implementación para secuenciar de manera precisa el ADNg de Mtb mejorando así las capacidades diagnósticas.

## **2 OBJETIVOS**

### **2.1 Objetivo general**

Evaluar del efecto de la temperatura de inactivación, métodos de extracción y fragmentación del ADNg de *Mycobacterium tuberculosis* (Mtb) en el secuenciamiento empleando la tecnología Oxford Nanopore.

### **2.2 Objetivos específicos**

- Evaluar el efecto de la temperatura de inactivación (80°C por 20 min y 95°C por 1 hora) en la cantidad, calidad e integridad del ADNg de Mtb.
- Evaluar el efecto de cuatro métodos de extracción en la cantidad, calidad e integridad del ADNg de Mtb.
- Evaluar el efecto de la temperatura de inactivación (80°C por 20 min) de Mtb en el rendimiento del secuenciamiento ONT del ADNg de Mtb.
- Evaluar el efecto de cuatro métodos de extracción de ADNg de Mtb en el rendimiento del secuenciamiento ONT.
- Evaluar el efecto de la fragmentación a 10 kpb del ADNg de Mtb en el rendimiento del secuenciamiento ONT.

### 3 MARCO TEÓRICO

#### 3.1 Tuberculosis y su impacto en la salud pública

La tuberculosis (TB) es una enfermedad infecciosa antigua y aún prevalente. Esta es causada por el bacilo *Mycobacterium tuberculosis* (Mtb) y considerada como la principal causa de muerte por agente infeccioso único (3,26). La TB sigue constituyendo una de las amenazas más graves para la salud pública, siendo considerada una de las 10 principales causas de muerte a nivel mundial (1,2). Aproximadamente más de 10 millones de personas siguen enfermando de TB cada año; según el “Informe mundial sobre la tuberculosis 2024” de la Organización Mundial de la Salud (OMS), en 2023, aproximadamente 10.8 millones de personas enfermaron de TB en todo el mundo, lo que provocó 1.25 millones de muertes (3). Además, se estima que una cuarta parte de la población mundial ha sido infectada con Mtb (27), y que el riesgo de desarrollar TB es mayor en los primeros dos años luego de la infección (aproximadamente el 5 %) (28).

Mtb es un bacilo gram positivo, aerobio y de lento crecimiento, que se divide cada 18 a 20 horas a diferencia de la mayoría de las bacterias que se dividen cada 20 min (29). Por ello, su cultivo en medios sólido tarda aproximadamente 3 semanas (30). Esta bacteria se esparce cuando la persona que está enferma con TB expulsa bacterias dentro del aire (3). Su pared celular contiene una gran cantidad de lípidos, y ácidos micólicos, que le otorgan ácido-alcohol resistencia en las tinciones impidiendo el ingreso de colorantes en la tinción de Gram, siendo la tinción de Ziehl Neelsen la adecuada para la identificación de esta (30). Por ello, su pared celular es

considerada un mecanismo distintivo de resistencia a numerosos antibióticos, ya que su impermeabilidad dificulta el acceso de estos a su diana bacteriana (31).

Los antibióticos utilizados en el tratamiento contra la TB se dividen generalmente en dos grupos principales: los de primera y segunda línea. Los fármacos de primera línea son considerados esenciales debido a su alta eficacia, buena tolerancia y menor toxicidad, e incluyen isoniazida (INH), rifampicina (RIF), pirazinamida (PZA) y etambutol (EMB), utilizados en un régimen de 6 meses (los cuatro fármacos durante los primeros dos, seguidos de I y R durante los cuatro meses restantes) (3). Cuando se presenta resistencia a estos medicamentos, se requieren antibióticos de segunda línea, tales como las fluoroquinolonas (FQ, levofloxacin o moxifloxacin), aminoglucósidos (AMG) (amikacina (AMK) o estreptomina (STR)), bedaquilina (BDQ), cicloserina (CS), clofazimina (CLF) y linezolid (LZD) (32). Estos últimos presentan mayor toxicidad, efectos adversos más frecuentes y son generalmente menos eficaces, por lo que se reservan para casos específicos de tuberculosis resistente a múltiples medicamentos (33).

La TB resistente a medicamentos (TB-DR) se ha convertido nuevamente en una grave amenaza y reto para la salud pública mundial debido a la aparición de cepas resistentes a los fármacos (34). La OMS usa 5 categorías para clasificar los casos de TB – DR.: La TB con resistencia a la rifampicina (TB-RR), la TB resistente a la rifampicina e isoniazida que se define como TB multidrogo resistente (TB-MDR), la TB pre-XDR resistente (TB pre-XDR) definida como TB resistente a la rifampicina y a cualquier fluoroquinolona, y la XDR-TB, definida como TB

resistente a la rifampicina, más cualquier fluoroquinolona y al menos uno de los siguientes fármacos: bedaquilina o linezolid (3).

En el caso de la TB-MDR o TB-XDR, el tratamiento requiere combinaciones más complejas y prolongadas de fármacos de segunda línea, tales como fluoroquinolonas (levofloxacino, moxifloxacino), bedaquilina, linezolid, cicloserina y clofazimina, extendiéndose generalmente entre 6 y 20 meses (3,32). Esto último depende del perfil específico de resistencia, la tolerancia del paciente y la evolución clínica, haciendo necesario un monitoreo riguroso y frecuente para asegurar la eficacia terapéutica y evitar efectos adversos significativos.

Globalmente en el 2023, 175 923 personas fueron diagnosticadas y tratadas para TB - MDR/RR. Esto representa sólo el 44% de las 400 000 personas (95% IC: 360 000–440 000) que fueron estimadas por la OMS (3). Así mismo, en el 2022 en la Región de las Américas se estimaron 11 600 casos de TB-RR/MDR y sólo se diagnosticaron 5 428, con un incremento del 12% en comparación con el 2021, y de estos el 100% inició tratamiento (35). En la región de las Américas, El Perú en conjunto con Brasil, concentra el 50 % de los casos estimados de TB-RR/MDR, siendo a nivel nacional Lima Metropolitana y el Callao las ciudades con mayor carga (36). En el 2023, en el Perú se han identificado 33 113 casos de TB, de los cuales 1424 fueron TB-MDR, 635 TB-RR y 5 TB-XDR (37).

La tasa de éxito del tratamiento contra la TB sensible a los medicamentos sigue siendo alta (88%) y ha mejorado al 68% para los casos de TB-MDR/RR. Sin

embargo, a pesar de los esfuerzos, persisten desafíos en la detección y tratamiento de la TB, TB-MDR, TB-RR y TB-XDR que aún necesitan ser mejorados (3). La brecha entre los casos estimados y los efectivamente diagnosticados y tratados evidencian deficiencias en la detección temprana y acceso limitado a pruebas diagnósticas y barreras en la provisión de tratamientos adecuados.

Para abordar esta problemática, es esencial el fortalecimiento de la detección y el diagnóstico de la TB en los sistemas de salud de primer nivel de atención con el fin de mejorar la búsqueda de casos de manera temprana y cortar la cadena de transmisión en la comunidad. Para ello se requiere la implementación de pruebas rápidas de detección molecular, asegurar el acceso a regímenes terapéuticos efectivos y robustecer los programas de vigilancia epidemiológica .

### **3.2 Diagnóstico de *M. tuberculosis***

Personas que son diagnosticadas con TB haciendo uso de test moleculares rápidos recomendados por la OMS, ensayos de lipoarabinomano en orina de flujo lateral, la baciloscopía, y el cultivo son definidas como casos de TB “bacteriológicamente confirmados” (3). Esta detección microbiológica no sólo permite un diagnóstico correcto, sino que también asegura la selección temprana de un régimen de tratamiento adecuado. En muchos países con alta carga de TB, países de bajos y medianos recursos, la baciloscopia de esputo sigue siendo aún el método de diagnóstico principal en los niveles primarios de atención para evaluar a las personas que presentan signos y síntomas de TB (38). Sin embargo, esta prueba presenta una sensibilidad variable y baja (entre 20% - 80% dependiendo

principalmente de la carga de micobacterias) (39,40) y además no permite distinguir entre cepas drogosensibles (DS) y drogoresistentes (DR). La OMS recomienda que los programas de TB sustituyan la baciloscopía como prueba diagnóstica inicial por pruebas moleculares rápidas que detecten el complejo Mtb.

El cultivo es el método de referencia (*gold standard*) para la confirmación bacteriológica de TB y para el diagnóstico de susceptibilidad a fármacos (*Drug Susceptibility Testing*, DST), excepto para etambutol, etionamida, protionamida, ácido para-aminosalicílico, imipenem-cilastatina y meropenem (38). Sin embargo, este no es usado como una prueba de diagnóstico primaria en muchos países con alta carga de TB debido al costo, requerimientos de infraestructura (nivel de bioseguridad 3) y el crecimiento lento de Mtb que termina impactando en el largo tiempo requerido para generar resultados (1 a 3 semanas para un resultado positivo y hasta 6 semanas para un resultados negativo para el diagnóstico de Mtb y de semanas hasta 3 meses para generar un perfil de DR completo) (6,9,10,35,38). No obstante, la baciloscopia y el cultivo siguen siendo necesarios para monitorear la respuesta del paciente al tratamiento.

Con el propósito de obtener resultados precisos y rápidos para el diagnóstico inicial de la TB y la detección de resistencia a fármacos, la OMS ha recomendado el uso de pruebas moleculares rápidas como Xpert® MTB/RIF Ultra (Cepheid), Truenat MTB, MTB Plus y MTB-RIF Dx assays, que permiten detectar no sólo Mtb sino también la resistencia a la RIF en menos de dos horas (34). Así mismo, sugiere pruebas de amplificación de ácidos nucleicos (*Nucleic Acid Amplification Tests*,

NAATs) automatizadas de moderada complejidad que permiten la detección simultánea de Mtb y de mutaciones asociadas a la farmoresistencia tanto a la RIF e INH: Abbott RealTime MTB, MTB RIF/INH, BD MAX MDR-TB, Bruker-Hain FluoroType MTBDR, y Roche cobas MTB y MTB-RIF/INH (34,38). Con respecto a la detección de Mtb, estas pruebas han demostrado una mayor sensibilidad y especificidad en comparación a la baciloscopía.

Para la detección de resistencia adicional a fármacos antituberculosos en personas con TB bacteriológicamente confirmada, la OMS recomienda el uso de pruebas de seguimiento que complementan el diagnóstico inicial. Entre ellas, los NAATs automatizados de baja complejidad, como Xpert MTB/XDR, que permiten detectar resistencia a INH y a fármacos de segunda línea, incluyendo fluoroquinolonas y aminoglucósidos en menos de 90 minutos (38,41). Por otro lado, los NAATs de reversa hibridización y alta complejidad, como Genoscholar PZA-TB, están diseñados para la detección rápida de resistencia a la PZA, ofreciendo una alternativa molecular más rápida al DST fenotípico (38).

Las pruebas rápidas recomendadas por la OMS han mejorado la detección de Mtb y la identificación de resistencia a fármacos, permitiendo un inicio más temprano del tratamiento. No obstante, estas pruebas están limitadas principalmente a RIF e INH y sólo permiten detectar mutaciones específicas en regiones génicas preseleccionadas asociadas a resistencia (40). En este contexto, las tecnologías de Secuenciación de Próxima Generación (Next-Generation Sequencing, NGS) han revolucionado el diagnóstico de la TB-DR, permitiendo la detección simultánea

precisa y rápida de mutaciones asociadas a la resistencia en múltiples genes y para distintos medicamentos en una sola prueba (43). NGS se refiere a las técnicas que se basan en la secuenciación de múltiples fragmentos de ADN en simultáneo, seguida de análisis bioinformáticos para ensamblar las secuencias. Estas tecnologías pueden utilizarse para determinar el secuenciamiento del genoma completo (WGS, Whole-Genome Sequencing) o de parte de un genoma, secuenciamiento dirigido (tNGS, Targeted NGS) en un único proceso de secuenciamiento (22). La WHO recomienda WGS y la tNGS para la vigilancia de la TB-DR. La WGS se utiliza en aislados de *Mtb* cultivados, mientras que la tNGS puede utilizarse directamente en muestras clínicas como las de esputo (44).

En el 2023 la WHO recomendó el uso de *Targeted Next-Generation Sequencing* (tNGS) como prueba de seguimiento para identificar mutaciones asociadas a la resistencia a múltiples fármacos en un sólo análisis. Estas pruebas, incluyen Deeplex® Myc-TB (GenoScreen) para detectar la resistencia para RIF, INH, PZA, EMB, FQ, BDQ, LZD, CFZ, AMK y STR; AmPORE TB (Oxford Nanopore Technologies) para RIF, INH, FQ, LZD, AMK y STR; y TBseq® (ShengTing Biotech) para EMB (38,45). La implementación de estas pruebas avanzadas mejora la detección temprana de resistencia a fármacos clave , optimizando la selección de tratamientos personalizados para pacientes con TB-DR. Sin embargo, el tNGS se limita al análisis de regiones genómicas específicas y no permite una caracterización más detallada de todas las regiones del genoma de las cepas *M. tuberculosis* .

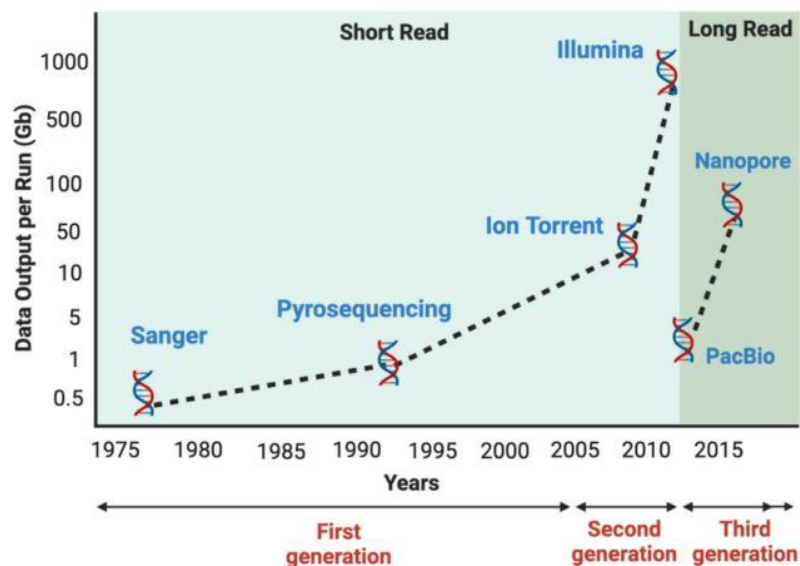
### **3.3 Importancia del WGS en TB**

En este contexto, la secuenciación del genoma completo (Whole-Genome Sequencing, WGS) ha surgido como una herramienta clave en la vigilancia y diagnóstico avanzado de Mtb, ofreciendo ventajas significativas sobre las pruebas moleculares convencionales (19,38). La WGS proporciona una caracterización integral del genoma bacteriano, permitiendo la identificación de todas las mutaciones asociadas a resistencia, incluyendo aquellas emergentes que aún no han sido incorporadas en los catálogos de mutaciones clínicas (44). Esta capacidad hace del WGS una estrategia más precisa para la detección temprana de la resistencia a múltiples fármacos de Mtb, contribuyendo así a la personalización del tratamiento y la reducción del riesgo a fallas terapéuticas.

Además de su capacidad diagnóstica, el uso del WGS ofrece ventajas en la epidemiología y control de la TB, facilitando el rastreo de la transmisión y la identificación de brotes (10). A diferencia de las pruebas moleculares convencionales, que se centran en la detección de resistencia específica, la WGS permite diferenciar cepas con alta resolución genómica, ayudando a determinar relaciones filogenéticas y cadenas de transmisión en poblaciones afectadas (10,19). Esto es particularmente relevante en entornos con alta prevalencia de TB-MDR y TB-XDR, donde la identificación rápida de variantes resistentes permite una respuesta sanitaria más efectiva (9). Asimismo, la integración del uso de la WGS con bases de datos globales mejora la vigilancia genómica y facilita la detección de nuevas mutaciones relevantes para la farmacorresistencia, reforzando el control de la enfermedad y optimizando estrategias de salud pública (44).

### 3.4 Tecnologías de secuenciamiento para *M. tuberculosis*

En las últimas décadas, las tecnologías de secuenciación de ADN han evolucionado significativamente, permitiendo avances clave en la investigación genómica de *Mtb*. Este desarrollo ha dado lugar a tres generaciones de tecnologías de secuenciación, desde el método Sanger hasta las plataformas de tercera generación (*Third-Generation Sequencing*, TGS) (22) (Figura 1), cada una con principios de funcionamiento distintos, ventajas y limitaciones específicas (22,46,47).



**Figura 1.** Evolución de tecnologías de secuenciamiento. Tomado de Satam *et al* (22).

La primera generación, representada por el método de Sanger, se basa en la terminación de cadena mediante dideoxinucleótidos marcados con fluorescencia. Durante la amplificación del ADN, estos nucleótidos modificados detuvieron la síntesis en diferentes posiciones, generando fragmentos de distintas longitudes que, tras ser separados por electroforesis capilar, permiten inferir la secuencia de bases (22,47). Aunque esta tecnología proporcionó las primeras secuencias de ADN y fue

fundamental en proyectos como el Genoma Humano, su baja escalabilidad y alto costo limitaron su aplicación en estudios de genómica masiva (19,46,47).

La segunda generación, con plataformas de secuenciamiento como Illumina e Ion Torrent, introdujo la secuenciación masivamente paralela que permitió procesar millones de fragmentos de ADN de manera simultánea en contraste a las plataformas de la primera generación (22). Illumina, la plataforma de secuenciamiento de alto rendimiento más ampliamente usada, se basa en la secuenciación por síntesis. Esta tecnología utiliza nucleótidos marcados con fluoróforos que se incorporan secuencialmente en fragmentos previamente amplificados en una celdade flujo (*flow cell*), emitiendo señales fluorescentes que son detectadas para determinar la secuencia nucleótidos en fragmento de ADN (48). En contraste, Ion Torrent se basa en la detección de cambios en el pH, en lugar de señales ópticas, para identificar la incorporación de nucleótidos (47). Estas tecnologías permitieron reducir drásticamente los costos y tiempos de secuenciación, aunque presentan limitaciones en la detección de variaciones estructurales grandes debido a sus lecturas cortas (49).

Finalmente, la tercera generación, representada por Pacific Biosciences (PacBio) y Oxford Nanopore Technologies (ONT), revolucionó la secuenciación al eliminar la necesidad de amplificación del ADN y permitir la secuenciación de moléculas individuales en tiempo real (22,49). PacBio SMRT se basa en la detección de nucleótidos marcados con fluorescencia a medida que son incorporados en una única molécula de ADN dentro de un nanopozo, permitiendo lecturas más largas y

la identificación de modificaciones epigenéticas como la metilación (50). Por otro lado, ONT emplea nanoporos biológicos insertados en una membrana, a través de los cuales pasa la hebra de ADN. A medida que las bases atraviesan el poro, generan alteraciones en la corriente eléctrica que son interpretadas para inferir la secuencia. Estas tecnologías permiten obtener lecturas ultralargas, facilitando el ensamblaje de novo de genomas completos y el análisis de regiones repetitivas (16).

En los últimos 10 años, las plataformas de secuenciación más ampliamente utilizadas para el WGS de *Mtb* han sido tanto de tecnologías de lectura corta y de lectura larga como Illumina y ONT respectivamente (51). Estas tecnologías han revolucionado la epidemiología genómica y el estudio de la resistencia a los antibióticos de *Mtb*, permitiendo análisis más precisos y accesibles en entornos clínicos y de investigación (46). Illumina pertenece a la segunda generación de secuenciación (NGS) y se basa en la secuenciación por síntesis (Sequencing by synthesis, SBS), donde nucleótidos marcados con fluoróforos se incorporan de manera secuencial a fragmentos de ADN (47). Illumina ofrece una amplia gama de plataformas de secuenciación, que se clasifican según su capacidad de procesamiento en sistemas de bajo a medio rendimiento, como MiniSeq y MiSeq, y plataformas de alto rendimiento, como *NextSeq 550*, *NextSeq 1000 and 2000* y *NovaSeqX* (52).

Estas plataformas del secuenciamiento Illumina han permitido una mayor producción a menor coste para la generación de lecturas cortas de alta calidad (19). MiSeq es una opción popular para el secuenciamiento de genoma completo de

bacterias, ya que ofrece un equilibrio entre costo y precisión con lecturas de hasta 300 pb en modo pareado (52,53). MiniSeq es una alternativa compacta y accesible para estudios pequeños (54), mientras que NextSeq y NovaSeq permiten un mayor rendimiento y profundidad de cobertura, lo que es ideal para estudios epidemiológicos (52,55,56).

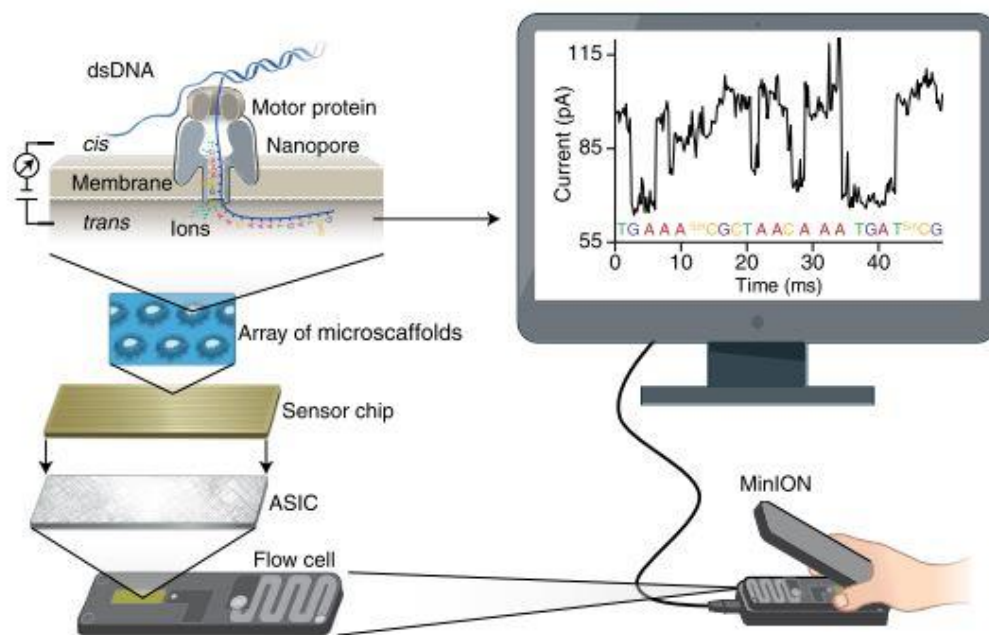
Las principales ventajas del secuenciamiento Illumina incluyen su alta precisión (>Q30), bajo costo por base y cobertura profunda, lo que permite detectar con gran confiabilidad mutaciones puntuales asociadas a resistencia antimicrobiana (13,43,57). Sin embargo, una de sus principales limitaciones es el tamaño de lectura corto (150–300 pb), lo que dificulta el ensamblaje *de novo* y la identificación de grandes reordenamientos genómicos, en especial en las regiones repetitivas y de alto contenido de GC de Mtb (16). Dentro de estas regiones se encuentran las familias génicas PE y PPE, exclusivas del complejo Mtb y caracterizadas por secuencias ricas en prolina y glutamato, así como por motivos repetitivos y una elevada identidad entre sí (12). Estas propiedades generan ambigüedad durante el mapeo de lecturas cortas (21), dado que múltiples fragmentos presentan alta identidad entre sí, lo que provoca fallas de alineamiento, pérdida de cobertura y una representación incompleta de estos loci. Por ello, los genes PE y PPE constituyen uno de los principales desafíos para la secuenciación Illumina y explican la necesidad de tecnologías de lectura larga capaces de atravesar repeticiones extensas y resolver estas regiones estructuralmente complejas.

ONT, que pertenece al secuenciamiento de la tercera generación, surge como una tecnología que nos permite resolver las limitaciones del secuenciamiento Illumina. La longitud media de lectura a menudo supera los 10 kb (22), y la longitud máxima de lectura alcanzada hasta el 2021 es de 4.2 Mb (58). Nanopore no sólo destaca por su capacidad de generar lecturas largas, sino también sus plataformas nos permiten una secuenciación en tiempo real y la detección directa de modificaciones epigenéticas sin necesidad de una amplificación previa (16). Una de las principales limitaciones de esta tecnología desde su lanzamiento ha sido la baja precisión de las lecturas en bruto, siendo más baja que la precisión de Illumina (11,43). No obstante, los avances en el diseño de las flow cells y en la química de los kits de preparación de librerías han mejorado significativamente su precisión en los últimos años. Estos desarrollos han permitido superar la calidad Q20 en lecturas en bruto y, mediante la generación de lecturas dúplex, alcanzar precisiones superiores a Q30 (59).

### **3.5 Secuenciamiento ONT, ventajas y limitaciones**

El principio de la tecnología de secuenciamiento de Oxford Nanopore utiliza un poro proteico a escala nanométrica, denominado "nanoporo", que actúa como biosensor y está incrustado en una membrana polimérica embebida en una solución electrolítica (60). La aplicación de un voltaje constante, dentro de esta solución, genera una corriente iónica a través del nanoporo que impulsa moléculas de ADN o ARN monocatenario desde el lado negativo ("cis") hacia el lado positivo ("trans") de la membrana. Esta translocación de los ácidos nucleicos es regulada por una

proteína motora (denominada adaptador), que los desplaza a través del nanoporo. Durante este proceso de desplazamiento ocurren cambios en la conductancia eléctrica cuando el ADN nativo atraviesa el nanoporo. Estas variaciones en la corriente iónica corresponden a la identidad y el orden de los nucleótidos presentes en la región de detección del nanoporo y a través de algoritmos computacionales estas señales eléctricas son procesadas y traducidas en tiempo real, permitiendo la reconstrucción de la secuencia del ácido nucleico analizado (Figura 2). Además de controlar la velocidad de translocación, la proteína motora posee actividad helicasa que facilita la separación del ADN bicatenario en hebras simples capaces de atravesar el nanoporo (16).



**Figura 2.** Principio del secuenciamiento Nanopore  
Tomado de Wang *et al* (16).

Este principio permitió secuenciar el ADN nativo en su longitud completa sin necesidad de fragmentación previa, lo que representa una ventaja clave sobre

tecnologías de lectura corta como Illumina (11). Actualmente, ONT frecuentemente obtiene una longitud de lectura promedio de 10 – 30 kb alcanzando valores de longitud máxima de lectura de 4.2 Mb (22,61). En consecuencia, la longitud de lectura en ONT depende directamente del tamaño de las moléculas presentes en la librería de secuenciación. Para ello, se han implementado diversas estrategias destinadas a la extracción y purificación de ADN de alto peso molecular (por sus siglas en inglés HMW, High Molecular Weight). Estas estrategias incluyen columnas de centrifugación (como el kit *Monarch Genomic DNA Purification* de New England Biolabs), columnas de flujo por gravedad (por ejemplo, *NucleoBond HMW DNA* de Takara Bio), perlas magnéticas (como el kit *MagAttract HMW DNA* de QIAGEN), y métodos clásicos como la extracción con fenol-cloroformo (16,62,63). Sin embargo, existe una compensación entre la longitud de lectura y el rendimiento total de secuenciación; es decir, a mayor longitud de fragmentos de ADN (HMW), menor suele ser el rendimiento en términos de cantidad de datos obtenidos (16). Esta relación aún no ha sido completamente evaluada de forma sistemática bajo diferentes condiciones experimentales, lo que resalta la necesidad de estandarizar protocolos que equilibren la integridad molecular con la eficiencia de la secuenciación.

El ADN HMW también puede fragmentarse al tamaño deseado mediante sonicación, cizallamiento hidrodinámico, extrusión con aguja o escisión por transposasa. Sin embargo, la presencia excesiva de fragmentos pequeños fuera del rango de tamaño deseado puede reducir el rendimiento del secuenciamiento, ya que estos fragmentos tienen una mayor eficiencia tanto en la ligación de adaptadores

como en su paso por los nanoporos, en comparación con fragmentos largos (16). Para eliminar estos fragmentos pequeños sobrerrepresentados, se han utilizado diversos métodos de selección por tamaño (por ejemplo, el sistema en gel *BluePippin* de Sage Science, perlas magnéticas y el kit *Short Read Eliminator* de Circulomics) con el objetivo de obtener una distribución de datos más adecuada y mejorar el rendimiento del secuenciamiento.

Una de las principales limitaciones históricas del secuenciamiento ONT ha sido su menor precisión en comparación con tecnologías de lectura corta como Illumina. Muchos de los esfuerzos recientes con los nuevos kits de secuenciación se han centrado en mejorar la calidad de lectura (Qscore), mediante actualizaciones en el enfoque de *basecalling*, así como optimizaciones en los kits de preparación de bibliotecas y en la química de las celdas de flujo (64). En sus primeras versiones, como con el poro R7.3, las lecturas 1D alcanzaban una precisión promedio de apenas el 65%, utilizando modelos estadísticos clásicos como los Hidden Markov Models (HMM) (65). Posteriormente, se desarrollaron algoritmos de *basecalling* más avanzados como Nanocall y DeepNano, que permitieron mejorar esta precisión al 70% y 78%, respectivamente (66,67). Paralelamente, ONT exploró estrategias basadas en la lectura secuencial de ambas hebras de una molécula de ADN, como en el método 2D, donde se utilizaba un adaptador en horquilla para unir ambas cadenas. Este enfoque generó una lectura de consenso (2D) con una precisión promedio de hasta el 94%. Más adelante, se introdujo la tecnología 1D<sup>2</sup>, que evitaba el uso del adaptador en horquilla pero permitía la captura consecutiva de ambas hebras mediante adaptadores especiales, alcanzando precisiones de hasta 95% con

el poro R9.5 (16). Sin embargo, debido a limitaciones técnicas y de eficiencia, ambos métodos (2D y 1D<sup>2</sup>) fueron discontinuados, y ONT centró su enfoque en una estrategia más simple y escalable: la secuenciación 1D de hebras independientes.

A pesar de la simplificación del protocolo, la precisión del secuenciamiento no se estancó. ONT impulsó mejoras significativas mediante la implementación de nuevos algoritmos de basecalling basados en redes neuronales profundas, como Guppy, Bonito y Dorado, que permitieron interpretar con mayor exactitud las señales eléctricas generadas por el paso del ADN a través del poro (16,23,68–70). Estas mejoras computacionales se potenciaron aún más con las mejoras en la preparación de su biblioteca y la evolución de la química de los poros destacando la introducción del poro R10.4 (lanzada a la venta el 2022), el cual incorpora un diseño estructural de doble constricción secuencial. Esta configuración permite que cada base nucleotídica sea leída dos veces de forma consecutiva, lo que aumenta la resolución de la señal y mejora la discriminación entre bases, especialmente en regiones difíciles como los homopolímeros (64,71,72).

Esta nueva arquitectura del nanoporo R10.4 ha contribuido significativamente a aumentar la precisión de lectura al 99% (Q20). Además, la adopción de la lectura dúplex moderna, en la que ambas hebras de una misma molécula se secuencian de forma independiente y luego se fusionan digitalmente para generar una secuencia de consenso, ha permitido alcanzar precisiones superiores al 99% (Q30) (59). En conjunto, estos avances han consolidado a ONT como una tecnología altamente

competitiva, extendiendo su aplicabilidad a proyectos genómicos de alta complejidad, como el ensamblaje *de novo* de genomas bacterianos y la identificación precisa de variantes estructurales complejas.

Adicionalmente a la longitud y precisión de la lectura de secuenciamiento, el rendimiento es otra consideración importante para aplicaciones de secuenciamiento ONT. Para responder a las necesidades de diferentes escalas de proyecto, ONT lanzó varias plataformas. Entre las plataformas de secuenciamiento de ONT más utilizadas destacan MinION, GridION y PromethION (16). MinION es un secuenciador portátil de bajo costo, compacto y fácil de llevar, que permite la secuenciación en tiempo real en diversos entornos. Esta plataforma fue diseñada para conectarse a portátiles modernos mediante su puerto USB-C y es compatible con las celdas de flujo MinION y Flongle (61). Las celdas de flujo de MinION contienen 512 canales con 4 nanopores por canal, y ha trabajado con el secuenciador MiniON Mk1B, MiniON Mk1C y actualmente con el MiniON Mk1D. Esta plataforma es ideal para estudios en campo y vigilancia epidemiológica (11,16).

Por otra parte, GridION es una plataforma de mediana escala que permite el procesamiento paralelo de hasta 5 celdas de flujo MinION. Esta plataforma es ideal para estudios metagenómicos o genómicos más demandantes (16). Por último, el PromethION es una plataforma de alto rendimiento para secuenciación masiva de proyectos de larga escala con diferentes opciones de dispositivos desde para 2 hasta 48 celdas de flujo PromethION. Las celdas de flujo PromethION contiene 3000

canales con 4 nanoporos por cada canal, lo que representa una densidad significativamente mayor en comparación con las celdas de flujo MinION (73).

Gracias a este diseño, las celdas de flujo PromethION son capaces de generar rendimientos de secuenciación mucho más alto que las celdas MinION. Un estudio independiente ha reportado hasta 153 GB de datos generados a partir de una sola celda PromethION (74) a comparación de los 10 – 15 Gb de una celda MinION (16). Esto hace que la celda de flujo PromethION sea ideal para proyectos de genómica de poblaciones, estudios del transcriptoma a gran escala, o análisis epigenómicos profundos. Esta plataforma requiere mayor capacidad computacional y está pensada para centros de investigación o instituciones que manejan grandes volúmenes de datos genómicos, consolidándose como la opción más robusta dentro del portafolio de ONT (75).

Finalmente, un componente clave que influye directamente en el rendimiento y la calidad de los datos obtenidos mediante la secuenciación ONT es el enfoque de preparación de librerías. ONT ofrece principalmente dos estrategias: el método por ligación (Ligation kit) y el método rápido mediante el uso de transposasa (Rapid kit) (64). El primero es el más robusto y se recomienda cuando se busca preservar la longitud de los fragmentos y maximizar la calidad de las lecturas, ya que emplea una ligasa para incorporar adaptadores a fragmentos de ADN de alto peso molecular. No obstante, este enfoque requiere mayor cantidad de ADN de alta pureza, entre 400 a 1000 ng con ratios de pureza 260/280 ~ 1.8 y 260/230 entre 2.0 – 2.2 (76), y un mayor tiempo de preparación (15,64,77). Por otro lado, el método

Rapid permite una preparación rápida y sencilla mediante el uso de una transposasa que fragmenta e incorpora adaptadores en un solo paso, siendo útil cuando el material biológico de alta pureza es limitado con un requerimiento de 200 ng por ADNg (78) o se requiere acelerar el flujo de trabajo, aunque con un compromiso en la longitud de las lecturas y el rendimiento total (15,64,77). La elección entre estos dos métodos debe estar alineada con los objetivos del estudio, el tipo de muestra y la escala del proyecto, lo que refuerza la versatilidad de la plataforma Nanopore para adaptarse a diversos contextos experimentales y necesidades de investigación.

### **3.6 Factores que afectan el secuenciamiento ONT**

El rendimiento del secuenciamiento con ONT puede verse significativamente afectado por etapas críticas del flujo de trabajo previo a la secuenciación del ADNg de un patógeno, siendo el proceso de inactivación térmica, el método de extracción y la fragmentación de ADNg factores críticos (17,18,20). Aunque se reconoce que la inactivación a altas temperaturas - necesaria por razones de bioseguridad en la manipulación de Mtb - degrada el ADNg (17), pocos estudios han cuantificado el efecto de esta sobre el nivel de integridad del ADNg de Mtb. Cuantificar y garantizar una alta integridad del ADNg antes del secuenciamiento ONT permite obtener lecturas largas, mayor cobertura, mejor resolución estructural, ensamblajes de alta calidad y un uso más eficiente de los recursos del flujo de secuenciación. Por ello, medir la integridad por ejemplo, mediante DIN (*DNA Integrity Number*) es una etapa crítica de control de calidad previa al secuenciamiento (14,79). El

trabajo de Chitale et. al. publicado en el 2022 ha utilizado como punto de corte un DIN score  $\geq 7$  para el secuenciamiento ONT de Mtb (14).

Por otro lado, la elección del método de extracción de ADNg es determinante para el rendimiento y la calidad de los datos de secuenciación ONT. Los métodos con fenol-cloroformo han sido históricamente valorados por su capacidad de recuperar ADN de alta integridad y alto peso molecular, lo cual es ideal para obtener lecturas largas (16). Sin embargo, este método implica riesgos de contaminación residual con fenol, una sustancia que puede interferir con la actividad de los nanoporos dentro de la celda de flujo, afectando la estabilidad de la señal iónica y fallos en la corrida (18). Por otro lado, los métodos libres de fenol, como los basados en columnas de sílica o resinas magnéticas, ofrecen protocolos más rápidos, de mejor calidad y con menor riesgo de interferencia química dentro de la celda de flujo (16). Aunque en algunos casos pueden generar fragmentos ligeramente más cortos o con menor rendimiento total, su compatibilidad directa con ONT y su menor interferencia química los convierte en una opción más robusta y confiable para secuenciación en entornos clínicos y de alta exigencia técnica. Hasta la fecha, no se han publicado estudios que demuestren de forma concluyente que las muestras extraídas mediante métodos con fenol-cloroformo presentan un menor rendimiento en el secuenciamiento ONT en comparación con los métodos libres de fenol, lo que sugiere que ambos enfoques pueden ser viables dependiendo del contexto experimental. En este sentido, es necesario poder evaluar el efecto del tipo de método de extracción en el rendimiento del secuenciamiento ONT para optimizar el flujo de trabajo de secuenciación con esta tecnología.

Finalmente, la fragmentación controlada del ADNg podría ser un paso clave para maximizar el rendimiento del secuenciamiento con ONT, particularmente en el caso de Mtb, cuyo genoma presenta un alto contenido de GC y una fuerte propensión a formar estructuras secundarias. Estas estructuras podrían bloquear los poros, afectando la continuidad de la secuenciación y disminuyendo la eficiencia de lectura (19,20). Por ello, mantener una fragmentación equilibrada podría ser esencial para el rendimiento del secuenciamiento de Mtb, considerando obtener fragmentos lo suficientemente largos para aprovechar la capacidad de ONT de generar lecturas extensas, pero también lo necesariamente cortos para evitar el bloqueo del nanoporo. Sin embargo, hasta la fecha no se han reportado estudios que evalúen de forma específica el efecto de la fragmentación del ADNg de Mtb en el rendimiento del secuenciamiento ONT, lo que resalta la necesidad de investigaciones dirigidas en este punto si se pretende mejorar el rendimiento del secuenciamiento ONT del ADNg de Mtb.

En conjunto, la optimización experimental de los tres factores clave —inactivación térmica, método de extracción y fragmentación del ADNg— no solo permitiría mejorar significativamente el rendimiento del secuenciamiento ONT, sino que también sería fundamental para su aplicación en estudios de vigilancia genómica, al generar datos de mayor calidad para el análisis integral del genoma de Mtb. Evaluar el impacto de estos parámetros permitiría establecer una condición óptima de inactivación asociada a un método de extracción y a un grado de fragmentación adecuado, que garanticen tanto un desempeño eficiente del secuenciamiento como

el cumplimiento de los estándares de bioseguridad microbiológica durante el flujo de trabajo.

### **3.7 Avances en el secuenciamiento ONT de *Mycobacterium tuberculosis***

En estos últimos 10 años, la secuenciación de Mtb mediante la tecnología de nanoporos ha surgido como una herramienta para el estudio genómico de esta bacteria, enfocándose en la vigilancia epidemiológica, detección de resistencia a fármacos y análisis estructurales del genoma. A diferencia de otras tecnologías de secuenciación, ONT permite la obtención de lecturas largas en tiempo real, lo que ha abierto nuevas posibilidades para el WGS y análisis genómico de cepas clínicas de *M. tuberculosis*, incluso en contextos de baja infraestructura o en tiempo de respuesta acelerado.

Nueve meses después del lanzamiento del programa de acceso MinION, la comunidad científica ya evidenciaba el potencial transformador de la tecnología de secuenciación por nanoporos en la investigación de la TB. En el 2015, Bradley et al. (80) llevaron a cabo un estudio de prueba de concepto en el que implementaron un enfoque basado en WGS para la predicción rápida de resistencia antimicrobiana en aislados clínicos de *M. tuberculosis* y *Staphylococcus aureus*. Este enfoque permitió detectar perfiles de resistencia directamente a partir de los datos de secuenciación, y generar un informe clínicamente interpretable en menos de tres minutos utilizando sólo una laptop y el secuenciador portátil MinION. Este trabajo marcó un precedente para la aplicación de WGS como herramienta de diagnóstico

rápido y preciso en la identificación de mutaciones asociadas a resistencia a fármacos antimicrobianos.

En el 2017, Votintseva et al. (5) evaluaron la viabilidad del uso de WGS directamente a partir de muestras respiratorias sin necesidad de cultivo previo, empleando la plataforma MinION de ONT. Este estudio utilizó un método de preparación de librerías basado en una PCR inespecífica, y lograron secuenciar con éxito el ADN de *Mycobacterium bovis* BCG a partir de una muestra clínica simulada (compuesta por ADNg extraído de esputo negativo combinado con ADN purificado de *M. bovis* BCG a diferentes concentraciones). El estudio demostró que es posible generar datos de identificación del patógeno y detección de variantes de resistencia a fármacos en un plazo aproximado de 7.5 horas, con resultados de resistencia completos disponibles en menos de 12 horas. Estos hallazgos resaltaron el potencial de la secuenciación ONT en muestras clínicas directas como herramienta de diagnóstico rápido y vigilancia genómica en TB, reduciendo significativamente los tiempos de entrega de resultados frente a los métodos tradicionales basados en cultivo. Sin embargo, los autores señalan como principales limitaciones la sensibilidad reducida en muestras con baja carga bacilar y la necesidad de optimizar la proporción de ADN micobacteriano frente al ADN humano y de otras bacterias, lo que puede impactar la profundidad de cobertura y la precisión en la detección de variantes. A pesar de estas limitaciones, el estudio constituye uno de los primeros en demostrar la aplicabilidad del secuenciamiento ONT para la obtención de datos genómicos de Mtb directamente de muestras clínicas.

En 2018, investigadores en Madagascar lograron realizar por primera vez en el país la secuenciación de ADN mediante el dispositivo portátil MinION, lo que les permitió identificar rápidamente Mtb y su perfil de resistencia a fármacos. Este éxito fue posible gracias a la portabilidad, simplicidad operativa y bajo costo del secuenciador MinION. Hasta entonces, la secuenciación de ADN aplicada a muestras de TB había estado mayormente limitada a países de altos ingresos, debido a los elevados costos y la infraestructura técnica que requieren otras plataformas de secuenciación. La secuenciación por ONT ha cambiado este panorama al ofrecer una alternativa accesible y descentralizada, permitiendo que países de ingresos bajos y medios puedan integrar la secuenciación genómica como parte de sus estrategias de diagnóstico e investigación en TB (81).

En 2018, Bainomugisa et al. (5) demostraron el valor de la tecnología de secuenciación por nanoporos al utilizar el dispositivo portátil MinION para secuenciar y ensamblar un aislado de Mtb extensamente resistente a fármacos (XDR), perteneciente a una cepa del sublinaje Beijing moderno, prevalente en la provincia Occidental de Papúa Nueva Guinea. Mediante una única celda de flujo, se obtuvo una profundidad de la cobertura (Depth of Coverage) de 238× que permitió realizar un ensamblaje *de novo* completamente contiguo con una precisión del 99,92 %, incluyendo regiones repetitivas ricas en GC como los genes *PE/PPE*. El estudio no solo evidenció la capacidad del MinION para identificar con alta certeza mutaciones de resistencia, variaciones estructurales y elementos de virulencia, sino que también resalta que la aplicabilidad de este secuenciador en

entornos con alta carga de TB podría tener un impacto positivo en el control de la TB y en los manejos de casos. Sin embargo, es importante indicar que la secuenciación de esta cepa se realizó usando una sola muestra en una sola celda de flujo lo cuál no sería práctico por el alto costo involucrado.

En el 2021, Smith et. al. (82) evaluaron la viabilidad del uso del dispositivo MinION para el diagnóstico clínico de Mtb, comparándolo con Illumina MiSeq en términos de identificación de especies, tipificación genómica, predicción de resistencia antimicrobiana y análisis filogenético. Este estudio se consideró el estudio clínico más largo evaluando 431 aislados clínicos. El rendimiento del MinION alcanzó una concordancia del 98,1 % con Illumina en predicción de resistencia antimicrobiana y una concordancia del 96 % con los resultados fenotípicos de susceptibilidad a fármacos. Además, este tipo de secuenciamiento presentó ventajas logísticas significativas, incluyendo un menor costo por muestra (~\$63 vs. \$130 de Illumina) y menor tiempo de respuesta. La data del secuenciamiento ONT destacó por su precisión en genotipificación y detección de linajes. Sin embargo, se presentaron limitaciones para detectar variantes heterocigotas o inserciones/deleciones pequeñas. Estas debilidades se atribuyen a su mayor tasa de error de lectura propia de la celda de flujo R9.4.1. Pese a ello, los hallazgos sugieren que con flujos de trabajo bioinformáticos adecuados, la secuenciación Nanopore tiene el potencial de complementar o incluso reemplazar a las plataformas de lectura corta en entornos de salud pública con limitaciones de recursos. Este trabajo se consolidó como un antecedente relevante para investigaciones centradas en la implementación de esta tecnología en la vigilancia genómica de Mtb.

En el 2020 Cervantes et. al. (83) tenían como objetivo validar el kit de secuenciación rápida ONT para la identificación genómica y el perfil genético de mutaciones a drogas de aislados

clínicos de Mtb y ADN proveniente de muestras de esputo, con resultado positivo de baciloscopía, utilizando el secuenciador portátil MinION. Entre los hallazgos más relevantes, se demuestra que el sistema ONT en conjunto con el kit de preparación de librería fue capaz de detectar mutaciones conocidas de resistencia incluso con un número bajo de lecturas, evidenciando una sensibilidad adecuada y rapidez en la obtención de resultados. Sin embargo, se identificaron limitaciones, como la baja proporción de ADN de Mtb en esputo debido a la abundancia de ADN humano, la alta tasa de error inherente a la tecnología Nanopore, y la falta de curación en la base de datos de referencia utilizada para el análisis automatizado, lo cual generó resultados con posibles falsos positivos. A pesar de esto, el estudio concluyó que mejoras futuras en la tasa de error a través de análisis estadísticos, algoritmos de predicción de resistencia a fármacos y bases de datos de referencia harían de esta plataforma una opción adecuada para el entorno clínico. El tamaño pequeño, el costo relativamente bajo del dispositivo, así como su protocolo de preparación de librería con el kit rapid y análisis rápido y simple, lo convierten en una opción atractiva para entornos con infraestructura de laboratorio limitada.

En el 2020 George et. al. (17) realizaron el primer estudio que utilizó secuenciación ONT en muestras clínicas (sin cultivo previo) como lavados broncoalveolares y aspirados de ganglios linfáticos. El estudio logró una cobertura genómica de Mtb que varió entre 0.55 X y 81X, dependiendo de la muestra. Cuando usaron un programa especializado llamado Nanopolish, que analiza los datos de ONT a un nivel más detallado (nivel de señal), consiguieron una precisión de consenso altísima de 99.9%. Por otro lado, el estudio enfrentó problemas cuando intentaron multiplexar (secuenciar varias muestras juntas en una sola corrida o celda de flujo).

Entre el 5% y el 47% de las lecturas no pudieron ser asignadas con confianza a ninguna muestra específica, lo cual afectó la cobertura del genoma. Los mejores resultados se obtuvieron sólo cuando se usó una celda de flujo por muestra individual, lo cual hace que el método sea demasiado costoso para aplicarlo de manera rutinaria en clínicas u hospitales.

En 2020 se publicaron tres estudios clave que evaluaron el uso de la secuenciación por Nanopore (MinION) para la identificación dirigida de mutaciones asociadas a la resistencia a fármacos en *Mtb* (84–86). Tafess et al. (84) desarrollaron y aplicaron paneles personalizados de 19 genes relacionados con resistencia. Estos fueron secuenciados bajo dos flujos de trabajo ONT (MinION) e Illumina (MiSeq) con una concordancia del 100 % entre ambos en la detección de variantes resistentes, excluyendo las variantes notificadas por ONT con una frecuencia alélica menor al 40 %. Ambos flujos de trabajo alcanzaron una sensibilidad y especificidad clínica media del 94,8% y 98,0% respectivamente, en comparación con la prueba fenotípica de sensibilidad a fármacos (pDST, phenotypic drug susceptibility tests). Los tiempos de respuesta de MiSeq y MinION fueron de 38 y 15 horas, al menos 17-18 días antes que la pDST, respectivamente. El costo de secuenciamiento por muestra con MinION (\$71.56) frente a MiSeq (\$67.83) se debió a su menor capacidad de procesamiento por lotes. El estudio evidenció que ambas plataformas son intercambiables para obtener resultados precisos y útiles en el tratamiento de la TB.

Por otra parte, Chan et al. (85) desarrollaron y aplicaron paneles personalizados de 10 genes relacionados con resistencia a fármacos antituberculosos, aplicado directamente a 23 muestras clínicas con baja carga bacilar. Al menos 8 loci fueron amplificados con éxito en el 60.9 % de las muestras, y las 12 muestras que lograron la amplificación de los 10 genes fueron secuenciadas mediante ONT MinION y MiSeq. El estudio mostró una concordancia del 95.1 % para variantes detectadas por ONT con una frecuencia alélica del 100 % notificadas por MiSeq. Además, se destacó que el tiempo de respuesta fue significativamente más corto con MinION (15 h) en comparación con MiSeq (38 h), lo que representa una ventaja clínica relevante en contextos donde se requiere una toma de decisiones rápidas. Aunque este enfoque mostró buena concordancia, se evidenció una menor sensibilidad en la detección de variantes con frecuencias alélicas bajas, especialmente en muestras con carga bacilar limitada. No obstante, el bajo tiempo de ensayo, el costo razonable de reactivos (\$64 por muestra al procesar 24 muestras por celda de flujo) y las mejoras continuas en la química de secuenciación y en los algoritmos de corrección de señal (como Nanopolish) respaldan el potencial clínico del método basado en Nanopore para el perfilamiento directo de resistencia en TB.

Por su parte, en el 2020 Cabibbe et al. (86) evaluaron el desempeño del ensayo GenoScreen Deeplex Myc-TB, una solución basada en tNGS, aplicada a 104 muestras de esputo bacilosκόpico positivo previamente analizadas con Illumina MiniSeq. Luego de una preparación personalizada de librerías, las muestras fueron secuenciadas con ONT MinION, alcanzando un porcentaje medio de identidad de las lecturas MinION asignadas al genoma diana de H37RV de 90,8% (comparado

con el 99,5 % en MiniSeq) y una cobertura media de profundidad superior a 4.100× en ambas plataformas. A pesar de la mayor tasa de error bruta en ONT, se observó una concordancia completa en la detección de marcadores clínicamente relevantes de resistencia a fármacos cuando se aplicó un umbral de frecuencia alélica del 80 %. Además, los costos por muestra fueron similares entre ambas tecnologías, aproximadamente 100 euros, destacándose las ventajas del MinION en cuanto a portabilidad, versatilidad y bajo costo de capital. Este estudio respalda el potencial de la secuenciación dirigida con MinION como herramienta viable para el diagnóstico molecular rápido de TB-MDR en entornos descentralizados.

Por otro lado en el 2023, un estudio realizado por Murphy *et al.* (6). Presentaron un enfoque de secuenciación dirigida y multiplex utilizando la plataforma ONT, comparando sus resultados con los obtenidos mediante WGS. Los autores demostraron la utilidad de un ensayo de secuenciación dirigida y multiplex capaz de predecir la resistencia a los fármacos contra Mtb directamente en muestras respiratorias clínicas. Para ello se diseñó una PCR multiplex para amplificar un grupo de trece genes completos y regiones promotoras con mutaciones asociadas a la resistencia a fármacos contra Mtb, estos genes fueron secuenciados mediante ONT e interpretados en tiempo real mediante un pipeline bioinformático propio. El análisis se aplicó a 72 muestras clínicas y 55 cultivos positivos para MTBC, y los resultados fueron comparados con los obtenidos por WGS. El tNGS permitió generar perfiles completos o parciales de resistencia en el 82 % de las muestras con baciloscopia positiva, mostrando una concordancia del 100 % con los resultados de WGS. Entre sus limitaciones destaca una menor aplicabilidad en muestras con

carga bacilar baja. Este enfoque permitió el diagnóstico de farmacoresistencia sin necesidad de cultivo, reduciendo significativamente el tiempo de respuesta a 15 días antes que los resultados de WGS derivados de cultivo. En conclusión, este estudio demuestra que la secuenciación dirigida mediante ONT es una herramienta confiable y rápida para predecir la resistencia a fármacos en Mtb, ofreciendo información crítica y oportuna en casos TB - DR.

Los últimos cuatro estudios han demostrado la viabilidad de la tecnología Nanopore, particularmente mediante secuenciación dirigida con MinION, para predecir la farmacoresistencia en Mtb directamente desde muestras clínicas. Esta estrategia permite identificar de forma rápida y precisa mutaciones en genes asociados a resistencia, siendo especialmente útil en contextos con alta prevalencia de TB - MDR. Sin embargo, su enfoque limitado al análisis de regiones específicas no permite una visión genómica completa. En contraste, WGS ofrece una perspectiva más amplia al detectar no solo variantes de resistencia, sino también al permitir el estudio de la evolución genómica, la diversidad poblacional, la transmisión y otros aspectos epidemiológicos esenciales. Además, el WGS posibilita la exploración de regiones poco estudiadas como los genes PE/PPE, implicados en la interacción hospedero-patógeno y en la respuesta inmune. Estas regiones, difíciles de cubrir con tecnologías como Illumina, pueden ser mejor analizadas mediante tecnologías de lectura larga. Por tanto, aunque la secuenciación dirigida con Nanopore es una herramienta valiosa para el diagnóstico clínico, el WGS continúa siendo indispensable para la investigación integral de la TB.

En este sentido, en el 2022 Gómez-González et al. (21) compararon 10 aislamientos clínicos de Mtb secuenciados con ONT e Illumina. El análisis mostró que ONT, utilizando celdas de flujo R10.3 y un kit de preparación de librerías basado en ligación, permitió una detección confiable de polimorfismos de nucleótido único, pese a su alta tasa de error. Además, este análisis de datos ONT pudo incorporar regiones del genoma que normalmente se excluyen (por ejemplo, los genes *pe/ppE*). Esta resolución adicional puede mejorar la caracterización de clústeres de transmisión y su dinámica, al permitir inferencias más precisas entre aislamientos estrechamente relacionados. La alta concordancia en variantes en loci asociados con resistencia a fármacos respalda su uso para la detección rápida de mutaciones resistentes. En conjunto, la secuenciación ONT representa una herramienta prometedora para investigaciones genómicas de TB, especialmente para apoyar decisiones clínicas y de vigilancia que contribuyan a reducir la carga de la enfermedad.

Un año después, en el 2023 Di Marco *et al.* (13) analizaron las ventajas y limitaciones de la secuenciación de lecturas cortas (SLC) de Illumina, secuenciación de lecturas largas (SLL) de ONT y un enfoque híbrido (HIBR) en la caracterización del genoma de Mtb. Se evaluó cómo cada método influye en la estimación de la cobertura genómica, llamada de variantes y el análisis de cluster, la detección de resistencia a fármacos y el ensamblaje de novo del genoma. El estudio demostró que ONT ofrece una excelente cobertura del genoma de Mtb,

incluyendo regiones repetitivas como los genes PE/PPE. En la llamada de variantes y el análisis de clústeres, ONT identificó correctamente mutaciones y agrupamientos epidemiológicos, aunque su tasa de error puede limitar la detección de variantes de baja frecuencia; esto puede mitigarse mediante una mayor profundidad de secuenciación. En cuanto a la detección de resistencia a fármacos, ONT mostró alta concordancia con los perfiles de resistencia definidos por SLC. Finalmente, en el ensamblaje *de novo*, ONT superó claramente a SLC, generando ensamblajes más completos y continuos. El estudio concluyó que ONT es una herramienta robusta y versátil para la investigación genómica de Mtb, siendo útil cuando se prioriza la cobertura total del genoma, la exploración de regiones complejas y la resolución de la arquitectura genómica. Sin embargo, su uso óptimo en entornos clínicos o de vigilancia requiere considerar la profundidad de secuenciación, y en algunos casos, combinarla con tecnologías complementarias para maximizar la precisión.

Los estudios realizados hasta la fecha sobre Mtb mediante tecnología ONT se han centrado principalmente en la en la detección de variantes genéticas y resistencia a fármacos, sin abordar la optimización de protocolos de secuenciación. A la fecha, no se han reportado investigaciones que evalúen el rendimiento de ONT usando las celdas de flujo R10.4.1 junto con la química de librería Kit 14, a pesar de su potencial para generar lecturas crudas con 99% de precisión y lecturas dúplex hasta 99.9%. La ausencia de estos estudios limita el aprovechamiento total de las capacidades actuales de esta plataforma. Optimizar el secuenciamiento permitiría mejorar la cobertura, la profundidad de cobertura genómica y la precisión en la

llamada de variantes. Además, estandarizar estas condiciones contribuiría a implementar protocolos de alto rendimiento. Esto favorecería la vigilancia genómica y facilitaría la adopción de ONT en salud pública. Por ello, los estudios de optimización son esenciales para consolidar el uso del WGS con ONT en TB.

## **4 MATERIALES Y MÉTODOS**

### **4.1 Diseño de estudio**

El estudio se ejecutó bajo un diseño exploratorio y experimental con el objetivo de evaluar el efecto de diferentes condiciones de inactivación térmica, métodos de extracción y la fragmentación de ADNg en el rendimiento del secuenciamiento de Mtb. Las muestras fueron procesadas en el siguiente orden: Inactivación térmica, Méodos de extracción y Fragmentación. La inactivación se ejecutó bajo dos temperaturas de inactivación 80°C por 20 min, 95°C por 60 minutos y el grupo de tratamiento control NT (no tratado) al cual no se le realizó inactivación térmica. Para cada condición térmica se usó 24 muestras (Tabla 1) las cuales se distribuyó equitativamente para cuatro métodos de extracción de ADNg, 6 muestras por cada método de extracción (Tabla 1).

La evaluación del efecto de la temperatura de inactivación en la cantidad, calidad e integridad del ADNg (objetivo 1) se realizó comparando los tres grupos de tratamiento térmico en los cuatro métodos de extracción, con un n= 6 réplicas por cada tratamiento térmico. Por otro lado, la evaluación del efecto del método de extracción en la cantidad, calidad e integridad del ADNg (objetivo 2) se realizó comparando los cuatro métodos de extracción en los tres tratamientos térmicos, con un n= 6 réplicas por cada método de extracción.

La evaluación del efecto de la temperatura de inactivación en el rendimiento del secuenciamiento (objetivo 3) se realizó comparando dos grupos de tratamiento térmico: 80°C por 20 min y el grupo control NT en los cuatro métodos de extracción. Por cada tratamiento térmico se usó un n= 3 réplicas. De la misma forma, la evaluación del efecto del método de extracción en el rendimiento de secuenciamiento (objetivo 4), se realizó comparando el rendimiento de los cuatro métodos de extracción de ADNg tanto para el tratamiento térmico de 80°C por 20 min como para el grupo control NT. Por cada método de extracción se usó un n= 3 réplicas.

Por último, para la evaluación del efecto de la fragmentación del ADNg a 10 kpb en el rendimiento del secuenciamiento (objetivo 5) se comparó dos grupos de condiciones : No Fragmentado (NF) y Fragmentado (F) en tres métodos de extracción. Por cada condición de fragmentación se usó un n= 6 réplicas.

El número limitado de réplicas en los análisis de secuenciamiento en el tercer y cuarto objetivo (n=3 por cada grupo de comparación) se justificó debido al costo que implica el proceso secuenciamiento. Por lo que este estudio al tener un enfoque exploratorio pretende brindar indicios clave del efecto de estas variables tanto en el rendimiento de la extracción del ADNg de Mtb como del secuenciamiento de genoma completo de este. En adelante se explicarán a más detalle las metodologías que se usaron en el desarrollo de este estudio.

## **4.2 Lugar de estudio**

El proyecto de investigación se realizó en el Laboratorio de Bioinformática y Biología Molecular y en el Laboratorio de Mycobacterium pertenecientes al LID (Laboratorios de Investigación y de Desarrollo) de la Universidad Peruana Cayetano Heredia (UPCH).

## **4.3 Material Biológico**

Para la ejecución del estudio se empleó la cepa de *M. tuberculosis* H37Rv, perteneciente al grupo de riesgo 3. Esta cepa de referencia se obtuvo del banco de cepas del Laboratorio de Bioinformática y Biología Molecular del LID de la UPCH.

## **4.4 Equipo de Protección Personal (EPP) e instalaciones**

Los procedimientos que implicaron el uso de Mtb se realizaron en el Laboratorio de Mycobacterium dentro de una cabina de bioseguridad de clase II tipo A/B3 (certificada) usando el siguiente equipo de protección : 3 pares de guantes, respirador N95, mandil azul, cubre-zapatos y gorro descartables, y mandil verde, cumpliendo con los protocolos de contención y seguridad establecidos. Las superficies y materiales de trabajo en el Laboratorio de Mycobacterium fueron descontaminadas con el desinfectante TEGO al 5%. Todos los procesos posteriores a la extracción de ADN se realizaron en el Laboratorio de Bioinformática y Biología Molecular usando el siguiente equipo de protección: Bata de laboratorio, 1 par de guantes, mascarilla descartable.

#### **4.5 Preparación de muestras**

La cepa Mtb H37Rv se descongeló y reactivó durante una semana en medio líquido (caldo Middlebrook 7H9 suplementado con OADC - ácido oleico, albúmina, dextrosa y catalasa- ) y posteriormente se sembró en placas 7H10 (BD Difco, EE.UU.), con 0.5% (v/v) de glicerol y 10% de OADC. Las placas fueron incubadas a 37 °C durante 3 semanas.

#### **4.6 Inactivación de la muestra**

Para la inactivación de la cepa H37Rv, 72 muestras fueron preparadas. Para cada muestra 1 asada de Mtb (~20 - 30 mg) se resuspendió en 500 µl de Buffer TE (10mM Tris HCl, pH 8 y 1mM EDTA pH 8). Todas las muestras fueron tratadas en un termobloque bajo dos temperaturas de inactivación -95 °C por 1h y 80 °C por 20 min- excepto su respectivo control al que no se le inactivó con calor (No Tratado: NT). La distribución del número de muestras por cada temperatura de inactivación se detalla en la Tabla 1. Para confirmar la eficacia de la inactivación térmica, se sembraron 30 µL de tres réplicas independientes de cada tratamiento, así como del control, en placas 7H10 (según el procedimiento previamente descrito). Posteriormente, las placas se incubaron a 37 °C durante 3 meses para evaluar cualitativamente la presencia de crecimiento bacteriano.

#### **4.7 Extracción y purificación del ADN**

El ADN genómico (ADNg) se extrajo empleando tres métodos diferentes, uno de los cuales presentó dos variaciones (Métodos: 1, 2, 3a y 3b), que se describen detalladamente más adelante. Cada uno de estos métodos se aplicó a 18 muestras, distribuidas en seis réplicas por cada tratamiento de temperatura. Tras la extracción, se evaluaron los siguientes parámetros: cantidad de ADN ( $\mu\text{g}$ ) mediante fluorimetría con Qubit 4; calidad del ADN, a través de las relaciones de absorbancia 260/280 y 260/230 en el espectrofotómetro NanoDrop 2000; e integridad del ADN, utilizando un TapeStation 4150, evaluada mediante tres métricas: DNA Integrity Number (DIN), tamaño modal de los fragmentos de ADN (Peak Size, kb) y porcentaje de fragmentos  $> 5$  kpb.

Para la secuenciación ONT se utilizaron muestras ADNg de H37Rv que cumplieran los siguientes criterios: cantidad de ADNg  $\geq 2$   $\mu\text{g}$  considerando que se necesite una purificación adicional para obtener 1  $\mu\text{g}$  de ADN necesario para ser procesado con el kit Native Barcoding basado en ligación para la preparación de librerías en el secuenciamiento ONT (15,87). Además de ratios 260/280 = 1.8-2.0; 260/230 = 2.0-2.2 y un DIN  $\geq 7$  (14,79) que fueron consideradas aptos para previos estudios del secuenciamiento ONT. Las muestras de ADNg fueron purificadas adicionalmente utilizando un kit DNA Clean & Concentrator -5 de acuerdo a las especificaciones del fabricante (Zymo Research Part Number D4004) para cumplir con los requisitos

de pureza establecidos para la secuenciación. El ADN purificado se almacenó a – 20 °C hasta su uso en la preparación de las bibliotecas de secuenciación.

**Tabla 1.** Número de muestras por tratamiento térmico

Tratamiento térmico	Número de muestras
80 °C durante 20 minutos	n=24 Método 1 (n= 6 réplicas) Método 2 (n= 6 réplicas) Método 3a (n= 6 réplicas) Método 3b (n= 6 réplicas)
95 °C durante 60 minutos	n=24 Método 1 (n= 6 réplicas) Método 2 (n= 6 réplicas) Método 3a (n= 6 réplicas) Método 3b (n= 6 réplicas)
Control no tratado (NT)	n=24 Método 1 (n= 6 réplicas) Método 2 (n= 6 réplicas) Método 3a (n= 6 réplicas) Método 3b (n= 6 réplicas)

### **Método 1: Método fenol-cloroformo con lisis enzimática**

Se utilizó el método estándar de extracción basado en fenol y bromuro de hexadeciltrimetilamonio (CTAB) (88) con ligeras modificaciones. Brevemente, para la lisis de la pared celular y de la membrana, se añadieron 50  $\mu$ L de lisozima (10 g/mL) a cada muestra, las cuales se incubaron a 37 °C durante toda la noche. Posteriormente, se agregaron 70  $\mu$ L de SDS (20 %) y 20  $\mu$ L de proteinasa K (20 mg/mL), y las muestras se incubaron a 65 °C durante 3 horas, realizando agitación en vórtex durante 5 segundos cada hora. A continuación, se añadieron 100  $\mu$ L de NaCl (5 M) y 100  $\mu$ L de solución CTAB/NaCl (0,7 M NaCl, 10 g de CTAB), incubándose la mezcla durante 10 minutos a 65 °C. La purificación del ADN se realizó añadiendo 750  $\mu$ L de fenol:cloroformo:alcohol isoamílico (25:24:1), seguido de centrifugación a 10.000 g durante 5 minutos. El sobrenadante obtenido se transfirió a un nuevo tubo, donde se realizó una segunda purificación utilizando 600  $\mu$ L de cloroformo:alcohol isoamílico (24:1), seguida de centrifugación a 10.000 g durante 5 minutos. El sobrenadante, que contenía el ADN, se transfirió nuevamente a un tubo limpio, se añadió 1 mL de etanol absoluto (-20 °C) y se homogeneizó manualmente. Posteriormente, las muestras se enfriaron a -70 °C durante 1 hora para precipitar el ADN. La precipitación se completó mediante centrifugación a 13.000 rpm durante 15 minutos; el sobrenadante fue eliminado y el pellet de ADN se conservó. El pellet se lavó añadiendo 1 mL de etanol frío al 70 %, seguido de centrifugación a 13.000 rpm durante 5 minutos. Después de eliminar el sobrenadante, el pellet se dejó secar a temperatura ambiente. Finalmente, el ADN se resuspendió en 100  $\mu$ L de buffer TE, las muestras se

incubaron a 55 °C durante 15 minutos, se homogeneizaron suavemente y se almacenaron a -20 °C hasta su uso posterior.

### **Método 2: Método fenol-cloroformo con lisis mecánica previo a lisis enzimática**

Se adaptó un método propuesto por Bouso et al. en el 2019 (63), el cual utilizó lisis mecánica mediante rompimiento con microesferas previo a un rompimiento enzimático y el fenol. Tras el tratamiento de las muestras con diferentes condiciones de calentamiento, estas se centrifugaron durante 10 minutos a 4500 rpm y se descartó el sobrenadante. Se añadieron 400 µL de buffer de lisis y 100 µL de SDS al 20%, y la mezcla se homogenizó mediante pipeteo. Posteriormente, las muestras se agitaron con 150 mg de microesferas de sílice de 0,1 mm de diámetro durante tres ciclos de 30 segundos a 10,5 m/s usando un equipo FastPrep. Se realizó una lisis adicional mediante la adición de 50 µL de lisozima (10 mg/mL), incubándose a 37°C durante 60 minutos. Luego, se añadieron 40 µL de proteinasa K (200 µg/mL) y se incubó a 37°C durante 90 minutos, mezclando por inversión cada 30 minutos. Las muestras se centrifugaron primero a 4500 rpm durante 10 minutos y luego a 14.000 rpm durante 2 minutos. El sobrenadante se transfirió a un nuevo tubo, al que se añadieron 500 µL de fenol:cloroformo:alcohol isoamílico (25:24:1, tamponado a pH 8,0), homogeneizándose por inversión durante 10 minutos para mezclar las fases acuosa y orgánica. La mezcla se centrifugó a 13.000 rpm durante 10 minutos y la fase acuosa superior (~200–450 µL), que contenía el ADN, se transfirió a un nuevo tubo de 2 mL. Dependiendo del volumen recuperado, se añadió 1/10 del volumen de NaCl 5M, seguido de 1 volumen de isopropanol, y se incubó a temperatura ambiente durante toda la noche. Posteriormente, se centrifugó a 14.000

rpm por 30 minutos a temperatura ambiente. El pellet obtenido se lavó con 700  $\mu$ L de etanol al 70% y se centrifugó nuevamente a 14.000 rpm por 10 minutos (repetiendo el lavado una vez más). Finalmente, se dejó evaporar el etanol restante a temperatura ambiente durante 15 minutos con las tapas de los tubos abiertas. El ADN fue eluido en 100  $\mu$ L de buffer EB previamente calentado a 65°C durante 10 minutos, y se almacenó inmediatamente a 4°C.

**Método 3: 3a y 3b. Kit comercial - Método 3a : Método del kit ZymoBIOMICS con lisis enzimática y Método 3b: Método del kit ZymoBIOMICS con lisis mecánica**

Se utilizó un kit comercial que emplea la lisis mecánica mediante rompimiento por microesferas y la purificación por columnas (89). Se evaluaron dos variaciones del paso de lisis de este protocolo: una modificación en la duración de la lisis mecánica y otra en la que esta fue sustituida por una lisis enzimática. Tras el tratamiento de las muestras con diferentes condiciones de calentamiento, estas se centrifugaron durante 10 minutos a 4500 rpm y se descartó el sobrenadante. Luego, se añadió 250  $\mu$ L de buffer TE y se mezclaron por pipeteo. La extracción de ADN se llevó a cabo utilizando el kit ZymoBIOMICS™ DNA Miniprep (Zymo Research, Part Number D4300), siguiendo las instrucciones del fabricante, con las siguientes modificaciones: el volumen de entrada fue de 250  $\mu$ L por muestra. Para el paso de lisis, se aplicaron dos variaciones. En el método 3a, se reemplazó la lisis mecánica por una lisis enzimática (no contemplada en el protocolo original y orientada a preservar la integridad del ADN). Las muestras se incubaron durante una noche con lisozima (concentración final de 1 mg/mL), seguida de una incubación de 3 horas

con proteínasa K (concentración final de 0,6 mg/mL). En el método 3b, se empleó lisis mecánica mediante las microesferas, proporcionadas por el kit, durante 20 minutos a 3200 rpm. Ambas variantes continuaron con el protocolo estándar del kit para las etapas subsiguientes de unión del ADN, lavado y elución mediante columna de centrifugación.

#### **4.8 Fragmentación del ADN**

Tres muestras por cada condición de inactivación térmica dentro de los métodos extracción de ADN<sub>g</sub> fueron fragmentadas a aproximadamente 10 kbp utilizando el kit de fragmentación g-TUBES de Covaris (número de parte 520079), siguiendo el protocolo del fabricante con la siguiente modificación: se centrifugaron 2 µg de ADN por muestra a 4200 rpm durante 1 minuto. Debido a la disponibilidad limitada de unidades g-TUBES, solo se incluyeron los métodos M1, M2 y M3b en la etapa de fragmentación. Posteriormente, las muestras fragmentadas se concentraron utilizando el kit DNA Clean & Concentrator, de acuerdo con las instrucciones del fabricante, y se evaluaron la pureza y la integridad tal como se describió previamente.

#### **4.9 Preparación de librerías para secuenciar**

Para la etapa de secuenciamiento, se diseñó una estrategia experimental orientada a aprovechar la capacidad de multiplexado de la celda de flujo, considerando un total de 12 barcodes por celda de flujo. Con este fin, se estableció un esquema de

comparación que incluyó dos métodos de extracción y dos condiciones térmicas por celda, empleando tres réplicas biológicas por condición. Debido a esta restricción de capacidad, únicamente se seleccionaron para secuenciamiento las condiciones de 80 °C/20 min (condición usada en nuestro grupo trabajo) y NT. La condición de 95 °C/1 h no fue incluida en esta fase, a fin de mantener la estructura experimental y la comparabilidad entre métodos bajo un diseño balanceado.

El efecto de la temperatura de inactivación (80 °C/20 min) y el método de extracción en el rendimiento del secuenciamiento nanopore del ADNg de Mtb se evaluó en los métodos de extracción planteados 1, 2, 3a y 3b sin fragmentación inducida a 10 kpb del ADNg. Por otro lado, el efecto de la fragmentación inducida a 10 kpb del ADNg de Mtb en el rendimiento del secuenciamiento se evaluó sólo en los métodos de extracción 1, 2 y 3b (Tabla 2).

Se preparó un total de 4 librerías con 12 muestras por librería, las muestras fueron distribuidas equitativamente entre los métodos de extracción y las condiciones de tratamiento térmico (80°C/20 min y NT). Las muestras se agruparon como se detalla en la Tabla 2. Cada librería se preparó utilizando el Native Barcoding Kit (Oxford Nanopore, SQK-NBD 114.24) de acuerdo con las instrucciones del fabricante. Se utilizó una celda de flujo R10.4.1 por cada librería en la plataforma GridION. Tras preparar y normalizar el pool de cada librería, se cargó 18 fmoles de cada pool en la celda de flujo 1, 2, 3 y 4 para la secuenciación (Tabla 2) hasta alcanzar un porcentaje de poros activos del 5%.

**Tabla 2.** Secuenciamiento de las flow cells

<b>ADNg</b>	<b>Extracción con Fenol</b> (Método 1 y Método 2)	<b>Extracción sin fenol</b> (Método 3a y Método 3b)
<b>No fragmentado a 10 kpb</b>	<b>Flow cell 1: Fenol No Fragmentado</b> Método 1 (80°C n=3 , NT n=3) Método 2 (80°C n=3 , NT n=3)	<b>Flow cell 2: Kit comercial No Fragmentado</b> Método 3a (80°C n=3 , NT n=3) Método 3b (80°C n=3 , NT n=3)
<b>Frgamentado a 10 kpb</b>	<b>Flow cell 3: Fenol Fragmentado</b> Método 1 (80°C n=3 , NT n=3) Método 2 (80°C n=3 , NT n=3)	<b>Flow cell 4: Kit comercial Fragmentado</b> Método 3b (80°C n=6 , NT n=6)

n: número de réplicas

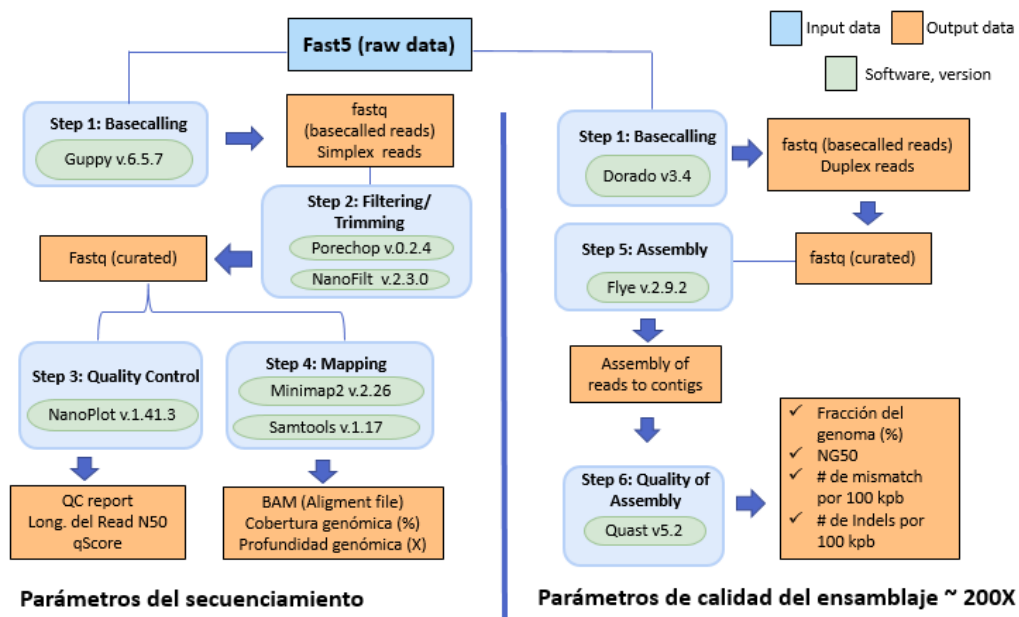
#### **4.10 Rendimiento del secuenciamiento**

El rendimiento del secuenciamiento se evaluó en dos niveles en función de los siguientes parámetros: en un primer nivel los parámetros del secuenciamiento en "*simplex reads*" : *qScore* (precisión de lectura), *read N50*, *Cobertura genómica*, *Profundidad genómica* y en un segundo nivel los parámetros de calidad de ensamblaje (en aquellas condiciones cuya profundidad genómica haya alcanzado los 200X): fracción del genoma cubierto, NG50, número de *mismatch* y número de *indels* en "*duplex reads*" con el objetivo de incrementar la precisión de lectura.

#### 4.11 Análisis bioinformático

Las métricas de análisis post secuenciamiento en "*simplex reads*" se obtuvieron de la siguiente manera: los archivos FASTQ fueron generados usando Guppy basecaller versión 6.5.7 de ONT, respetando criterios como una puntuación de qScore mínima de 10 y una longitud de 200 pb para garantizar la fiabilidad de los datos para el análisis bioinformático. El análisis de los datos FASTQ se realizó a través de un canal bioinformático especialmente diseñado en la plataforma NextFlow. El control de calidad inicial se llevó a cabo con la versión 1.41.3 de Nanoplot. Para el posterior mapeo de lecturas, se utilizó la versión 2.26 de Minimap2, elegida por su eficiencia de alineación y precisión con secuencias largas; y las lecturas fueron alineadas con el genoma de referencia bien anotado Mtb H37Rv (NC\_000962.3). La versión 1.17 de Samtools facilitó el análisis cuantitativo de la cobertura y profundidad genómica de las secuencias mapeadas (Figura 3).

Los parámetros del ensamblaje se trabajaron en "*duplex reads*". Estos fueron obtenidos utilizando Dorado basecaller versión 3.4, y fueron filtrados en función de las etiquetas "*duplex*" y "*simplex*" utilizando Samtools. El ensamblaje se realizó mediante Flye versión 2.9.2. Las métricas de ensamblaje fueron calculadas por Quast versión 5.2 utilizando H37Rv como genoma de referencia (Figura 3).



**Figura 3.** Pipeline de procesamiento y ensamblaje genómico de datos ONT (lecturas simplex y dúplex).

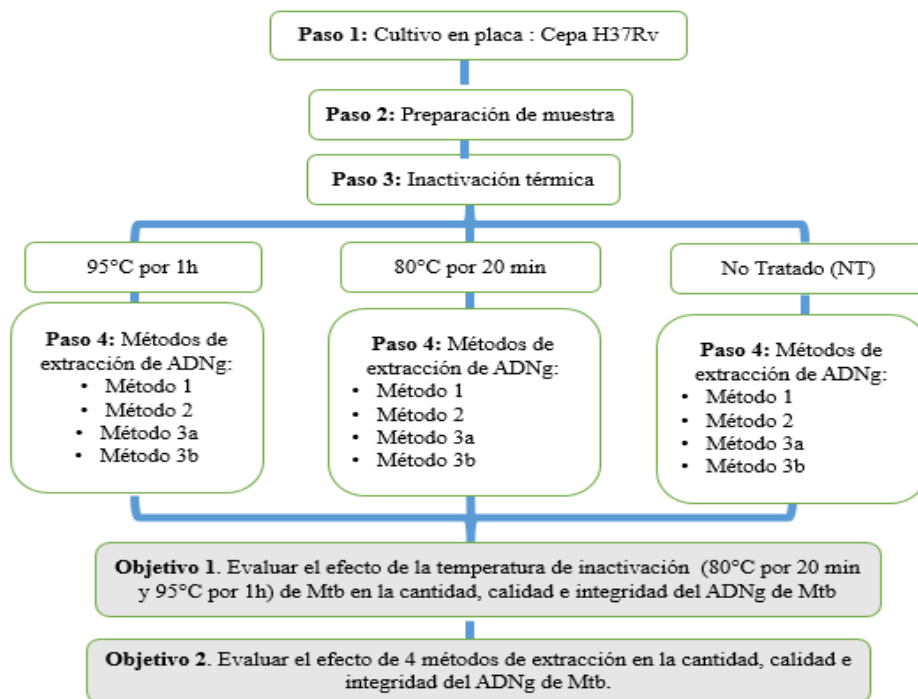
#### 4.12 Flujograma de trabajo y análisis

El siguiente flujograma de trabajo describe las etapas clave del proceso experimental que se realizará para evaluar el efecto de la temperatura de inactivación, extracción y fragmentación del ADNg de Mtb tanto en el rendimiento del ADNg de Mtb como en el secuenciamiento de este.

Para la evaluación del efecto de la temperatura de inactivación de Mtb en la cantidad, calidad e integridad del ADNg se consideraron dos temperaturas de inactivación: 80°C por 20 minutos, 95°C por 1 hora y el control NT. Esta evaluación se realizó en los cuatro métodos de extracción planteados: Método 1, 2, 3a y 3b (Figura 4). El objetivo fue determinar cómo la temperatura de inactivación afecta

la cantidad, calidad e integridad del ADNg, los cuales son factores fundamentales para la secuenciación de alto rendimiento.

Para la evaluación del efecto del método de extracción de ADNg de Mtb en la cantidad, calidad e integridad del ADNg se consideraron todos los métodos de extracción evaluados 1, 2, 3a y 3b. Esta evaluación se realizó en las muestras en las que se aplicó las temperaturas de inactivación de 95°C por 1 hora, 80°C por 20 minutos y al grupo control NT (Figura 4). El objetivo fue determinar cómo el método de extracción afecta la cantidad, calidad e integridad del ADNg.



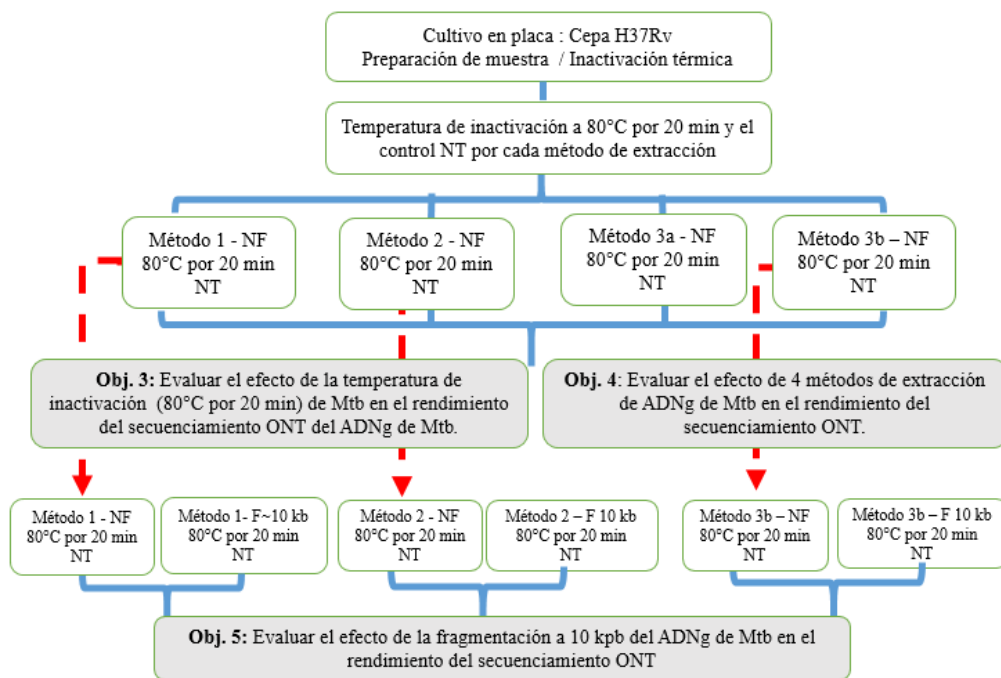
**Figura 4.** Flujograma del análisis de la evaluación del efecto de la temperatura de inactivación de Mtb (80°C por 20 min y 95°C por 1 hora) y el método de extracción de ADNg en la cantidad, calidad e integridad del ADNg.

Cada método de extracción se realizó con 6 réplicas.

La evaluación del efecto de la temperatura de inactivación 80°C por 20 min en el rendimiento del secuenciamiento del ADNg de Mtb usando ONT se realizó en los

métodos de extracción planteados 1, 2, 3a y 3b sin fragmentación inducida a 10 kpb del ADNg (Figura 5). Como control negativo se usó aquellas muestras a las que no se le aplicó el tratamiento térmico –NT–.

Para la evaluación del método de extracción en el rendimiento del secuenciamiento del ADNg de Mtb usando ONT se consideraron los métodos de extracción planteados 1, 2, 3a y 3b sin fragmentación inducida a 10 kpb del ADNg (Figura 5). La evaluación se realizó tanto para las muestras tratadas a 80°C por 20 min como aquellas no tratadas (NT). Por otro lado, para la evaluación del efecto de la fragmentación inducida a 10 kpb del ADNg de Mtb en el rendimiento del secuenciamiento usando ONT se consideraron los métodos de extracción 1, 2, 3b (Figura 5).



**Figura 5.** Flujograma del análisis de la evaluación del efecto de la temperatura de inactivación (80°C por 20 min), extracción y fragmentación de ADNg de Mtb en el rendimiento del secuenciamiento ONT.

Cada temperatura de inactivación para cada método de extracción de ADNg se evaluó con 3 réplicas  
Cada condición de fragmentación para cada método de extracción de ADNg se evaluó con 6 réplicas

#### 4.13 Análisis estadístico

Los análisis estadísticos se realizaron en Rstudio versión 4.3.2 para Windows. El paquete ggplot2 de R permitió una visualización detallada de los datos. Las medidas de cantidad, calidad e integridad de las distintas extracciones de ADNg se compararon mediante la prueba no paramétrica de Kruskal-Wallis y posteriormente una prueba de Dunnett para determinar dónde radican las diferencias. Para evaluar el efecto de la temperatura de inactivación (80°C por 20 min) de Mtb y la fragmentación en el rendimiento del secuenciamiento ONT del ADNg de Mtb con

respecto al control (NT y No Fragmentado a 10 kpb) se realizó la prueba de Mann-Whitney U para determinar diferencias entre ambas condiciones. Para evaluar el efecto del método de extracción en el rendimiento del secuenciamiento ONT del ADNg de Mtb se realizó una prueba de Kruskal-Wallis y posteriormente una prueba de Dunnett.

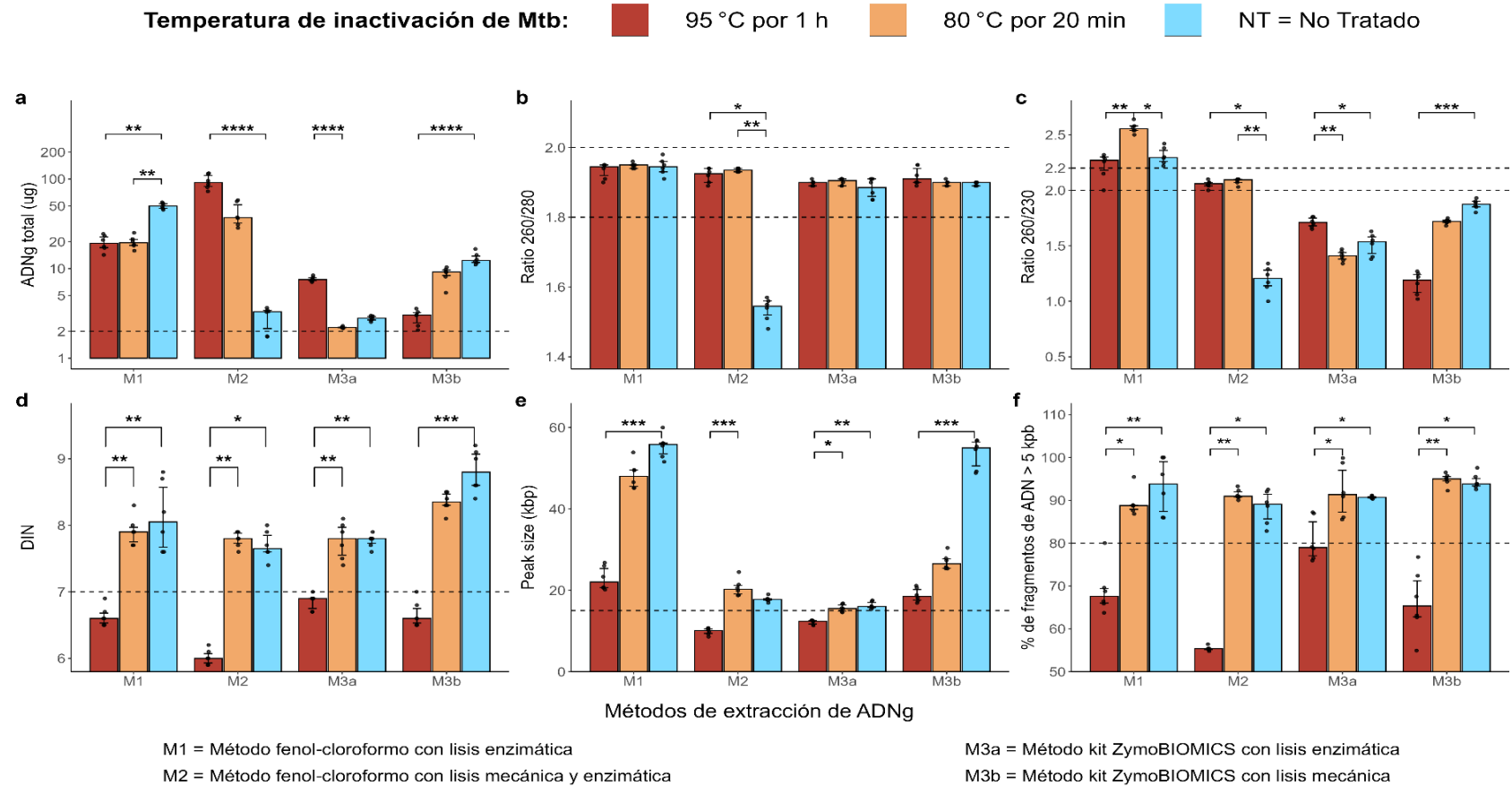
## 5 RESULTADOS

### 5.1 Efecto de la temperatura de inactivación (80 °C por 20 min y 95 °C por 1 hora) de Mtb en la cantidad, calidad e integridad del ADNg.

Los resultados del rendimiento de extracción bajo cada condición de inactivación se presentan en el Anexo 1. En la Figura 6a se observa que la temperatura de inactivación influye significativamente en la cantidad de ADNg recuperado ( $p < 0.05$ ) para todos los métodos, aunque en ningún caso se redujo por debajo del umbral requerido ( $\sim 2 \mu\text{g}$ ) para la preparación de bibliotecas por ONT (63). Para M1 y M3b, la condición NT produjo las mayores cantidades de ADNg (49.98  $\mu\text{g}$  y 12.34  $\mu\text{g}$ , mediana), con diferencias significativas frente a 95 °C/1 h y 80 °C/20 min. En contraste, en M2 y M3a la mayor recuperación se obtuvo a 95 °C/1 h (91.53  $\mu\text{g}$  y 7.55  $\mu\text{g}$ , mediana), con diferencias significativas únicamente frente a NT para M2 y frente a 80 °C/20 min para M3a.

En cuanto a la calidad (Figura 6b–6c), el ratio 260/280 se mantuvo dentro del rango óptimo (1.8–2.0) (14) en todas las condiciones, excepto en el control NT de M2 (1.55), donde se detectó contaminación proteica y diferencias significativas respecto a las condiciones de inactivación. Para el ratio 260/230, la temperatura afectó de manera significativa en todos los métodos; únicamente M2 logró valores dentro del rango recomendado (2.0–2.2) (63) tras la inactivación a 95 °C/1 h y 80 °C/20 min.

La integridad del ADNg se presentó en las Figuras 6d–6f. El DIN disminuyó de forma marcada tras la inactivación a 95 °C/1 h (6.0–6.9), ubicándose por debajo del umbral recomendado ( $DIN \geq 7$ ) (14). En contraste, 80 °C/20 min y la condición NT mantuvieron  $DIN \geq 7$ , con diferencias significativas frente a 95 °C en todos los métodos. De manera consistente y estadísticamente significativa, 95 °C/1 h redujo el Peak Size (Figura 6e) y el porcentaje de fragmentos  $\geq 5$  kpb (Figura 6f), mientras que 80 °C/20 min y NT conservaron fragmentos largos ( $>15$  kpb) y un porcentaje  $\geq 80\%$  de fragmentos  $\geq 5$  kb en todos los métodos.



**Figura 6.** Efecto de la temperatura de inactivación de Mtb en la cantidad, calidad e integridad del ADNg por cada método de extracción. Las barras indican los valores la mediana y los bigotes el rango intercuartílico. Los corchetes horizontales indican la significación estadística determinada por la prueba de Kruskal-Wallis, seguida de la prueba post-hoc de Dunn. Niveles de significancia:  $p < 0.05$  (\*),  $p < 0.01$  (\*\*),  $p < 0.001$  (\*\*\*). **a.** ADNg total (en escala log10). **b.** Relación de pureza 260/280. **c.** Relación de pureza 260/230. **d.** DIN. **e.** Peak size, parámetro que se define como el tamaño del fragmento de ADN con mayor intensidad dentro del perfil de distribución. **f.** Porcentaje de la muestra de ADN con fragmentos >5 kbp. Para a, b, c y d, las líneas discontinuas denotan los valores recomendados para la cantidad de ADN, rango de pureza y de integridad para ONT. Las pruebas se realizaron por sextuplicado.

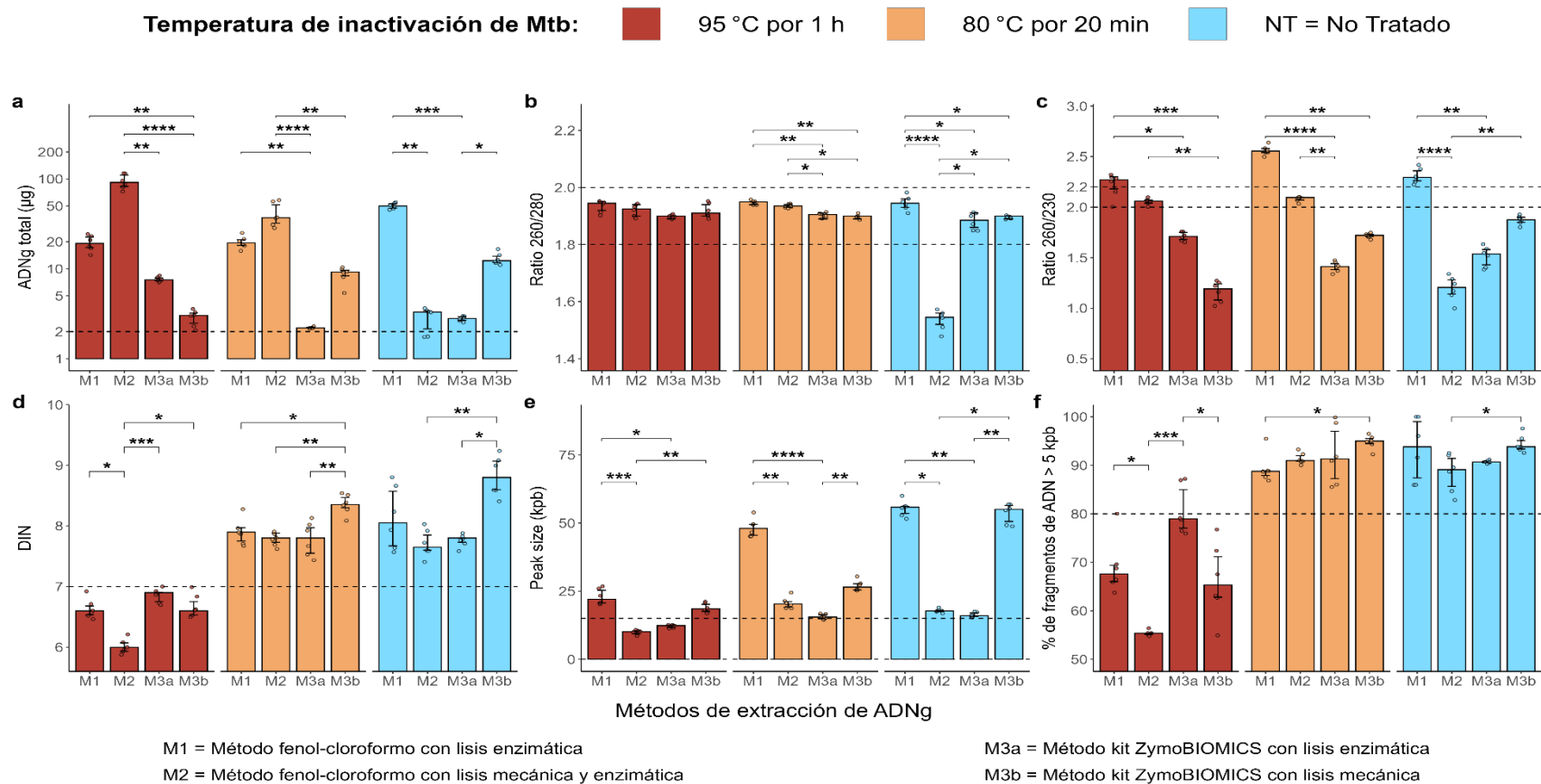
## **5.2 Efecto de cuatro métodos de extracción en la cantidad, calidad e integridad del ADNg de Mtb.**

Los resultados detallados del rendimiento de extracción bajo distintos métodos de extracción de ADNg se presentan en el Anexo 2. La Figura 7a muestra que el método de extracción influye significativamente en la cantidad de ADNg recuperado ( $p < 0.05$ ), aunque todos lograron valores suficientes para secuenciamiento ONT. El rendimiento fue dependiente de la temperatura de inactivación. En la condición NT, M1 obtuvo la mayor recuperación (49.98  $\mu\text{g}$ ), significativamente superior a los demás excepto M3b. A 80 °C/20 min y a 95 °C/1 h, M2 presentó el mayor rendimiento (36.99  $\mu\text{g}$  y 91.53  $\mu\text{g}$  respectivamente), superando significativamente a M3a y M3b, pero no frente a M1. En conjunto, M1 resultó más eficiente en muestras sin tratamiento térmico, mientras que M2 mantuvo mejor desempeño bajo inactivación.

En cuanto a la calidad del ADNg (Figuras 7b–7c), el método de extracción también tuvo un efecto significativo. Para el ratio 260/280, todos los métodos se ubicaron dentro del rango óptimo (1.8–2.0) excepto M2 en la condición NT (1.55). A 80 °C/20 min, los métodos basados en fenol-cloroformo (M1 y M2) presentaron valores superiores a los métodos basados en kit (M3a y M3b), mientras que a 95 °C/1 h las diferencias dejaron de ser significativas. Para el ratio 260/230, en las condiciones de inactivación térmica, M1 y M2 alcanzaron los valores más altos ( $\approx 2.3$ – $2.5$  para M1 y  $\approx 2.0$ – $2.2$  para M2), situándose dentro o incluso por encima

del rango recomendado para secuenciamiento ONT. En contraste, los métodos asociados al kit ZymoBIOMICS (M3a y M3b) mostraron valores inferiores bajo estas mismas condiciones. En la condición NT, solo M1 y M3b mantuvieron valores cercanos al rango óptimo, mientras que M2 y M3a presentaron una reducción más marcada.

Respecto a la integridad (Figuras 7d–7f), el método de extracción afectó significativamente tanto el DIN, Peak Size y el porcentaje de fragmentos > 5kpb. La inactivación a 95 °C/1 h redujo el DIN por debajo del umbral recomendado (umbral:  $\text{DIN} \geq 7$ ), en todos los métodos. En contraste, 80 °C/20 min y NT mantuvieron valores  $\geq 7$ , destacando M3b con los valores más altos (8.35 y 8.80). El Peak Size mostró un patrón similar: M1 generó los fragmentos más largos en todas las condiciones (22.04–55.80 kpb), con diferencias frente a M2 y M3a, pero no frente a M3b. Sólo M2 y M3a a 95 °C/1 h presentaron longitudes < 15 kpb. En cuanto al porcentaje de fragmentos > 5 kpb, bajo 95 °C/1 h, M3a conservó la mayor proporción, seguido de M1 y M3b, mientras que M2 mostró la pérdida más marcada. En 80 °C/20 min y en NT, todos los métodos mantuvieron porcentajes > 80%, con M3b tendiendo a valores ligeramente superiores. En conjunto, M3a destacó por su mayor tolerancia a 95°C/1h, mientras que M3b favoreció la preservación de fragmentos largos tanto bajo 80°C/20 min como NT.

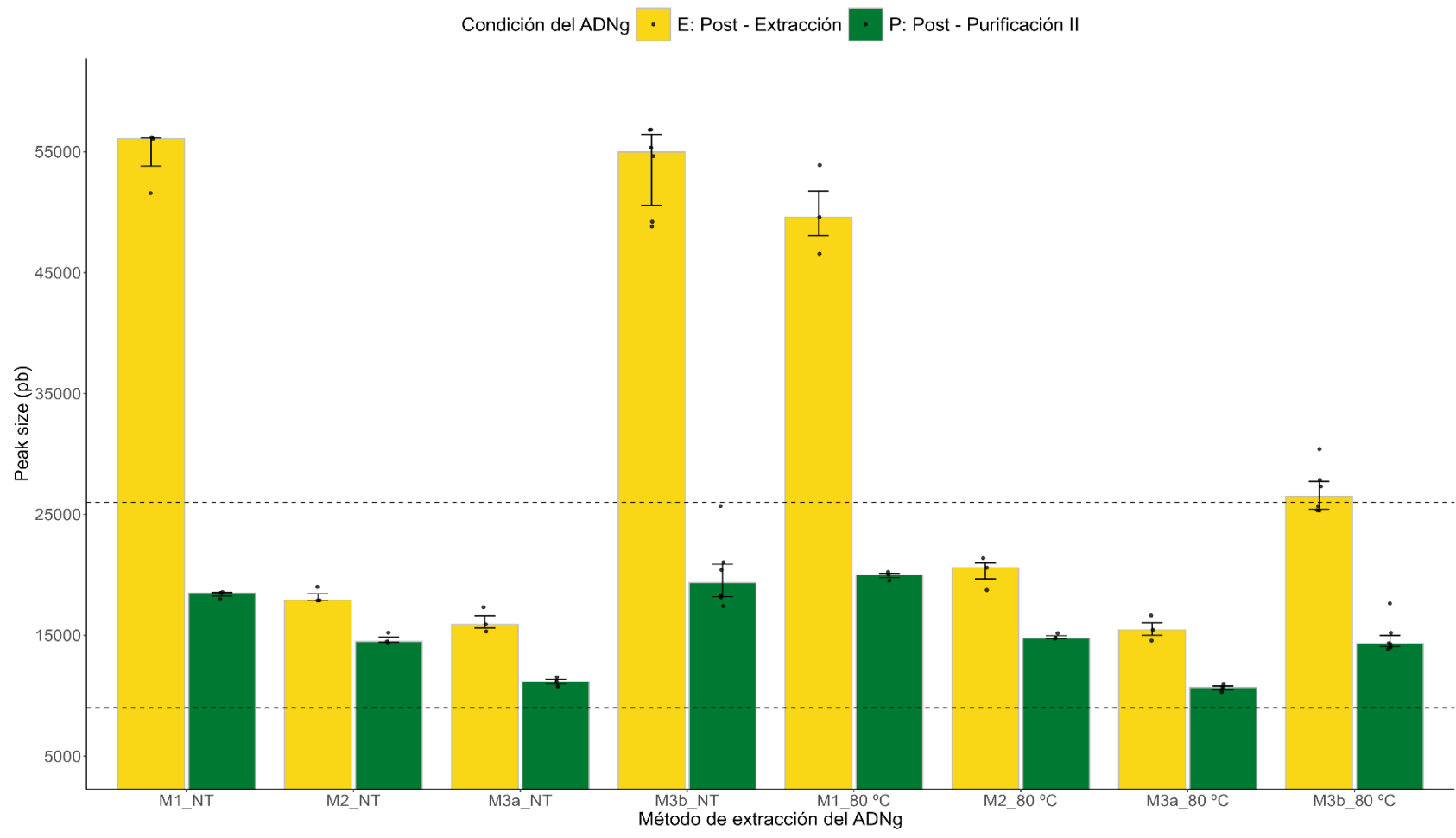


**Figura 7.** Efecto de cuatro métodos de extracción en la cantidad, calidad e integridad del ADNg por cada temperatura de inactivación de Mtb. Las barras indican los valores la mediana y los bigotes el rango intercuartílico. Los corchetes horizontales indican la significación estadística determinada por la prueba de Kruskal-Wallis, seguida de la prueba post-hoc de Dunn. Niveles de significancia:  $p < 0.05$  (\*),  $p < 0.01$  (\*\*),  $p < 0.001$  (\*\*\*) **a.** ADNg total (en escala log10). **b.** Relación de pureza 260/280. **c.** Relación de pureza 260/230. **d.** DIN. **e.** Peak size. **f.** Porcentaje de la muestra de ADN con fragmentos >5 kpb. Para a, b, c y d, las líneas discontinuas denotan los valores recomendados para la cantidad de ADN, rango de pureza y de integridad para ONT. Las pruebas se realizaron por sextuplicado.

En función a nuestros resultados previos se puede evidenciar que ningún método de extracción bajo la condición de inactivación térmica de 95 °C por 1h cumple con el requerimiento establecido de integridad para el secuenciamiento ONT: muestras con una puntuación  $DIN \geq 7$  (14,79). Para evaluar el rendimiento del secuenciamiento ONT se consideró la condición de inactivación térmica de 80 °C por 20 min y el control NT. Por otro lado, debido a que ninguna de las muestras sometidas a inactivación térmica de 80 °C por 20 min ni a la condición NT alcanzaron el valor de pureza requerido para el ratio 260/230 (2,0–2,2) (Figura 7d), se procedió a purificarlas con el kit *DNA Clean & Concentrator-5* (Zymo Research, cat. D4004) alcanzando el ratio 260/230 esperado para el secuenciamiento ONT.

### **5.3 Purificación de las muestras para el secuenciamiento ONT del ADN<sub>g</sub> de Mtb.**

Aunque la purificación con el kit DNA Clean & Concentrator-5 redujo el Peak Size de todas las muestras (Figura 8), la tendencia relativa en la condición de inactivación térmica (80 °C/20 min) y la condición NT permaneció: M1 y M3b generaron los fragmentos más largos, seguidos de M2 y M3a (Figura 8, Tabla 3).



**Figura 8.** Peak size de los fragmentos de ADN para cada método de extracción de ADNg y cada temperatura de inactivación de Mtb. Cada color indica una condición del ADNg: Amarillo (E): Post-Extracción y Verde (P): Post-Purificación. Las barras representan la mediana y los bigotes el rango intercuartílico. La línea discontinua representan el rango del tamaño de los fragmentos de ADN de acuerdo a la condición del ADNg.

#### 5.4 Preparación de librerías para el secuenciamiento ONT

En esta primera etapa dos librerías de secuenciamiento conteniendo ADN<sub>g</sub> de *Mtb* obtenido mediante los cuatro métodos de extracción bajo la inactivación térmica de 80 °C por 20 min y la condición NT se prepararon con el Native Barcoding Kit (Oxford Nanopore, SQK-NBD 114.24). La primera librería fue asignada a la celda de flujo 1 conteniendo ADN<sub>g</sub> de los métodos que usaron fenol – cloroformo (M1 y M2) y la segunda, asignada a la celda de flujo 2, conteniendo ADN<sub>g</sub> de los métodos que usaron el kit ZymoBIOMICS (M3a y M3b). Para cada método de extracción se consideró 3 réplicas tanto para la inactivación térmica de 80 °C por 20 min como para la condición NT. Por ende cada librería contenía 12 muestras.

Ambas corridas se extendieron hasta alcanzar el 5 % de poros activos lo que resultó en 40 h de corrida para ambas celdas de flujo, generando un rendimiento total muy similar entre ambas, con 4.45 Gb en la primera celda y 4.32 Gb en la segunda. Así mismo, el rendimiento normalizado por poro y hora fue muy similar entre ambas celdas de flujos con  $8.87 \times 10^{-4}$  Gb/poro·h y  $8.54 \times 10^{-4}$  Gb/poro·h (Tabla 3).

**Tabla 3.** Rendimiento del secuenciamiento ONT

Métricas del rendimiento del secuenciamiento ONT									
Celda de Flujo	Método extracción ADNg	Condición	Peak Size del ADNg (kpb) ***	# de réplicas	# de poros activos x flow cell	Rendimiento: Data total (Gb) x flow cell qScore > = 10	Tiempo de secuenciamiento (h)*	Rendimiento por poro y por hora **	Rendimiento: Data (Gb) qScore > = 10 ***
1	Método 1 (M1)	80 °C	20.00 (Q1: 19.76 , Q3: 20.12)	3	1254	4.45	40.00	0.0000887	0.30 (Q1: 0.28 - Q3: 0.33)
		NT	18.51 (Q1: 18.24 , Q3: 18,54)	3					0.15 (Q1: 0.14 - Q3: 0.25)
	Método 2 (M2)	80 °C	14.75 (Q1: 14.74 , Q3: 14.96)	3					0.24 (Q1: 0.24 - Q3: 0.26)
		NT	14.48 (Q1: 14.41 , Q3: 14.85)	3					0.33 (Q1: 0.32 - Q3: 0.45)
	Método 3a (M3a)	80 °C	10.68 (Q1: 10.49 , Q3: 10.80)	3					0.10 (Q1: 0.09 - Q3: 0.11)
		NT	11.17 (Q1: 10.96 , Q3: 11.35)	3					0.15 (Q1: 0.14 - Q3: 0.16)
Método 3b (M3b)	80 °C	14.34 (Q1: 14.30 , Q3: 14.77)	3	0.28 (Q1: 0.26 - Q3: 0.30)					
	NT	21.04 (Q1: 20.72 , Q3: 23.36)	3	0.56 (Q1: 0.43 - Q3: 0.65)					

1.\* Tiempo de secuenciamiento hasta que el 5% de poros secuenciando sea logrado

2. \*\* Calculado como el cociente entre la data total obtenida por flow cell y el número inicial de poros y el tiempo de corrida (Gb/poro/hora)

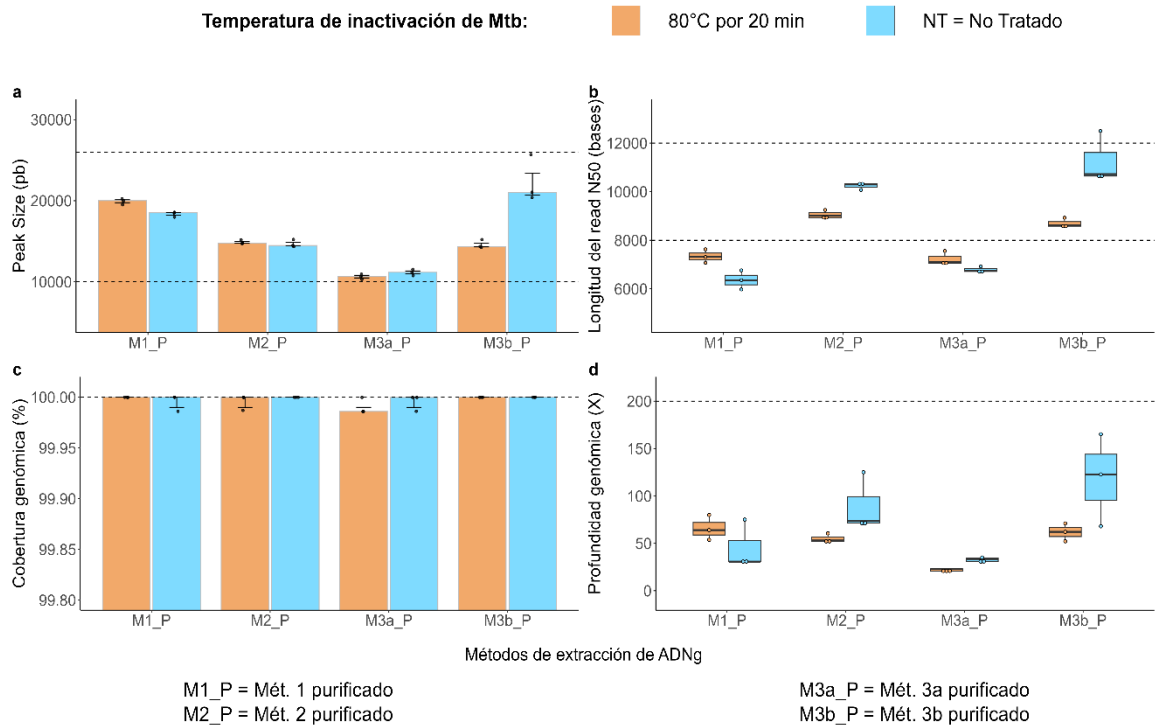
3. \*\*\* Los valores como mediana, Q1 (primer cuartil) y Q3 (tercer cuartil) se presentan con fines descriptivo.

## **5.5 Efecto de la temperatura de inactivación (80 °C por 20 min) en las métricas del secuenciamiento ONT del ADNg de Mtb.**

Para evaluar el impacto de la inactivación térmica a 80 °C por 20 min sobre el rendimiento del secuenciamiento del ADNg de Mtb con la plataforma ONT, se comparó dicha condición de inactivación frente al control NT para los cuatro métodos de extracción (M1–M3b).

La inactivación a 80 °C por 20 min no mostró un efecto estadísticamente significativo frente al control NT en el Peak Size después de la purificación ( $p > 0.05$ ), aunque se mantuvo la tendencia a una reducción marcada del tamaño de fragmentos en M3b respecto a su control NT, antes y después de la purificación (con una mediana de 21.04 kpb para el control NT y de 14.34 kpb para 80 °C por 20 min) (Figura 9a). Respecto a la longitud del read N50 (Figura 9b), tampoco se observaron diferencias significativas entre condiciones térmicas. Sin embargo, para M2 y M3b el control NT presentó valores mayores (con una mediana de 10.72 kb para M3b y 10.30 kb para M2) que la condición a 80 °C por 20 min (con una mediana de 8.60 kb para M3b y 9.02 kb para M2), mientras que M1 y M3a mostraron longitudes N50 menores a 8 kb en ambas condiciones. En términos de cobertura genómica (Figura 9c), 80 °C por 20 min no afectó significativamente los resultados, manteniéndose en ~100 % para todos los métodos, salvo una ligera reducción no significativa en M3a.

Con respecto a la profundidad genómica (Figura 9d), aunque no se alcanzaron diferencias significativas ( $p > 0.05$ ) entre ambas condiciones de inactivación y no se lograron valores de 200X para un ensamblaje completo, M2 y M3b mostraron mayor profundidad genómica en NT (con una mediana de 73.39 X y 122.73 X para M2 y M3b respectivamente), mientras que en M1 la condición térmica incrementó la profundidad genómica respecto a su control. Por otro lado, la inactivación térmica a 80 °C por 20 min no afectó de forma significativa la calidad de las lecturas de secuenciación ( $p > 0.05$ ) (Tabla 4). Los qScore se mantuvieron entre 17.40–18.20 y la precisión superó el 98 % tanto para la condición de inactivación de 80°C por 20 min como para su control NT en todos los métodos evaluados.



**Figura 9.** Efecto de la temperatura de inactivación 80°C/ 20 min de Mtb en el rendimiento del secuenciamiento ONT del ADNg de Mtb.

Las gráficas de barras muestran la mediana y el rango intercuartílico. Los gráficos de caja (boxplots) representan la distribución de los datos individuales, incluyendo la mediana, el rango intercuartílico y valores atípicos. **a.** Peak Size. **b.** Longitud del read N50 (pb). **c.** Porcentaje de cobertura. **d.** Profundidad de cobertura genómica (X). Para los paneles a y b, las líneas discontinuas representan el rango del tamaño de los fragmentos de ADN y las lecturas N50 de acuerdo a la condición de inactivación por cada método de extracción de ADNg. Para los paneles c y d, estas líneas indican el umbral objetivo de cobertura (100%) y profundidad de cobertura (200X) requerido para un rendimiento óptimo en secuenciamiento genómico.

**Tabla 4.** Efecto de la temperatura de inactivación en las métricas del secuenciamiento ONT por cada método de extracción de ADNg

Métricas de calidad del secuenciamiento ONT									
Condición	n	qScore				Precisión %			
		Mediana	Q1	Q3	p-value (Wilcoxon, BH- ajustado)	Mediana	Q1	Q3	p-value (Wilcoxon, BH- ajustado)
M1_NT	3	17.40	17.35	17.40	0.6733	98.18	98.16	98.18	0.6733
M1_80°C	3	17.40	17.40	17.40		98.18	98.18	98.18	
M2_NT	3	17.80	17.80	17.80	0.2200	98.34	98.34	98.34	0.2200
M2_80°C	3	17.60	17.60	17.65		98.26	98.26	98.28	
M3a_NT	3	17.80	17.80	17.80	1.0000	98.34	98.34	98.34	1.0000
M3a_80°C	3	17.80	17.75	17.85		98.34	98.32	98.36	
M3b_NT	3	18.20	18.15	18.20	0.2200	98.49	98.47	98.49	0.2200
M3b_80°C	3	18.00	18.00	18.05		98.42	98.42	98.44	

1. NT: No Tratado (sin inactivación térmica), 80 °C : 80 °C por 20 min

2. Los valores como mediana, Q1 (primer cuartil) y Q3 (tercer cuartil) se presentan con fines descriptivo.

3. La precisión es una función de la puntuación Phred, donde Precisión % =  $100 - 100 \cdot 10^{-(\text{Phred}/10)}$ .

4. La prueba de Wilcoxon, ajustada por el método de Benjamini-Hochbergs (BE), se aplicó por separado dentro de cada grupo de método de extracción para comparar el efecto de la inactivación térmica de Mtb.

Debido a que ninguna condición de inactivación térmica alcanzó los 200X de profundidad genómica (Figura 9d) -profundidad necesaria para un ensamblaje de alta calidad del genoma de Mtb- no se realizó los análisis del efecto de la temperatura de inactivación de Mtb en las métricas de la calidad de ensamblaje post secuenciamiento.

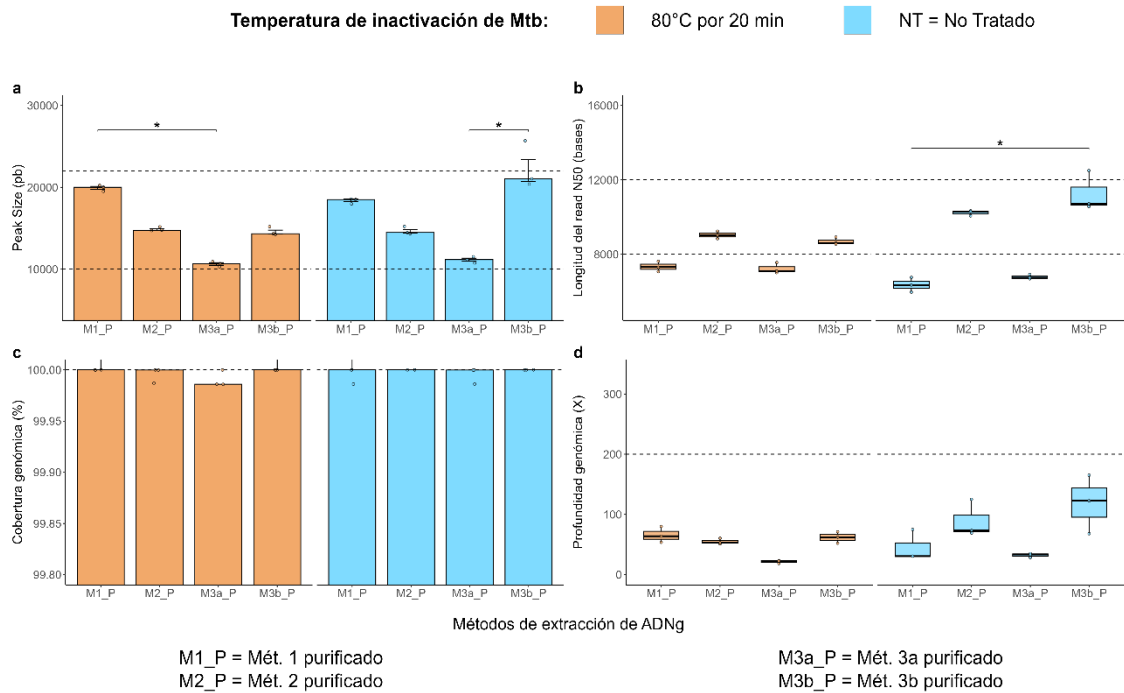
## **5.6 Efecto de cuatro métodos de extracción de ADNg de Mtb en las métricas del secuenciamiento ONT.**

El Peak Size de los fragmentos de ADNg mostró un efecto estadísticamente significativo ( $p < 0.05$ ) del método de extracción tanto para la inactivación a 80 °C/20 min como en la condición sin tratamiento NT (Figura 10a). Para la inactivación a 80 °C/20 min, M1 (20.00 kpb, mediana) produjo fragmentos más largos que M3a (10.68 kpb) con diferencias estadísticamente significativas, mientras que M2 y M3b se ubicaron en rangos intermedios. Para la condición NT, M3b generó los fragmentos más extensos (21.04 kbp) en comparación con M3a (11.17 kbp) con diferencias estadísticamente significativas, con M1 y M2 nuevamente en posiciones intermedias.

En las métricas de secuenciamiento, el método de extracción del ADNg también tuvo efectos estadísticamente significativos ( $p < 0.05$ ) en la longitud del read N50 (Figura 10b), donde M2 y M3b obtuvieron valores superiores tanto a 80 °C por 20 min como en NT. Sin embargo, solo se detectaron diferencias significativas entre M3b (10.72 kb, mediana) y M1 (6.35 kb, mediana) en la condición NT. A pesar de estas variaciones, la cobertura genómica permaneció cercana al 100 % en todos los métodos de extracción evaluados (Figura 10c), lo que confirma que los métodos de extracción de ADNg evaluados no comprometieron la representación completa del genoma.

La profundidad genómica fue más variable (Figura 10d), destacando nuevamente M3b (mediana, 122.73 X) en la condición NT, aunque sin diferencias estadísticas entre los métodos M1, M2 y M3a frente M3b. Para la condición de inactivación térmica de 80 °C por 20 min, la profundidad de lectura varía entre 22.78 X y 63.98 X según el método de extracción, con diferencias marcadas, pero no estadísticamente significativas entre los métodos M1, M2 y M3b frente al método M3a., siendo este método el que obtuvo la menor profundidad genómica. En cuanto a la calidad de lectura (Tabla 5), M3b mostró los valores más altos de qScore (18.20 - 18.00) y precisión (98.49 % - 98.42 %) en ambas condiciones térmicas, con diferencias estadísticamente significativas frente a M1 ( $p < 0.05$ ) (17.40 y 98.18%). Los métodos M2 y M3a presentaron valores intermedios.

Debido a que ningún método de extracción alcanzó los 200X de profundidad genómica (Figura 10d) -profundidad necesaria para un ensamblaje de alta calidad del genoma de Mtb- no se realizó los análisis del efecto de los métodos de extracción de ADNg en las métricas de la calidad de ensamblaje post secuenciamiento.



**Figura 10.** Efecto de cuatro métodos de extracción de ADNg de Mtb en el rendimiento del secuenciamiento ONT.

Las gráficas de barras muestran la mediana y el rango intercuartílico. Los gráficos de caja (boxplots) representan la distribución de los datos individuales, incluyendo la mediana, el rango intercuartílico y posibles valores atípicos. **a.** Peak Size. **b.** Longitud del read N50 (pb). **c.** Cobertura genómica (%). **d.** Profundidad de cobertura genómica (X). Para los paneles a y b, las líneas discontinuas representan el rango del tamaño de los fragmentos de ADN y las lecturas N50 de acuerdo a cada método de extracción de ADNg para la condición de inactivación térmica de 80°C por 20 min y el control negativo (NT). Para los paneles c y d, estas líneas indican el umbral objetivo de cobertura (100%) y profundidad de cobertura (200X) requerido para un rendimiento óptimo en secuenciamiento.

**Tabla 5.** Efecto de cuatro métodos de extracción de ADNg en las métricas del secuenciamiento ONT por temperatura de inactivación.

Métricas de calidad del secuenciamiento ONT									
Condición	n	qScore				Precisión %			
		Mediana	Q1	Q3	p-value (Kruskal-Wallis)	Mediana	Q1	Q3	p-value (Kruskal-Wallis)
M1_NT	3	17.40	17.35	17.40	0.0132	98.18	98.16	98.18	0.0132
M2_NT	3	17.80	17.80	17.80		98.34	98.34	98.34	
M3a_NT	3	17.80	17.80	17.80		98.34	98.34	98.34	
M3b_NT	3	18.2 **	18.15	18.20		98.49 **	98.47	98.49	
M1_80°C	3	17.40	17.40	17.40	0.0153	98.18	98.18	98.18	0.0153
M2_80°C	3	17.60	17.60	17.65		98.26	98.26	98.28	
M3a_80°C	3	17.80	17.75	17.85		98.34	98.32	98.36	
M3b_80°C	3	18.00 *	18.00	18.05		98.42 *	98.42	98.44	

1. NT: No Tratado (sin inactivación térmica), 80 °C : 80 °C por 20 min

2. Los valores como mediana, Q1 (primer cuartil) y Q3 (tercer cuartil) se presentan con fines descriptivo.

3. La precisión es una función de la puntuación Phred, donde Precisión % =  $100 - 100 \cdot 10^{-(\text{Phred}/10)}$ .

4. Se aplicó la prueba de Kruskal-Wallis dentro de cada grupo de inactivación térmica para evaluar diferencias globales entre los métodos de extracción

5. \* Indica diferencia estadísticamente significativa ( $p < 0.05$ , prueba de Dunn ajustada por el método de Benjamini-Hochberg) en comparación con M1 dentro de la misma condición de inactivación térmica de 80°C por 20 min.

6. \*\* Indica diferencia estadísticamente significativa ( $p < 0.01$ , prueba de Dunn ajustada por el método de Benjamini-Hochberg) en comparación con M1 dentro de la misma condición del control negativo de la inactivación térmica NT

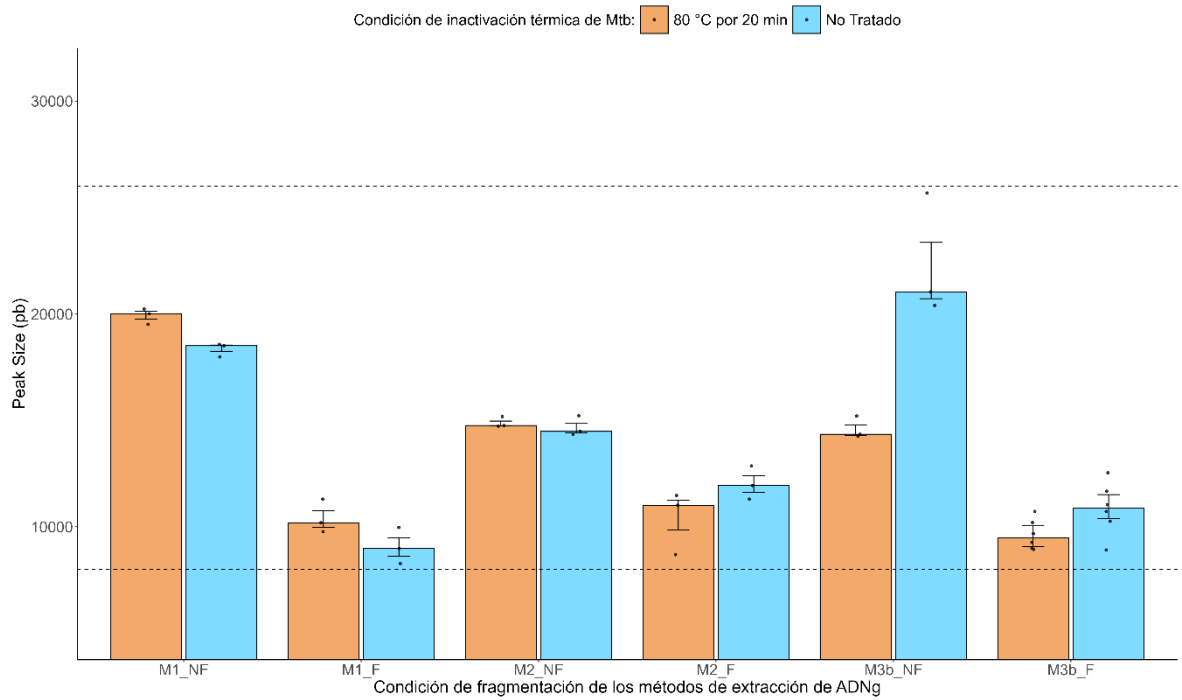
### **5.7 Peak Size de fragmentos de ADNg y rendimiento de celdas de flujo tras la fragmentación controlada 10 kpb.**

Solo se fragmentaron los métodos de extracción M1, M2 y M3b, mientras que el método M3a no fue considerado en esta etapa debido a que presentó la menor profundidad genómica en la condición no fragmentada. La Figura 11 compara el tamaño modal de los fragmentos de ADNg para cada método de extracción bajo dos condiciones (80 °C por 20 min y NT) y dos estados de fragmentación -no fragmentado (NF) y fragmentado (F) a 10 kpb. En las muestras NF, el Peak Size se ubicó entre ~14 y 21 kpb, mientras que en las muestras F a ~10 kpb se redujo este rango a ~8–12 kbp. Además, el análisis estadístico no mostró diferencias significativas entre las condiciones térmicas (80 °C/20 min vs NT) (Figura 11), por lo que estas se agruparon para centrarse en el efecto de la fragmentación inducida a ~ 10 kpb.

El rendimiento del secuenciamiento mostró que la fragmentación influyó positivamente en la generación de datos (Tabla 6). En las muestras NF, las celdas de flujo 1 y 2, ambas con un número similar de poros activos, generaron ~4.3–4.5 Gb en 40 horas, con rendimientos por poro-tiempo entre 8.5 y  $8.9 \times 10^{-5}$  Gb/poro•h. En contraste, para las muestras F se emplearon las celdas 3 y 4, que diferían notablemente en su número de poros activos (754 y 1101, respectivamente), lo que explica la mayor capacidad de producción observada en la celda 4. Esta última, asociada al método M3b, alcanzó el rendimiento más alto con 11.59 Gb en 62 h

( $17.00 \times 10^{-5}$  Gb/poro•h), mientras que la celda 3, asociada al M1 y M2 produjo 3.72 Gb en 35 h ( $14.10 \times 10^{-5}$  Gb/poro•h).

Debido a las diferencias en el número de poros y en la duración de las corridas entre las celdas de flujo (Tabla 6), se procedió a normalizar el rendimiento de secuenciamiento de cada celda tomando como referencia la celda 4 —la de mayor desempeño— con el fin de homogenizar las condiciones y minimizar sesgos atribuibles a la capacidad física de cada celda. Tras este ajuste, los rendimientos proyectados fueron de 3.91, 3.76, 9.62 y 11.59 Gb para las celdas 1, 2, 3 y 4, respectivamente (Tabla 6), observándose que, aunque la celda 4 mantuvo el mayor rendimiento, la diferencia relativa respecto a la celda 3 se redujo, indicando que parte de la variabilidad inicial se debía a diferencias instrumentales.



**Figura 11.** Peak size del ADNg para cada condición de fragmentación en tres métodos de extracción de ADNg de Mtb.

Cada color indica la condición de temperatura de inactivación. Las barras representan la mediana y los bigotes el rango intercuartílico. La línea discontinua representan el rango del tamaño de los fragmentos de ADNg

**Tabla 6.** Rendimiento del secuenciamiento ONT para las muestras No Fragmentadas vs Frgamentadas a 10 kpb

Métricas del rendimiento del secuenciamiento ONT								
Celda de flujo	Método extracción ADNg	Condición	# de réplicas	Número de poros activos de la flow cell	Rendimiento: Data total (Gb) x flow cell qScore > = 10	Tiempo de secuenciamiento (h)*	Rendimiento por poro y por hora **	Rendimiento: Data total (Gb) x flow cell qScore > = 10 proyectada ***
<b>1</b>	Método 1	No	6	1254	4.45	40.00	0.000089	3.91 <sup>a</sup>
	Método 2	Fragmentado a 10 kpb	6					
<b>2</b>	Método 3a	No	6	1264	4.32	40.00	0.000085	3.76 <sup>a</sup>
	Método 3b	Fragmentado a 10 kpb	6					
<b>3</b>	Método 1	Fragmentado a 10 kpb	6	754	3.72	35.00	0.000141	9.62 <sup>b</sup>
	Método 2		6					
<b>4</b>	Método 3b	Fragmentado a 10 kpb	12	1101	11.59	62.00	0.000170	11.59 <sup>b</sup>

1.\* Tiempo de secuenciamiento hasta que el 5% de poros secuenciando sea logrado

2. \*\* Calculado como el cociente entre la data total obtenida por flow cell y el número inicial de poros activos y el tiempo de corrida (Gb/poro/hora)

3. \*\*\* Data proyectada en función al rendimiento por poro y por hora acorde a los 1101 poros disponibles y el tiempo de

a: para la condición no fragmentado a 10 kpb con 40 horas de corrida de secuenciamiento

b: para la condición fragmentado a 10 kpb con 62 horas de corrida de secuenciamiento

Los parámetros de rendimiento —Data (Gb) con qScore  $\geq 10$ , longitud de read N50, cobertura genómica (%), profundidad (X), qScore y precisión— se presentan en las Tablas 7 y 8. Tras la normalización, el método M3b mantuvo la mayor producción de datos, con 0.98 Gb en la condición fragmentada (F a ~10 kpb) y 0.32 Gb en la condición no fragmentada (NF). Bajo el mismo criterio, M2 alcanzó valores estimados de 0.98 Gb (F) y 0.26 Gb (NF), mientras que M1 registró 0.61 Gb (F) y 0.24 Gb (NF). Este ajuste permitió comparar los métodos bajo condiciones equivalentes de secuenciamiento, separando el efecto real del método y de la fragmentación del efecto asociado a diferencias entre celdas.

Dado que la profundidad genómica es directamente proporcional al volumen de datos generado, los valores normalizados de la data proyectada impactaron de forma directa en las profundidades genómicas proyectadas para cada método y condición. En la Tabla 7 se presentan conjuntamente la data proyectada y la profundidad genómica tanto obtenida como proyectada, lo que permite una comparación equitativa del desempeño de cada método.

**Tabla 7.** Efecto de la fragmentación de ADNg a 10 kb en las métricas del secuenciamiento ONT.

Métricas del rendimiento del secuenciamiento ONT								
Método extracción ADNg	Condición	# de réplicas	Data (Gb) qScore > = 10	Data (Gb) proyectada * qScore > = 10	Cobertura genómica (%)	Longitud del read N50 (bases)	Profundidad genómica (X)	Profundidad genómica (X) proyectada **
Método 1	Fragmentado a 10 kpb	6	0.23 (Q1: 0.19 - Q3: 0.28)	0.61 b (Q1: 0.49 - Q3: 0.73)	99.99	5398 Q1: 4996 Q3: 5799	50.98 Q1: 41.17 Q3: 61.64	131.83 b Q1: 106.47 Q3: 159.40
	No Fragmentado a 10 kpb	6	0.28 (Q1: 0.17 - Q3: 0.34)	0.24 a (Q1: 0.15 - Q3: 0.29)	100	6914 Q1: 6456 Q3: 7253	58.84 Q1: 36.48 Q3: 72.23	51.70 a Q1: 32.05 Q3: 63.46
Método 2	Fragmentado a 10 kpb	6	0.38 (Q1: 0.31 - Q3: 0.46)	0.98 b (Q1: 0.80 - Q3: 1.20)	100	8045 Q1: 7850 Q3: 8137	84.56 Q1: 68.89 Q3: 103.36	218.68 b Q1: 178.16 Q3: 267.30
	No Fragmentado a 10 kpb	6	0.29 (Q1: 0.25 - Q3: 0.33)	0.26 a (Q1: 0.22 - Q3: 0.29)	100	9656 Q1: 9075 Q3: 10239	64.88 Q1: 54.89 Q3: 72.33	57.01 a Q1: 48.23 Q3: 63.56
Método 3b	Fragmentado a 10 kpb	12	0.98 (Q1: 0.87 - Q3: 1.07)	0.98 b (Q1: 0.87 - Q3: 1.07)	100	7331 Q1: 6531 Q3: 7984	214.70 Q1: 191.38 Q3: 237.28	214.70 b Q1: 191.38 Q3: 237.28
	No Fragmentado a 10 kpb	6	0.32 (Q1: 0.29 - Q3: 0.50)	0.28 a (Q1: 0.25 - Q3: 0.43)	100	9745 Q1: 8685 Q3: 10681	69.49 Q1: 63.58 Q3: 109.80	60.48 a Q1: 55.34 Q3: 95.56

1. Los valores son representados por la mediana, Q1 (primer cuartil) y Q3 (tercer cuartil) y se presentan con fines descriptivo.
2. \* La data (Gb) por condición fue proyectada para la data (Gb) por flow cell proyectada para una flow cell con 1101 poros y el tiempo de a: para la condición no fragmentado a 10 kpb con 40 horas de corrida de secuenciamiento  
b: para la condición fragmentado a 10 kpb con 62 horas de corrida de secuenciament
3. \*\* La profundidad genómica (X) por condición fue proyectada de acuerdo al rendimiento proyectado

## **5.8 Evaluación del efecto de la fragmentación a 10 kpb del ADNg de Mtb en las métricas del secuenciamiento ONT.**

La Figura 12a muestra que la fragmentación a ~10 kbp redujo de forma significativa el Peak Size en todos los métodos de extracción ( $p < 0.05$ ), pasando de medianas entre 14.34–21.04 kbp en la condición NF a valores entre 8–12 kbp en la condición F a 10 kpb. Para evaluar el efecto de esta fragmentación en las métricas de secuenciamiento, las métricas como la longitud del read N50 y la cobertura genómica (%) se calcularon a partir de los datos de cada corrida, excepto la profundidad genómica, que se recalculó tras normalizar el rendimiento en Gb debido a su dependencia directa del volumen de datos.

En cuanto a la longitud del read N50 (Figura 12b), la fragmentación a 10 kpb también produjo una reducción significativa en todos los métodos ( $p < 0.05$ ). En M1, M2 y M3b las medianas disminuyeron de 6.91, 9.66 y 9.75 kb en la condición NF a 5.40, 8.05 y 7.33 kb en la condición F, respectivamente. Sin embargo, la cobertura genómica se mantuvo cercana al 100 % en todos los métodos de extracción de ADNg y condiciones de fragmentación (Figura 12c).

La profundidad genómica proyectada (Figura 12d), obtenida tras normalizar el rendimiento al de la celda de flujo que generó la mayor cantidad de datos, mostró incrementos significativos en la condición F ( $p < 0.05$ ), alcanzando medianas de 131.83, 218.68 y 214.70X para M1, M2 y M3b, en comparación con 51.70, 57.01

y 60.48X en NF. Solo M2 y M3b alcanzaron profundidades  $\geq 200X$ , umbral recomendado para ensamblajes completos del genoma con lecturas largas.

Es importante señalar que tanto el qScore como la precisión derivan del proceso de basecalling y están influenciados por la calidad de la señal eléctrica y por los parámetros del modelo de decodificación, por lo que no pueden normalizarse sólo en función del rendimiento de la celda de flujo. A diferencia de la profundidad genómica —que sí puede recalcularse a partir del volumen total de datos generados—, estas métricas reflejan propiedades intrínsecas del proceso de lectura y del algoritmo de interpretación.

Por lo previamente señalado, los valores de qScore y precisión se analizaron directamente sobre los datos brutos de cada corrida. En el caso del método M3b, la comparación entre la condición fragmentada y no fragmentada fue posible, observándose que la condición F a 10 kpb redujo tanto el qScore como la precisión (16.70 y 97.86) respecto a la condición NF (18.10 y 98.45). Para M1 y M2, en cambio, debido a las diferencias marcadas en las celdas de flujo empleadas, los valores se reportaron únicamente de forma descriptiva, con medianas que variaron entre 17.40 y 18.35 en qScore y entre 98.18 y 98.53 en precisión (Tabla 8).



**Tabla 8.** Efecto de la fragmentación de ADNg a 10 kpb en las métricas de precisión del secuenciamiento ONT por cada método de extracción.

Métricas de calidad del secuenciamiento ONT									
Condición	n	qScore				Precisión %			
		Mediana	Q1	Q3	p-value (Wilcoxon, BH-adjusted)	Mediana	Q1	Q3	p-value (Wilcoxon, BH-adjusted)
Método 1 Fragmentado a 10 kpb	6	17.95	17.90	18.00	0.00413	98.40	98.38	98.42	0.00413
Método 1 No Fragmentado a 10 kpb	6	17.40	17.40	17.40		98.18	98.18	98.18	
Método 2 Fragmentado a 10 kpb	6	18.35	18.30	18.40	0.00413	98.53	98.52	98.55	0.00413
Método 2 No Fragmentado a 10 kpb	6	17.75	17.62	17.80		98.32	98.27	98.34	
Método 3b Fragmentado a 10 kpb	12	16.70	16.70	16.70	0.00146	97.86	97.86	97.86	0.00146
Método 3b No Fragmentado a 10 kpb	6	18.10	18.02	18.17		98.45	98.43	98.48	

1. Los valores como mediana, Q1 (primer cuartil) y Q3 (tercer cuartil) se presentan con fines descriptivo.
2. La precisión es una función de la puntuación Phred, donde Precisión % =  $100 - 100 \cdot 10^{-(\text{Phred}/10)}$ .
3. La prueba de Wilcoxon, ajustada por el método de Benjamini-Hochbergs (BE), se aplicó por separado dentro de cada grupo de método de extracción para comparar el efecto de la condición de fragmentación del ADN.

## **5.9 Evaluación del efecto de la fragmentación a 10 kpb del ADNg de Mtb en las métricas de la calidad del ensamblaje post secuenciamiento ONT con read dúplex**

Para el análisis de la calidad de ensamblaje se consideraron exclusivamente las lecturas dúplex, dado que ofrecen mayor exactitud en la inferencia de bases. La proporción de lecturas dúplex fue similar entre métodos, representando entre el 3.9 % y 5.7 % del total de lecturas (Tabla 9). Sin embargo, su contribución en la calidad fue determinante. Mientras las lecturas *simplex* presentaron medianas de qScore en el rango de 16.70–18.35 (precisión de 97.86–98.54), las lecturas dúplex alcanzaron valores significativamente más altos, con medianas entre 26.20 y 28.10 (precisión de 99.76–99.85) en los tres métodos evaluados (Tabla 9), lo que se traduce en una mejora sustancial en la precisión base a base y, por tanto, en la confiabilidad del ensamblaje resultante.

Las métricas de ensamblaje se evaluaron directamente sobre las lecturas dúplex de cada corrida, sin aplicar normalización adicional. Bajo este criterio, la comparación entre las condiciones F y NF solo fue factible para el método M3b, ya que en ambos casos se emplearon celdas de flujo con capacidades comparables. En los métodos M1 y M2, las métricas de ensamblaje no pudieron compararse porque se obtuvieron a partir de celdas de flujo con capacidades distintas (celda 1 vs. celda 3), lo que impidió una comparación directa entre condiciones. No obstante, las métricas se

presentan en la Figura 13 con fines descriptivos de las métricas obtenidas bajo estas condiciones.

La Figura 13a evidencia que la fragmentación a ~10 kbp incrementó significativamente ( $p < 0.05$ ) la fracción del genoma cubierto en M3b, alcanzando una cobertura del 99.83 %. En el caso de M2, si bien se observó una cobertura de 95.67 % con la data obtenida, dicho valor debe interpretarse con cautela, pues no refleja una comparación equivalente entre condiciones debido a las limitaciones mencionadas.

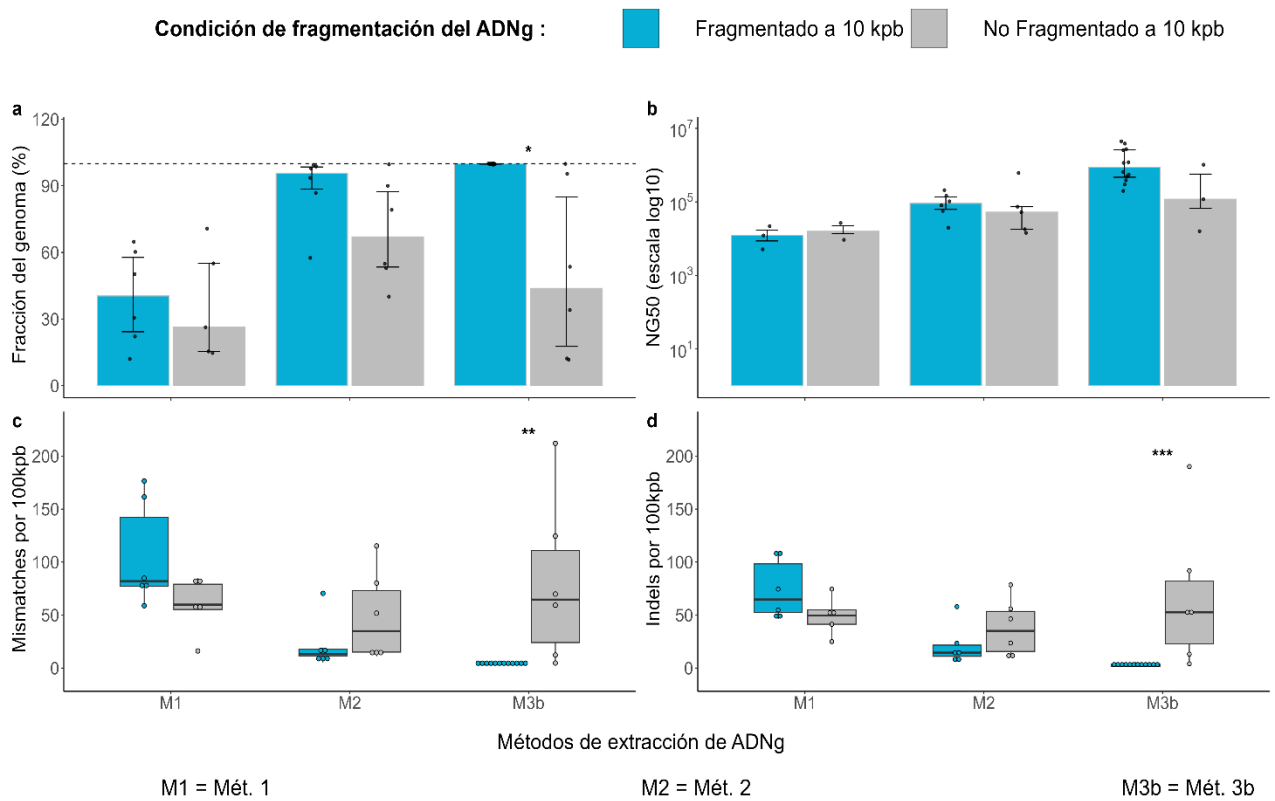
Asimismo, la fragmentación no produjo diferencias estadísticamente significativas en el NG50 cuando se comparó con la condición NF del método M3b (Figura 13b) ( $p > 0.05$ ). No obstante, M3b en condición F a 10 kpb presentó la mayor mediana de NG50 0.89 Mb (Q1: 0.47 Mb - Q3: 2.59 Mb) con valores extremos de 3.86 Mb y 4.41 Mb (Figura 13b, Anexo3), indicando una mayor continuidad del ensamblaje. Finalmente, la fragmentación redujo de manera marcada los errores de ensamblaje (mismatches e indels) en M2 y M3b, con una reducción estadísticamente significativa únicamente en M3b ( $p < 0.05$ ) (Figuras 13c y 13d, Anexo 4 y 5).

**Tabla 9.** Métricas de la data dúplex del secuenciamiento ONT

Métricas de calidad del secuenciamiento ONT									
Método extracción ADNg	Condición	# de réplicas	Data (Gb) qScore > = 10	Data (Gb) proyectada * qScore > = 10	# reads	qScore	# reads duplex	qScore - duplex	reads duplex (%)
Método 1	Fragmentado	6	0.23 (Q1: 0.19 - Q3: 0.28)	0.61 (Q1: 0.49 - Q3: 0.73)	55782 Q1: 50223.00 Q3: 72459.0	17.95 Q1: 17.90 Q3: 18.00	2630.5 Q1: 2215.50 Q3: 3353.00	27.45 Q1: 27.30 Q3: 27.60	4.54 Q1: 4.50 Q3: 4.73
	No Fragmentado	6	0.28 (Q1: 0.17 - Q3: 0.34)	0.24 (Q1: 0.15 - Q3: 0.29)	59775.5 Q1: 40777.50 Q3: 81128.5	17.40 Q1: 17.40 Q3: 17.40	2346.5 Q1: 1411.25 Q3: 3410.75	27.00 Q1: 26.90 Q3: 27.17	3.91 Q1: 3.31 Q3: 4.04
Método 2	Fragmentado	6	0.38 (Q1: 0.31 - Q3: 0.46)	0.98 (Q1: 0.80 - Q3: 1.20)	67465 Q1: 54296.75 Q3: 78415.5	18.35 Q1: 18.30 Q3: 18.40	3898.0 Q1: 3203.50 Q3: 4341.25	27.95 Q1: 27.90 Q3: 28.00	5.74 Q1: 5.57 Q3: 5.88
	No Fragmentado	6	0.29 (Q1: 0.25 - Q3: 0.33)	0.26 (Q1: 0.22 - Q3: 0.29)	47158.5 Q1: 44083.75 Q3: 50057.0	17.75 Q1: 17.62 Q3: 17.80	2360.5 Q1: 2199.00 Q3: 2532.50	27.55 Q1: 27.50 Q3: 27.60	5.19 Q1: 5.02 Q3: 5.32
Método 3b	Fragmentado	12	0.98 (Q1: 0.87 - Q3: 1.07)	0.98 (Q1: 0.87 - Q3: 1.07)	174664 Q1: 154354.50 Q3: 191953.5	16.70 Q1: 16.70 Q3: 16.70	8611.0 Q1: 8183.00 Q3: 9872.50	26.20 Q1: 26.20 Q3: 26.30	5.13 Q1: 4.97 Q3: 5.30
	No Fragmentado	6	0.32 (Q1: 0.29 - Q3: 0.50)	0.28 (Q1: 0.25 - Q3: 0.43)	46171.5 Q1: 35767.00 Q3: 60699.5	18.10 Q1: 18.02 Q3: 18.17	1881.0 Q1: 1243.25 Q3: 2889.25	28.10 Q1: 27.95 Q3: 28.10	4.05 Q1: 3.46 Q3: 4.73

1. Los valores son representados por la mediana, Q1 (primer cuartil) y Q3 (tercer cuartil)

2. \* La data (Gb) fue proyectada al rendimiento de la celda de flujo con 1101 poros y 62 h de corrida

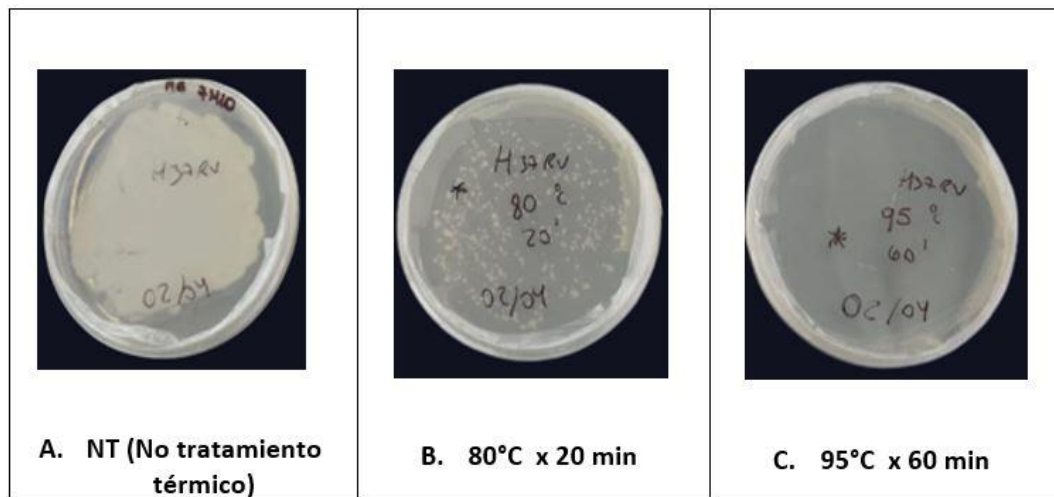


**Figura 13.** Efecto de la fragmentación a 10 kpb del ADNg de Mtb sobre las métricas de la calidad del ensamblaje post secuenciamiento ONT con read dúplex.

Las barras representan la mediana y el rango intercuartílico; los diagramas de caja (boxplots) muestran la distribución de los datos individuales, incluyendo la mediana, el rango intercuartílico y valores atípicos. **a.** Fracción del genoma (%) cubierta por los contigs ensamblados; la línea discontinua indica el umbral ideal (100%) para cobertura completa del genoma. **b.** NG50 (escala logarítmica base 10), como indicador de la continuidad del ensamblaje. **c.** Número de mismatches por 100 kbp. **d.** Número de indels por 100 kbp.

### 5.10 Confirmación de la inhibición del crecimiento de Mtb ante dos condiciones de inactivación térmica: 80 °C/20 min y 95 °C/1 hora

La confirmación de la inactivación térmica de Mtb nos indicó que tras 90 días de incubación, los controles NT mostraron una alta confluencia de crecimiento; las muestras tratadas a 80°C/ 20 min mostraron un crecimiento considerablemente reducido; y las muestras tratadas a 95°C/1 hora no mostraron signos de crecimiento (Figura 14)



**Figura 14.** Evaluación del crecimiento de Mtb inactivadas a 80°C/20 min y 95°C/1h tras 90 días de incubación a 37°C

## 6 DISCUSIÓN

Este estudio se desarrolló en tres etapas. En la primera, se evaluaron dos condiciones de inactivación térmica de Mtb - 80 °C/20 min (90–93) y 95 °C/1 h (17,63) -, además del control sin tratamiento (NT), en combinación con cuatro métodos de extracción de ADNg, con el fin de determinar su impacto sobre el rendimiento, la pureza y la integridad del ADNg, y seleccionar las condiciones más adecuadas para la preparación de bibliotecas y el secuenciamiento mediante ONT. La temperatura de inactivación modificó el rendimiento de forma dependiente del método de extracción: incrementó en M2 y M3a y redujo en M1 y M3b (Figuras 6a y 7a). Aunque todas las combinaciones superaron los 2 µg, M2 requiere desnaturalización térmica para una lisis eficiente y un mejor rendimiento, ya que temperaturas altas favorece a la disrupción de la envoltura (94,95). M3b fue particularmente sensible a 95 °C/1 h, probablemente por la generación de fragmentos cortos que favorecieron la pérdida de fragmentos largos de ADN durante la purificación en columnas.

En cuanto a la pureza, el ratio 260/280 se mantuvo en el rango óptimo (~1.8–2.0) para todos los métodos y condiciones de inactivación (Figuras 6b y 7b), excepto para M2 en NT (1.55), lo que indica que este método requiere la desnaturalización térmica para eliminar proteínas de forma eficiente. A 95 °C/1 h, todas las variantes convergieron al rango óptimo, lo que sugiere que esta temperatura proporciona una desnaturalización suficiente para homogeneizar la purificación independientemente del método. Este comportamiento concuerda con la naturaleza altamente lipídica y

poco permeable de la envoltura de Mtb, que exige pasos de lisis y desnaturalización más enérgicos para remover macromoléculas intracelulares (96–98). En conjunto, M2 requiere el calentamiento para lograr la pureza 260/280 adecuada a diferencia de los demás métodos.

Respecto a contaminantes orgánicos, el ratio 260/230 mostró mayor variabilidad. Solo M2 bajo 95 °C/1 h y 80 °C/20 min alcanzó el rango óptimo (2.0–2.2). Este resultado coincide con los valores reportados en un estudio previo que usa un método similar al M2 para la extracción de ADN micobacteriano (63). En M1, los valores permanecieron > 2.2, coherentes con la separación orgánica y los múltiples lavados etanólicos del protocolo, y aunque exceden el rango típico para ONT, siguen siendo considerados aceptables según criterios de pureza reportados en 2019 por New England Biology (2.3-2.4) (99). En contraste, los métodos basados en columnas de sílice (M3a/M3b) tendieron a valores < 2.0, probablemente por arrastre residual de guanidina o etanol (99), lo que podría mitigarse mediante lavados adicionales con etanol al 70 % y tiempos de secado más prolongados.

La temperatura de inactivación y el método de extracción afectaron significativamente la integridad del ADNg (Figuras 6d-f y 7d-f). La condición de 95 °C/1h redujo de forma marcada el Peak Size (< 15 kbp), el porcentaje de fragmentos > 5 kpb (< 80%) y no alcanzó un DIN  $\geq$  7, valor mínimo recomendado para secuenciación de lecturas largas (14), por lo que no resultó adecuada para secuenciamiento de lecturas largas. En contraste, 80 °C/ 20 min y la condición NT preservaron fragmentos más largos e íntegros. Esto coincide con lo mencionado en previos trabajos en donde se indica que temperaturas altas de manera prolongada

como 95°C/1h pueden degradar el ADNg (17,94,100). Sin embargo, no se conocía el impacto a nivel cuantitativo que podría tener esta variable en la integridad del ADNg, siendo este el primer trabajo que cuantifica este efecto.

El método M3b presentó los valores de DIN más altos, con diferencias estadísticamente significativas frente a M1, M2 y M3a tanto en 80 °C/20 min como en NT. Este resultado fue inesperado, pues a pesar de incluir una lisis mecánica prolongada combinada con lisis química (101), M3b preservó mejor la integridad del ADNg que los demás métodos, los cuales no emplean una disrupción mecánica sostenida. Si bien M1 alcanzó el Peak Size más elevado, M3b proporcionó un mejor equilibrio entre longitud e integridad. En contraste, M3a y M2 produjeron fragmentos más cortos, posiblemente debido, en el caso de M3a, a la saturación de la membrana de sílica que limita la recuperación de ADN de alto peso molecular (102), y en el caso de M2, al cizallamiento mecánico generado durante la etapa de FastPrep previa a la lisis enzimática.

En síntesis, la inactivación a 95 °C durante 1h afectó negativamente la integridad del ADNg en todos los métodos, por lo que no se considera técnicamente adecuada para preservar la integridad del ADNg para secuenciación de lecturas largas. En contraste, el método M2 bajo 80 °C durante 20 min fue el único método que alcanzó cantidad, pureza e integridad suficientes para el secuenciamiento ONT sin requerir pasos de purificación adicional. No obstante, con el fin de estandarizar el procedimiento y permitir comparaciones válidas entre métodos y condiciones, se aplicó una purificación adicional a todas las muestras bajo 80 °C/20 min y NT.

Aunque este paso redujo ligeramente el Peak Size, la tendencia de las diferencias en el tamaño de los fragmentos entre métodos y condiciones térmicas se mantuvo tras la purificación.

En la segunda etapa, dado que el fenol puede interferir en la preparación de librerías y en el rendimiento del secuenciamiento ONT (tolerancia ~1%) (18), las muestras se agruparon para el secuenciamiento según el método de extracción: M1 y M2 (protocolos con fenol) en una misma celda de flujo, mientras que M3a y M3b (purificación en columnas de sílica) se agruparon en otra, con tres réplicas por condición térmica evaluada. Esta organización permitió evaluar de manera comparativa el efecto de la inactivación térmica y el método de extracción sobre el rendimiento del secuenciamiento. Ambas celdas de flujo alcanzaron rendimientos similares tras 40 horas de corrida (4.45 Gb y 4.32 Gb), lo que sugiere que los niveles residuales de fenol se mantuvieron dentro de los límites tolerables para una preparación de librerías y un secuenciamiento eficientes en ONT.

Al evaluar el efecto de la inactivación térmica sobre las métricas de secuenciamiento, se observó que bajo la condición NT, M2 y M3b tendieron a presentar mayores valores de read N50 y profundidad genómica en comparación con 80 °C/20 min, aunque sin diferencias significativas, probablemente debido al número limitado de réplicas. No obstante, la cobertura genómica y el qScore se mantuvieron estables en ambas condiciones, lo que indica que 80 °C/20 min no compromete la representación del genoma ni la calidad de lectura. Estas variaciones podrían estar asociadas a la interacción entre el tratamiento térmico y los protocolos

de extracción, afectando en diferente grado la preservación de fragmentos largos y la eficiencia en la preparación de librerías (103).

El análisis comparativo entre métodos evidenció diferencias significativas en la longitud del read N50 y en el qScore (Figura 10, Tabla 5), con M3b mostrando el mejor desempeño, seguido de M2, particularmente bajo NT. En contraste, M1 presentó el rendimiento más bajo, posiblemente debido a la coprecipitación de compuestos de la envoltura micobacteriana asociada al uso de CTAB (bromuro de cetiltrimetilamonio) - comúnmente empleado en la extracción de ADN genómico de *Mtb* (103,104) - lo que puede interferir con la preparación de librerías y afectando la eficiencia del secuenciamiento, en concordancia con lo reportado por Elton et al. (2025) (103). M3a también mostró desempeño reducido, probablemente por la sustitución de lisis mecánica por una lisis enzimática insuficiente para disrumpir eficazmente la envoltura celular de *Mtb*, favoreciendo la retención de contaminantes que afectan la eficacia del secuenciamiento (105). A pesar de estas diferencias, la cobertura genómica permaneció en 100 % en todos los métodos, lo que sugiere que la representación del genoma no se vio comprometida.

En conjunto, los resultados indican que, bajo la condición NT, los métodos M2 y M3b son los más adecuados para preparar ADN<sub>g</sub> para secuenciación ONT, ya que favorecen la recuperación de fragmentos largos y reducen la presencia de inhibidores que afectan la preparación de librerías. Ambos métodos mostraron un mejor desempeño global, no solo en cantidad, pureza e integridad del ADN<sub>g</sub> — considerando la purificación adicional— sino también en el rendimiento del

secuenciamiento, reflejado en un mayor número de bases obtenidas, read N50 más largos, mayor profundidad genómica y qScores superiores. No obstante, ningún método bajo ninguna condición térmica alcanzó profundidades cercanas a 200X, valor recomendado como umbral óptimo para ensamblajes de alta resolución con lecturas largas (72).

En la tercera etapa del estudio se evaluó el efecto de la fragmentación controlada del ADNg a ~10 kpb sobre el rendimiento del secuenciamiento ONT. Dado que las celdas de flujo presentaban variaciones en el número de poros y en la duración de las corridas, los resultados se normalizaron según el número de poros y el tiempo de secuenciamiento, tomando como referencia la celda con mayor rendimiento. Esta normalización permitió eliminar sesgos instrumentales y realizar comparaciones equitativas entre métodos y condiciones. Bajo estas condiciones equivalentes, para el método M1, M2 y M3b el efecto significativo de la fragmentación: incrementó la profundidad genómica a costa de una reducción en la longitud del read N50, mientras que la cobertura genómica se mantuvo en 100 %.

En este contexto, entre los métodos evaluados M3b fragmentado a ~10 kpb aumentó la profundidad genómica a 214X –valor considerado óptimo para los ensamblajes de alta resolución usando lecturas largas (72), mientras que disminuyó el read N50 y el qScore. Un comportamiento similar se observó en M2, con profundidades proyectadas de 218X y reducciones moderadas en el N50. No obstante, estas disminuciones se mantuvieron dentro de rangos compatibles con ensamblajes de alta resolución, y se compensan con la elevada profundidad obtenida. Esto se

demuestra en estudios previos donde una profundidad de 238X permitió un ensamblaje completo de Mtb aún con un N50 reducido (5.07 kb) y una celda de flujo con menor precisión de base (R.9.4) reportado en el 2018 por Bainomugisa et al. (12). Bajo este escenario, la fragmentación a ~ 10 kpb posiciona a M3b (experimentalmente) y a M2 (según datos normalizados) como métodos particularmente adecuados para análisis genómicos de alta resolución.

En términos de ensamblaje, el uso de lecturas dúplex en M3b incrementó la precisión de basecalling (26.20; 99.76 %) y permitió recuperar el 99.83 % del genoma, asegurando una reconstrucción casi completa. Aunque en M2, pese a no contar con la data proyectada (0.98 Gb, 218 X) para calcular métricas de ensamblaje con lecturas dúplex, logró recuperar el 95.67 % del genoma con la data disponible (0.38 Gb, 84.56X), lo que sugiere que bajo condiciones instrumentales óptimas podría alcanzar un desempeño comparable al de M3b. Además, M3b obtuvo el mayor NG50 (0.897 Mb) - en concordancia con lo reportado en el 2023 para protocolos basados en purificación por columna por Di Marco et al. (106) - y las menores tasas de mismatches e indels, incluso inferiores a los reportados para secuenciación Illumina en el 2021 por Peker et al. (107). Esto refleja una preservación adecuada de la arquitectura genómica, un aspecto crítico en Mtb, donde variaciones mínimas pueden asociarse con resistencia antimicrobiana.

En conjunto, estos resultados demuestran que la fragmentación controlada del ADNg a ~10 kbp  $\pm$  2 kbp constituye una estrategia eficaz para optimizar el rendimiento del secuenciamiento ONT, al incrementar la profundidad sin

comprometer la capacidad de ensamblaje en los métodos evaluados. Además, esta estrategia permite pasar de la secuenciación de una sola muestra por celda de flujo (12) a un esquema de multiplexado de 12–24 muestras por celda de flujo, reduciendo el costo de la celda de flujo por muestra de aproximadamente USD 1000 (11) (una muestra por celda) a ~\$ 83–42 mediante multiplexado, valor sustancialmente inferior a estimaciones previas para WGS en Mtb (~\$ 280) (63).

Nuestros resultados confirmaron que la inactivación a 95 °C/1h eliminó completamente la viabilidad de Mtb, mientras que 80 °C/20 min solo la redujo parcialmente. No obstante, para que la condición de 95 °C/1 h pueda evaluarse en la secuenciación ONT, el método de extracción en conjunto con esta condición debe preservar la integridad del ADNg en niveles al menos cercanos a los adecuados para un buen rendimiento del secuenciamiento; de acuerdo a nuestros resultados obtenidos: DIN cercano a 7, Peak Size ~10 kpb y un porcentaje de fragmentos mayores a 5 kpb cercano al 80%. Entre los métodos evaluados, M3a fue el único que se aproximó simultáneamente a estos parámetros sin requerir fragmentación.

No obstante, nuestros resultados muestran que, aunque el método M3a bajo 80 °C/20 min y NT generó una pureza e integridad comparable al obtenido con el método M2, su rendimiento de secuenciamiento fue inferior a este y al método M3b (Figura 10). Por tanto, aunque M3a bajo 95 °C/1 h conserva una integridad del ADNg cercana a la compatible con lecturas largas, su rendimiento del secuenciamiento observado en 80 °C/20 min y NT indica que no optimizaría el secuenciamiento ONT en términos de eficiencia.

Finalmente, los resultados bajo NT posicionan a M3b y a M2 , on una purificación reforzada, como las opciones más robustas para la obtención de ADNg de alta calidad destinado a buen rendimiento del secuenciamiento ONT del ADNg de Mtb. Mientras que la fragmentación del ADNg a ~10 kpb permitió para M3b obtener las mejores métricas globales del secuenciamiento permitiendo ensamblajes casi completos con lecturas dúplex, y para M2 mostró un rendimiento competitivo aun con menor cantidad de datos, lo que sugiere que ambos métodos pueden sostener ensamblajes de alta cobertura cuando se dispone de condiciones instrumentales equivalentes. Esta estrategia de fragmentación a ~10 kpb tanto para M3b como para M2 se establece para maximizar el rendimiento y obtener ensamblajes de genomas completos, y así viabilizar esquemas de vigilancia genómica y aplicaciones de diagnóstico molecular escalables y sostenibles en contextos de recursos limitados.

## 7 LIMITACIONES

Limitaciones: Este trabajo es una evaluación comparativa con un número limitado de réplicas por condición, lo que reduce la potencia estadística y aumenta la incertidumbre en las estimaciones. El uso de la cepa de referencia H37Rv favorece la reproducibilidad, pero limita la extrapolación a aislados clínicos de linajes diversos. Algunas comparaciones de profundidad se basaron en proyecciones normalizadas por poro-tiempo, las cuales, aunque facilitan la estandarización entre celdas de flujo, pueden sobreestimar el rendimiento in situ y deben interpretarse con cautela. Finalmente, el ensamblaje y la precisión derivados de lecturas dúplex dependen del tamaño del subconjunto disponible, pudiendo ser insuficientes cuando el rendimiento bruto es bajo.

Posibles sesgos : Las métricas de calidad (qScore, tasas de error puntual e índices) son sensibles a la versión del basecaller y al modelo utilizado -p. ej. *High accuracy (HAC)* vs *Super accurate (SUP)*-; cambios de versión podrían desplazar cuantitativamente los resultados. Los residuos químicos potenciales (p. ej., fenol/cloroformo o sales caotrópicas) y variaciones menores en reparación de extremos, *end-prep* y *clean-ups* pueden alterar la distribución real de tamaños, desplazando el N50 respecto del objetivo nominal. Finalmente, el sesgo de selección inherente al uso de lecturas dúplex —más precisas pero menos abundantes— puede favorecer la exactitud del ensamblaje a costa de la cobertura global.

## 8 CONCLUSIONES

- Bajo la inactivación térmica a 80 °C/20 min y 95 °C/1 h se obtuvo cantidades y calidades de ADNg de *M. tuberculosis* comparables, con variaciones dependientes del método de extracción. No obstante, la integridad se vio reducida a 95 °C, con un DIN < 7 y fragmentos cortos.
- El método fenol-cloroformo con lisis mecánica seguida de lisis enzimática, bajo la inactivación a 80 °C/20 min, obtuvo la mayor cantidad, calidad e integridad del ADNg de *M. tuberculosis* frente a los demás métodos de extracción.
- El rendimiento del secuenciamiento ONT del ADNg de *M. tuberculosis* sometido a la inactivación de 80 °C/20 min fue comparable al de la condición sin inactivación térmica, mostrando métricas similares entre ambas.
- Los cuatro métodos de extracción de ADNg de *M. tuberculosis* mostraron métricas de secuenciamiento ONT similares, con excepción de una mayor longitud de read N50 (10.72 kb) y qScore (18.20) en el método basado en el Kit ZymoBIOMICS con lisis mecánica y sin inactivación térmica frente al método basado en fenol cloroformo con lisis enzimática.
- La fragmentación del ADNg a ~10 kpb incremento la profundidad genómica en comparación a la condición no fragmentada en todos los métodos de extracción de ADNg de *M. tuberculosis*, superando los 200 X en los métodos de extracción basados con lisis mecánica con el Kit ZymoBIOMICS (M3b) y con el método fenol cloroformo (M2). Aunque la fragmentación redujo el read N50, se logró un ensamblaje con 99.83% cobertura y con bajas tasas de errores e indels.

## 9 RECOMENDACIONES

De acuerdo a nuestros resultados obtenidos se recomienda las siguientes acciones:

- Evaluar el rendimiento del secuenciamiento ONT para los cuatro métodos de extracción bajo la condición de inactivación térmica a 95 °C/1 h, con el fin de determinar las métricas del secuenciamiento bajo estas condiciones y en un contexto de bioseguridad.
- Dado que 80 °C/20 min no garantiza la inactivación completa y que 95 °C/1 h compromete la integridad del ADNg, se sugiere explorar tiempos de exposición intermedios de 45, 30 o 15 min a 95 °C que permitan equilibrar bioseguridad y preservación de fragmentos largos.
- Evaluar métodos de inactivación química validados, como aquellos basados en guanidina tiocianato (GITC) (108). También pueden considerarse Tris/EDTA, cloroformo o etanol al 70% (109), los cuales han demostrado capacidad de inactivación. El uso de fenol:cloroformo:alcohol isoamílico (25:24:1) es recomendable para inactivar, pero debe acompañarse de una verificación experimental de inviabilidad.
- Evaluar alternativas a la fragmentación mediante g-TUBE - cuyo costo por unidad es \$ 38.9 (110) - para reducir costos operativos. En particular, se recomienda estandarizar la lisis mecánica del método M2, cuyo Peak Size (17.78 kbp) se aproxima al rango de ~10 kpb, o ajustar la lisis mecánica del método M3b para obtener fragmentos de tamaño similar.

- Explorar estrategias de fragmentación que generen tamaños de 5, 10 y 15 kpb ( $\pm 2$  kpb), con el fin de identificar el equilibrio óptimo entre rendimiento, longitud de fragmentos y calidad, según se priorice vigilancia genómica o ensamblaje *de novo*.
- Validar el flujo de trabajo en aislados clínicos de diversos linajes (Beijing, LAM, EAI) y distintas matrices clínicas, evaluando métricas como N50, profundidad genómica, cobertura de genes de resistencia y exactitud del llamado de variantes, asegurando su robustez más allá de H37Rv.
- Realizar comparaciones con lecturas cortas (Illumina) y ensamblajes híbridos para evaluar concordancia de variantes y cobertura de loci de resistencia. Incluir un análisis de costo-rendimiento por muestra para orientar decisiones en contextos con restricciones de presupuesto.

## 10 REFERENCIAS BIBLIOGRÁFICAS

1. World Health Organization. Global Tuberculosis Report 2023. World Health Organization. 2023.
2. World Health Organization. Global Tuberculosis Report 2022. World Health Organization. 2022. p. 68.
3. World Health Organization. Global Tuberculosis Report 2024. World Health Organization. 2024.
4. Yu MC, Hung CS, Huang CK, Wang CH, Liang YC, Lin JC. Integrative utility of long read sequencing-based whole genome analysis and phenotypic assay on differentiating isoniazid-resistant signature of *Mycobacterium tuberculosis*. *J Biomed Sci* [Internet]. 2021;28(1):1–9. Available from: <https://doi.org/10.1186/s12929-021-00783-x>
5. Votintseva AA, Bradley P, Pankhurst L, del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. Tang Y-W, editor. *J Clin Microbiol* [Internet]. 2017 May;55(5):1285–98. Available from: <https://journals.asm.org/doi/10.1128/JCM.02483-16>
6. Murphy SG, Smith C, Lapierre P, Shea J, Patel K, Halse TA, et al. Direct detection of drug-resistant *Mycobacterium tuberculosis* using targeted next generation sequencing. *Front Public Heal*. 2023;11(June).
7. Centers for Disease Control and Prevention. Whole Genome Sequencing | TB Genotyping | TB | CDC [Internet]. Division of Tuberculosis Elimination, National Center for HIV, Viral Hepatitis, STD, and TB Prevention. 2024 [cited 2024 Jan 20]. Available from:

<https://www.cdc.gov/tb/programs/genotyping/genome-sequencing.htm>

8. Vogel M, Utpatel C, Corbett C, Kohl TA, Iskakova A, Ahmedov S, et al. Implementation of whole genome sequencing for tuberculosis diagnostics in a low-middle income, high MDR-TB burden country. *Sci Reports* 2021 111 [Internet]. 2021 Jul 28 [cited 2024 Jan 20];11(1):1–9. Available from: <https://www.nature.com/articles/s41598-021-94297-z>
9. Shea J, Halse TA, Lapierre P, Shudt M, Kohlerschmidt D, Van Roey P, et al. Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York State. *J Clin Microbiol.* 2017;55(6):1871–82.
10. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol.* 2019;17(9):533–45.
11. Zheng P, Zhou C, Ding Y, Liu B, Lu L, Zhu F, et al. Nanopore sequencing technology and its applications. *MedComm.* 2023;4(4):1–34.
12. Bainomugisa A, Duarte T, Lavu E, Pandey S, Coulter C, Marais BJ, et al. A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. *Microb Genomics.* 2018;4(7).
13. Di Marco F, Spitaleri A, Battaglia S, Batignani V, Cabibbe AM, Cirillo DM. Advantages of long- and short-reads sequencing for the hybrid investigation of the *Mycobacterium tuberculosis* genome. *Front Microbiol.* 2023;14(February):1–9.
14. Chitale P, Lemenze AD, Fogarty EC, Shah A, Grady C, Odom-Mabey AR,

- et al. A comprehensive update to the *Mycobacterium tuberculosis* H37Rv reference genome. *Nat Commun.* 2022;13(1):1–12.
15. Eagle SHC, Robertson J, Bastedo DP, Liu K, Nash JHE. Evaluation of five commercial DNA extraction kits using *Salmonella* as a model for implementation of rapid Nanopore sequencing in routine diagnostic laboratories. *Access Microbiol.* 2023;5(2):1–20.
  16. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021;39(11):1348–65.
  17. George S, Xu Y, Rodger G, Morgan M, Sanderson ND, Hoosdally SJ. DNA Thermo-Protection Facilitates Whole-Genome Sequencing of *Mycobacteria* Direct from Clinical Samples. *J Clin Microbiol.* 2020;58(10):1–15.
  18. Oxford Nanopore Technologies. Community - Knowledge - Contaminants [Internet]. Oxford Nanopore Technologies website. 2024 [cited 2024 Jan 20]. Available from: [https://community.nanoporetech.com/contaminants#phenol\\_nanodrop\\_spectra=](https://community.nanoporetech.com/contaminants#phenol_nanodrop_spectra=)
  19. Cabibbe AM, Walker TM, Niemann S, Cirillo DM. Whole genome sequencing of *Mycobacterium tuberculosis*. *Eur Respir J* [Internet]. 2018;52(5):1–5. Available from: <http://dx.doi.org/10.1183/13993003.01163-2018>
  20. Oxford Nanopore Technologies. Community - Knowledge - Extraction methods - Optional fragmentation of gDNA [Internet]. Oxford Nanopore Technologies website. 2024 [cited 2024 Jan 20]. Available from:

[https://community.nanoporetech.com/extraction\\_method\\_groups/optional-fragmentation-of-gdna](https://community.nanoporetech.com/extraction_method_groups/optional-fragmentation-of-gdna)

21. Gómez-González PJ, Campino S, Phelan JE, Clark TG. Portable sequencing of *Mycobacterium tuberculosis* for clinical and epidemiological applications. *Brief Bioinform.* 2022;23(5):1–10.
22. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel).* 2023;12(7):1–25.
23. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One* [Internet]. 2021;16(10 October). Available from: <http://dx.doi.org/10.1371/journal.pone.0257521>
24. Dippenaar A, Goossens SN, Grobbelaar M, Oostvogels S, Cuypers B, Laukens K, et al. Nanopore Sequencing for *Mycobacterium tuberculosis* : a Critical Review of the Literature, New Developments, and Future Opportunities. *J Clin Microbiol.* 2022;60(1).
25. Psifidi A, Dovas CI, Bramis G, Lazou T, Russel CL, Arsenos G, et al. Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. *PLoS One.* 2015;10(1).
26. Koch A, Mizrahi V. *Mycobacterium tuberculosis*. *Trends Microbiol.* 2018;26(6):555–6.
27. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med.* 2016;13(10):1–13.

28. Menzies NA, Wolf E, Connors D, Bellerose M, Sbarra AN, Cohen T, et al. Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions. *Lancet Infect Dis* [Internet]. 2018;18(8):e228–38. Available from: [http://dx.doi.org/10.1016/S1473-3099\(18\)30134-8](http://dx.doi.org/10.1016/S1473-3099(18)30134-8)
29. González-martin J. Microbiología de la tuberculosis. *Semin Fund Esp Reum* 2014;15(1)25–33. 2014;15(1):25–33.
30. Lado L, Hospitalario C, Coru L. Diagnóstico microbiológico de la tuberculosis. 2002;39(5):207–15.
31. Briffotiaux J, Liu S, Gicquel B. Genome-wide transcriptional responses of *Mycobacterium* to antibiotics. *Front Microbiol*. 2019;10(FEB):1–14.
32. Pan American Health Organization (PAHO) & World Health Organization (WHO). WHO Operational Handbook on Tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update. In: WHO Operational Handbook on Tuberculosis [Internet]. 2022. p. 118. Available from: [https://iris.paho.org/bitstream/handle/10665.2/55927/9789275325377\\_spa.pdf?sequence=1&isAllowed=y](https://iris.paho.org/bitstream/handle/10665.2/55927/9789275325377_spa.pdf?sequence=1&isAllowed=y)
33. Santos-Lazaro D, Gavilan RG, Solari L, Vigo AN, Puyen ZM. Whole genome analysis of extensively drug resistant *Mycobacterium tuberculosis* strains in Peru. *Sci Rep* [Internet]. 2021;11(1):1–12. Available from: <https://doi.org/10.1038/s41598-021-88603-y>
34. Caminero JA, Cayla JA, García-García JM, García-Pérez FJ, Palacios JJ, Ruiz-Manzano J. Diagnóstico y tratamiento de la tuberculosis con resistencia a fármacos. *Arch Bronconeumol*. 2017;53(9):501–9.

35. Pan American Health Organization (PAHO) & World Health Organization (WHO). Factsheet: Drug-Resistant Tuberculosis in the Americas Region. 2022 [Internet]. Internet. 2024. Available from: <https://www.paho.org/es/documentos/hoja-informativa-tuberculosis-resistente-medicamentos-region-americas-2022>
36. Ministerio de Salud del Perú (MINSA). Minsa inicia tratamiento totalmente oral para la tuberculosis estrictamente multidrogorresistente (MDR) en el país [Internet]. Internet. 2023. Available from: <https://www.gob.pe/institucion/minsa/noticias/697909-minsa-inicia-tratamiento-totalmente-oral-para-la-tuberculosis-estrictamente-multidrogorresistente-mdr-en-el-pais>
37. Ministerio de Salud del Perú (MINSA). Minsa intensifica búsqueda de casos de tuberculosis en poblaciones vulnerables [Internet]. Internet. 2024. Available from: <https://www.gob.pe/institucion/minsa/noticias/996224-minsa-intensifica-busqueda-de-casos-de-tuberculosis-en-poblaciones-vulnerables>
38. World Health Organization (WHO). WHO operational handbook on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection, third edition. [Internet]. World Health Organization. 2024. 56 p. Available from: <https://iris.who.int/>.
39. Steingart, KR, Ramsay, A., y Pai M. Optimizing sputum smear microscopy for the diagnosis of pulmonary tuberculosis. *Expert Rev Anti Infect Ther.* 2007;5(3):327–331.
40. Das PK, Ganguly SB MB. Sputum Smear Microscopy in Tuberculosis: It Is

- Still Relevant in the Era of Molecular Diagnosis When Seen from the Public Health Perspective. *Biomed Biotechnol Res J.* 2019;3(1):77–9.
41. Cao Y, Parmar H, Gaur RL, Lieu D, Raghunath S, Via N, et al. Xpert MTB / XDR : a 10-Color Reflex Assay Suitable for Point-of- Care Settings To Detect Isoniazid , Fluoroquinolone , and Second-Line-Injectable-Drug Resistance Directly from. *J Clin Microbiol.* 2021;59(3):1–18.
  42. Xiao YX, Liu KH, Lin WH, Chan TH, Jou R. Whole-genome sequencing-based analyses of drug-resistant *Mycobacterium tuberculosis* from Taiwan. *Sci Rep* [Internet]. 2023;13(1):1–11. Available from: <https://doi.org/10.1038/s41598-023-29652-3>
  43. World Health Organization. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide [Internet]. World Health Organization. 2018. Available from: <https://apps.who.int/iris/handle/10665/274443>
  44. World Health Organization. The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual [Internet]. World Health Organization. 2023. p. 134. Available from: <https://www.who.int/publications/i/item/9789240078079>
  45. World Health Organization (WHO). Use of targeted next-generation sequencing to detect drug-resistant tuberculosis: rapid communication [Internet]. World Health Organization. 2023. Available from: <https://www.who.int/publications/i/item/9789240076372>
  46. Ness TE, DiNardo A, Farhat MR. High Throughput Sequencing for Clinical

- Tuberculosis: An Overview. *Pathogens*. 2022;11(11):1–9.
47. Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. *Microorg Sustain*. 2019;17:313–41.
  48. Slatko, B. E., Gardner, A. F., & Ausubel FM. Overview of Next Generation Sequencing Technologies. *Curr Protoc Mol Biol*. 2018;176(1):139–48.
  49. Mandlik JS, Patil AS SS. Next-generation sequencing (NGS): Platforms and applications. *J Pharm Bioallied Sci*. 2024;16(4):1–5.
  50. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics, Proteomics Bioinforma* [Internet]. 2015;13(5):278–89. Available from: <http://dx.doi.org/10.1016/j.gpb.2015.08.002>
  51. Thorpe J, Sawaengdee W, Ward D, Campos M, Wichukchinda N, Chaiyasirinroje B, et al. Multi-platform whole genome sequencing for tuberculosis clinical and surveillance applications. *Sci Rep* [Internet]. 2024;14(1):1–9. Available from: <https://doi.org/10.1038/s41598-024-55865-1>
  52. Illumina. Sequencing platforms [Internet]. Internet. 2024. Available from: <https://www.illumina.com/systems/sequencing-platforms.html>
  53. Ravi RK, Walton K, Khosroheidari M. Miseq: A next generation sequencing platform for genomic analysis. *Methods Mol Biol*. 2018;1706:223–32.
  54. Twyford AD. Will benchtop sequencers resolve the sequencing trade-off in plant genetics? *Front Plant Sci*. 2016;7(APR2016):6–8.
  55. Stefan CP, Hall AT, Graham AS, Minogue TD. Comparison of Illumina and Oxford Nanopore Sequencing Technologies for Pathogen Detection from

- Clinical Matrices Using Molecular Inversion Probes. *J Mol Diagnostics* [Internet]. 2022;24(4):395–405. Available from: <https://doi.org/10.1016/j.jmoldx.2021.12.005>
56. The Genomics Services Company. Fact Sheet: Illumina NextSeq Sequencing Systems [Internet]. Internet. 2024. Available from: <https://www.cd-genomics.com/blog/nextseq/>
  57. Illumina. Quality Scores for Next-Generation Sequencing [Internet]. Internet. 2011. p. 1–2. Available from: [https://www.illumina.com/documents/products/technotes/technote\\_Q-Scores.pdf](https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf)
  58. Oxford Nanopore Technologies. From kilobases to “whales”: a short history of ultra-long reads and high-throughput genome sequencing [Internet]. Oxford Nanopore Technologies website. 2021. Available from: <https://nanoporetech.com/blog/news-blog-kilobases-whales-short-history-ultra-long-reads-and-high-throughput-genome>
  59. Tianyuan Zhang, Hanzhou Li, Mian Jiang, Huiyu Hou, Yunyun Gao, Yali Li, Fuhao Wang, Jun Wang, Kai Peng Y-XL. Nanopore sequencing: flourishing in its teenage years. *J Genet Genomics* [Internet]. 2024;51(12):1361–74. Available from: <https://www.sciencedirect.com/science/article/pii/S1673852724002443>
  60. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet* [Internet]. 2018;34(9):666–81. Available from: <https://doi.org/10.1016/j.tig.2018.05.008>
  61. Oxford Nanopore Technologies. MinION Mk1D [Internet]. Oxford

- Nanopore Technologies website. 2024. Available from: <https://store.nanoporetech.com/minion.html>
62. Pinzauti D, Iannelli F, Pozzi G, Santoro F. DNA isolation methods for Nanopore sequencing of the *Streptococcus mitis* genome. *Microb Genomics*. 2022;8(2):1–8.
  63. Bouso JM, Planet PJ. Complete nontuberculous mycobacteria whole genomes using an optimized DNA extraction protocol for long-read sequencing. *BMC Genomics*. 2019;20(1):793.
  64. Lermينياux N, Fakharuddin K, Mulvey MR, Mataseje L. Do we still need Illumina sequencing data ? Evaluating Oxford Nanopore Technologies R10 .4 . 1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. *Can J Microbiol*. 2024;70:178–89.
  65. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res*. 2015;25(11):1750–6.
  66. David M, Dursi LJ, Yao D, Boutros PC, Simpson JT. Nanocall: An open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics*. 2017;33(1):49–55.
  67. Boža V, Brejová B, Vinař T. DeepNano: Deep recurrent neural networks for base calling in MinION Nanopore reads. *PLoS One*. 2017;12(6):1–13.
  68. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. 2019;20(1):1–10.
  69. Oxford Nanopore Technologies. Nanopore sequencing accuracy [Internet].

- Oxford Nanopore Technologies website. 2025. Available from:  
<https://nanoporetech.com/es/platform/accuracy>
70. Oxford Nanopore Technologies. How basecalling works [Internet]. Oxford Nanopore Technologies website. 2025. Available from:  
<https://nanoporetech.com/es/platform/technology/basecalling>
71. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. Oxford Nanopore R10 . 4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*. 2022;19:823–6.
72. Wick, R.R., Judd, L.M., and Holt KE. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* [Internet]. 2023;1–12. Available from:  
<http://dx.doi.org/10.1371/journal.pcbi.1010905>
73. De Coster W, De Rijk P, De Roeck A, De Pooter T, D’Hert S, Strazisar M, et al. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res*. 2019;29(7):1178–87.
74. Nicholls, S. M., Quick, J. C., Tang, S. & Loman NJ. Ultra-deep , long-read nanopore sequencing of mock microbial community standards. *Gigascience*. 2019;8(December 2018):1–9.
75. Oxford Nanopore Technologies. PromethION: Flexible, large-scale, direct DNA and RNA sequencing [Internet]. Oxford Nanopore Technology website. Available from:  
<https://nanoporetech.com/products/sequence/promethion>

76. Oxford Nanopore Technologies. Native Barcoding Kit 24 V14 SQK-NBD114.24 [Internet]. Oxford Nanopore Technologies website. Available from: <https://store.nanoporetech.com/uk/native-barcoding-kit-24-v14.html>
77. Oxford Nanopore Technologies. Chemistry Technical Document [Internet]. Oxford Nanopore Technologies website. 2024. Available from: <https://nanoporetech.com/es/document/chemistry-technical-document>
78. Oxford Nanopore Technologies. Rapid Barcoding Kit 24 V14 SQK-RBK114.24 [Internet]. Oxford Nanopore Technologies website. Available from: <https://store.nanoporetech.com/uk/rapid-barcoding-sequencing-kit-24-v14.html>
79. Jaudou S, Tran ML, Vorimore F, Fach P, Delannoy S. Evaluation of high molecular weight DNA extraction methods for long-read sequencing of Shiga toxin-producing *Escherichia coli*. *PLoS One* [Internet]. 2022;17(7 July):1–12. Available from: <http://dx.doi.org/10.1371/journal.pone.0270751>
80. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun*. 2015;6.
81. Oxford Nanopore Technologies. Celebrating 10 years of innovation in TB research by the Nanopore Community [Internet]. Oxford Nanopore Technologies website. 2024. Available from: <https://nanoporetech.com/blog/celebrating-10-years-of-innovation-in-tb-research-by-the-nanopore-community>
82. Smith C, Halse TA, Shea J, Modestil H, Fowler RC. Assessing Nanopore

- Sequencing for Clinical Diagnostics : a for Mycobacterium tuberculosis. *J Clin Microbiol.* 2021;59(1):1–14.
83. Cervantes J, Yokobori N, Hong BY. Genetic identification and drug-resistance characterization of mycobacterium tuberculosis using a portable sequencing device. A pilot study. *Antibiotics.* 2020;9(9):1–15.
84. Tafess K, Ng TTL, Lao HY, Leung KSS, Tam KKG, Rajwani R, et al. Targeted-Sequencing Workflows for Comprehensive Drug Resistance Profiling of Mycobacterium tuberculosis Cultures Using Two Commercial Sequencing Platforms: Comparison of Analytical and Diagnostic Performance, Turnaround Time, and Cost. *Clin Chem.* 2020;66(6):809–20.
85. Chan WS, Au CH, Chung Y, Leung HCM, Ho DN, Wong EYL, et al. Rapid and economical drug resistance profiling with Nanopore MinION for clinical specimens with low bacillary burden of Mycobacterium tuberculosis. *BMC Res Notes* [Internet]. 2020;13(1):1–7. Available from: <https://doi.org/10.1186/s13104-020-05287-9>
86. Cabibbe AM, Spitaleri A, Battaglia S, Colman RE, Suresh A, Uplekar S, et al. Application of targeted next-generation sequencing assay on a portable sequencing platform for culture-free detection of drug-resistant tuberculosis from clinical samples. *J Clin Microbiol.* 2020;58(10).
87. Su J, Lui WW, Lee YL, Zheng Z, Siu GKH, Ng TTL, et al. Evaluation of Mycobacterium tuberculosis enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Sci Rep* [Internet]. 2023;13(1):1–10. Available from: <https://doi.org/10.1038/s41598-023-32378-x>

88. Somerville W, Thibert L, Schwartzman K, Behr MA. Extraction of *Mycobacterium tuberculosis* DNA: A question of containment. *J Clin Microbiol.* 2005;43(6):2996–7.
89. Zymo Research. ZymoBIOMICS™ DNA Miniprep Kit Table of Contents [Internet]. Zymo Research website. 2023. Available from: [https://files.zymoresearch.com/protocols/\\_d3024\\_d3025\\_quick-dna\\_miniprep\\_kit.pdf](https://files.zymoresearch.com/protocols/_d3024_d3025_quick-dna_miniprep_kit.pdf)
90. Van Soolingen D, Hermans PWM, De Haas PEW, Soll DR, Van Embden JDA. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: Evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol.* 1991;29(11):2578–86.
91. Cauna Y. Estandarización de una técnica de complementación del gen *pncA* en *Mycobacterium tuberculosis* *pncA*-knockout: Herramienta para el estudio de la relación entre mutaciones en *pncA* y parámetros fenotípicos. Tesis para optar el grado de Maestro en Bioquímica y Biología molecular. Universidad Peruana Cayetano Heredia; 2019.
92. Rodriguez J, Alcántara R, Rodríguez J, Vargas J, Roncal E, Antiparra R, et al. Evaluation of three alternatives cost-effective culture media for *Mycobacterium tuberculosis* detection and drug susceptibility determination using the microscopic observation drug susceptibility (MODS) assay. *Tuberculosis.* 2022;137.
93. Alcántara et al. 2019. crossm MODS-Wayne , a Colorimetric Adaptation of the Microscopic- of *Mycobacterium tuberculosis* Pyrazinamide Resistance

- from. 2019;(August 2018):1–12.
94. Wang C-H, Putri DU, Lee J-C, Liao C-C, Tsao S, Hsiao A-L, et al. Biosafety and Proteome Profiles of Different Heat Inactivation Methods for *Mycobacterium tuberculosis*. *Microbiol Spectr*. 2021;9(3):1–12.
  95. Castro C, González L, Rozo JC, Puerto G, Ribón W. Biosafety evaluation of the DNA extraction protocol for *Mycobacterium tuberculosis* complex species, as implemented at the Instituto Nacional de Salud, Colombia. *Biomedica*. 2009;29(4):561–6.
  96. Käser M, Ruf MT, Hauser J, Pluschke G. Optimized DNA preparation from mycobacteria. *Cold Spring Harb Protoc*. 2010;5(4).
  97. Epperson LE, Strong M. A scalable, efficient, and safe method to prepare high quality DNA from mycobacteria and other challenging cells. *J Clin Tuberc Other Mycobact Dis* [Internet]. 2020;19:100150. Available from: <https://doi.org/10.1016/j.jctube.2020.100150>
  98. Jacobo-Delgado YM, Rodríguez-Carlos A, Serrano CJ, Rivas-Santiago B. *Mycobacterium tuberculosis* cell-wall and antimicrobial peptides: a mission impossible? *Front Immunol*. 2023;14(May):1–18.
  99. Koetsier G, Cantor E. A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers. New Engl Biolabs Inc [Internet]. 2019;(1):1–8. Available from: [www.BioTek.com](http://www.BioTek.com)
  100. Doig C, Seagar AL, Watt B, Forbes KJ. The efficacy of the heat killing of *Mycobacterium tuberculosis*. *J Clin Pathol*. 2002;55(10):778–9.
  101. Gand M, Bloemen B, Vanneste K, Roosens NHC, De Keersmaecker SCJ. Comparison of 6 DNA extraction methods for isolation of high yield of high

- molecular weight DNA suitable for shotgun metagenomics Nanopore sequencing to detect bacteria. *BMC Genomics* [Internet]. 2023;24(1):1–14. Available from: <https://doi.org/10.1186/s12864-023-09537-5>
102. Kalendar R, Ivanov KI, Samuilova O, Kairov U, Zamyatnin AA. Isolation of High-Molecular-Weight DNA for Long-Read Sequencing Using a High-Salt Gel Electroelution Trap. *Anal Chem*. 2023;95(48):17818–25.
  103. Elton L, Aydin A, Stoker N, Rofael S, Wildner LM, Abdul JBPAA, et al. A pragmatic pipeline for drug resistance and lineage identification in *Mycobacterium tuberculosis* using whole genome sequencing. *PLOS Glob Public Heal* [Internet]. 2025;5(2 February):1–18. Available from: <https://doi.org/10.1371/journal.pgph.0004099>
  104. Kent L, McHugh TD, Billington O, Dale JW, Gillespie SH. Demonstration of homology between IS6110 of *Mycobacterium tuberculosis* and DNAs of other *Mycobacterium* spp. *J Clin Microbiol*. 1995;33(9):2290–3.
  105. Boughattas S, Albatash D, Al-Khater A, Giraldez BW, Althani AA, Benslimane FM. Whole genome sequencing of marine organisms by Oxford Nanopore Technologies: Assessment and optimization of HMW-DNA extraction protocols. *Ecol Evol*. 2021;11(24):18505–13.
  106. Alhakami H, Mirebrahim H, Lonardi S. A comparative evaluation of genome assembly reconciliation tools. *Genome Biol* [Internet]. 2017;18(1):1–14. Available from: <http://dx.doi.org/10.1186/s13059-017-1213-3>
  107. Peker N, Schuele L, Kok N, Terrazos M, Neuenschwander SM, De Beer J, et al. Evaluation of whole-genome sequence data analysis approaches for short- and long-read sequencing of *Mycobacterium tuberculosis*. *Microb*

- Genomics. 2021;7(11).
108. Mtafya B, Qwaray P, John J, Sichone E, Shoo A, Gillespie SH, et al. A practical approach to render tuberculosis samples safe for application of tuberculosis molecular bacterial load assay in clinical settings without a biosafety level 3 laboratory ☆. *Tuberculosis* [Internet]. 2023;138(November 2022):102275. Available from: <https://doi.org/10.1016/j.tube.2022.102275>
  109. Vigo AN, Puyén ZM, Santos-lázaro D, Perea ML, Solari L. Methods for the Inactivation of *Mycobacterium tuberculosis* : A Systematic Review of the Literature. 2024;237–46.
  110. Covaris Inc. g-TUBE for DNA Shearing [Internet] [Internet]. Woburn (MA): Covaris. 2025. Available from: <https://www.covaris.com/products-services/consumables/g-tube>

## 11 ANEXOS

**ANEXO 1 :** Variables de la cantidad, calidad e integridad del ADN<sub>g</sub> obtenido bajo dos tratamientos térmicos (95 °C por 1h y 80 °C por 20 min) y el control NT en diferentes métodos de extracción.

Método de extracción	Condición de inactivación térmica	ADN total (ug)	Ratio 260/280	Ratio 260/230	DIN	Peak Size (Kpb)	Fragmentos ≥ 5kpb (%)
M1	95 °C por 1h	19.26 (17.16 - 22.59)	1.95 (1.92 - 1.95)	2.27 (2.18 - 2.30)	6.60 (6.53 - 6.68)	22.04 (20.66 - 25.33)	67.59 (66.02 - 69.38)
	80 °C por 20min	19.46 (18.2 - 21.14)	1.95 (1.94 - 1.95)	2.56 (2.54 - 2.58)	7.90 (7.75 - 7.98)	48.00 (45.53 - 49.55)	88.77 (87.88 - 88.85)
	NT	49.98 (47.33 - 52.96)	1.95 (1.93 - 1.96)	2.30 (2.26 - 2.36)	8.05 (7.68 - 8.58)	55.80 (53.51 - 56.15)	93.83 (87.40 - 99.01)
M2	95 °C por 1h	91.53 (82.58 - 109.94)	1.93 (1.91 - 1.94)	2.06 (2.04 - 2.07)	6.00 (5.93 - 6.08)	10.11 (9.40 - 10.51)	55.36 (55.17 - 55.40)
	80 °C por 20min	36.99 (32.34 - 51.60)	1.94 (1.93 - 1.94)	2.10 (2.08 - 2.10)	7.8 (7.73 - 7.88)	20.25 (19.10 - 21.18)	90.95 (90.80 - 91.98)
	NT	3.31 (2.14 - 3.43)	1.55 (1.52 - 1.56)	1.21 (1.14 - 1.28)	7.65 (7.60 - 7.85)	17.78 (17.55 - 17.90)	89.11 (85.67 - 91.42)
M3a	95 °C por 1h	7.55 (7.33 - 7.90)	1.90 (1.89 - 1.90)	1.71 (1.68 - 1.75)	6.90 (6.75 - 6.90)	12.37 (11.70 - 12.54)	78.96 (77.06 - 84.98)
	80 °C por 20min	2.21 (2.19 - 2.22)	1.91 (1.89 - 1.91)	1.41 (1.38 - 1.44)	7.80 (7.55 - 7.98)	15.53 (14.87 - 16.37)	91.34 (87.25 - 97.00)
	NT	2.80 (2.67 - 2.93)	1.89 (1.86 - 1.91)	1.54 (1.43 - 1.58)	7.80 (7.73 - 7.80)	16.01 (15.77 - 17.02)	90.67 (90.52 - 90.87)
M3b	95 °C por 1h	3.03 (2.48 - 3.22)	1.91 (1.90 - 1.94)	1.19 (1.09 - 1.24)	6.60 (6.53 - 6.75)	18.48 (17.59 - 20.18)	65.34 (62.79 - 71.18)
	80 °C por 20min	9.19 (8.37 - 9.66)	1.90 (1.89 - 1.90)	1.72 (1.71 - 1.73)	8.35 (8.30 - 8.48)	26.50 (25.43 - 27.73)	94.99 (94.49 - 95.55)
	NT	12.34 (11.74 - 13.80)	1.90 (1.89 - 1.90)	1.88 (1.85 - 1.90)	8.80 (8.60 - 9.08)	54.99 (50.57 - 56.44)	93.84 (93.36 - 95.10)

**ANEXO 2:** Variables de la cantidad, calidad e integridad del ADNg obtenido bajo cuatro métodos de extracción de ADNg en dos tratamientos de inactivación térmica (95 °C por 1h y 80 °C por 20 min) y el control NT.

Condición de inactivación térmica	Método de extracción	ADN total (ug)	Ratio 260/280	Ratio 260/230	DIN	Peak Size (Kpb)	Fragmentos $\geq$ 5kpb (%)
95 °C por 1h	M1	19.26 (17.16 - 22.59)	1.95 (1.92 - 1.95)	2.27 (2.18 - 2.30)	6.60 (6.53 - 6.68)	22.04 (20.66 - 25.33)	67.59 (66.02 - 69.38)
	M2	91.53 (82.58 - 109.94)	1.93 (1.91 - 1.94)	2.06 (2.04 - 2.07)	6.00 (5.93 - 6.08)	10.11 (9.40 - 10.51)	55.36 (55.17 - 55.40)
	M3a	7.55 (7.33 - 7.90)	1.90 (1.89 - 1.90)	1.71 (1.68 - 1.75)	6.90 (6.75 - 6.90)	12.37 (11.70 - 12.54)	78.96 (77.06 - 84.98)
	M3b	3.03 (2.48 - 3.22)	1.91 (1.90 - 1.94)	1.19 (1.09 - 1.24)	6.60 (6.53 - 6.75)	18.48 (17.59 - 20.18)	65.34 (62.79 - 71.18)
	M1	19.46 (18.2 - 21.14)	1.95 (1.94 - 1.95)	2.56 (2.54 - 2.58)	7.90 (7.75 - 7.98)	48.00 (45.53 - 49.55)	88.77 (87.88 - 88.85)
80 °C por 20min	M2	36.985 (32.34 - 51.60)	1.94 (1.93 - 1.94)	2.10 (2.08 - 2.10)	7.8 (7.73 - 7.88)	20.25 (19.10 - 21.18)	90.95 (90.80 - 91.98)
	M3a	2.21 (2.19 - 2.22)	1.91 (1.89 - 1.91)	1.41 (1.38 - 1.44)	7.80 (7.55 - 7.98)	15.53 (14.87 - 16.37)	91.34 (87.25 - 97.00)
	M3b	9.19 (8.37 - 9.66)	1.90 (1.89 - 1.90)	1.72 (1.71 - 1.73)	8.35 (8.30 - 8.48)	26.50 (25.43 - 27.73)	94.99 (94.49 - 95.55)
	M1	49.98 (47.33 - 52.96)	1.95 (1.93 - 1.96)	2.30 (2.26 - 2.36)	8.05 (7.68 - 8.58)	55.80 (53.51 - 56.15)	93.83 (87.40 - 99.01)
NT	M2	3.305 (2.14 - 3.43)	1.55 (1.52 - 1.56)	1.21 (1.14 - 1.28)	7.65 (7.60 - 7.85)	17.78 (17.55 - 17.90)	89.11 (85.67 - 91.42)
	M3a	2.8 (2.67 - 2.93)	1.89 (1.86 - 1.91)	1.54 (1.43 - 1.58)	7.80 (7.73 - 7.80)	16.01 (15.77 - 17.02)	90.67 (90.52 - 90.87)
	M3b	12.34 (11.74 - 13.80)	1.90 (1.89 - 1.90)	1.88 (1.85 - 1.90)	8.80 (8.60 - 9.08)	54.99 (50.57 - 56.44)	93.84 (93.36 - 95.10)

**ANEXO 3: Métricas de la calidad de ensamblaje: Estadísticas de la longitud del read NG50**

Método de extracción	Condición de fragmentación	n	Mediana (bases)	Q1	Q3	Promedio	DS (desviación estándar)
M1	F a 10 kpb	6	12151	8628.5	16984.5	13025	8390.21
M1	NF a 10 kpb	6	17903	13585	22221	17903	12213.15
M2	F a 10 kpb	6	91682.5	62390	135456	102273.83	67061.51
M2	NF a 10 kpb	6	52131	17995	73742	153955.8	257026
M3b	F a 10 kpb	12	896913	466303.25	2592616.25	1538008.33	1472969.09
M3b	NF a 10 kpb	6	117807	66845	569647.5	385059.33	553514.42

**ANEXO 4: Métrica de la calidad de ensamblaje : Estadísticas de número de Mismatches por 100 kpb**

Método de extracción	Condición de fragmentación	n	Mediana	Q1	Q3	Promedio	DS (desviación estándar)
M1	F a 10 kpb	6	81.98	77.36	142	106.26	49.61
M1	NF a 10 kpb	6	59.8	55.3	79.1	59.04	27.03
M2	F a 10 kpb	6	13.27	11.53	18.1	22.18	24.1
M2	NF a 10 kpb	6	34.88	15.05	73.1	48.42	42.37
M3b	F a 10 kpb	12	3.34	3.13	4.49	3.87	1.23
M3b	NF a 10 kpb	6	64.58	24.06	111	80.46	77.72

**ANEXO 5: Métrica de la calidad de ensamblaje : Estadísticas de número de Indels  
por 100 kpb**

Método de extracción	Condición de fragmentación	n	Mediana	Q1	Q3	Promedio	DS (desviación estándar)
M1	F a 10 kpb	6	4.54	4.5	4.73	4.59	0.22
M1	NF a 10 kpb	6	3.91	3.31	4.04	3.69	0.57
M2	F a 10 kpb	6	5.74	5.57	5.88	5.72	0.18
M2	NF a 10 kpb	6	5.19	5.02	5.32	5.25	0.49
M3b	F a 10 kpb	12	5.13	4.97	5.3	5.14	0.18
M3b	NF a 10 kpb	6	4.05	3.46	4.73	4.08	0.75