



# Introducción a las nuevas tecnologías de secuenciación: MinION, secuenciación en tiempo real

<b>Apellidos, nombre</b>	Arrones Olmo, Andrea (anarol1@etsiamn.upv.es) Vilanova Navarro, Santiago (sanvina@upvnet.upv.es) Plazas Ávila, Mariola (maplaav@btc.upv.es)
<b>Departamento</b>	Instituto de Conservación y Mejora de la Agrodiversidad Valenciana
<b>Centro</b>	Universitat Politècnica de València



## 1 Resumen de las ideas clave

Desde el descubrimiento de la estructura del ácido desoxirribonucleico (ADN) por Watson, Crick y Franklin en 1953, la comunidad científica ha realizado continuos esfuerzos para secuenciar el ADN. La primera tecnología de Sanger era increíblemente precisa, pero requería mucho tiempo de preparación, lo que aumentaba el coste del proceso y de los secuenciadores. Sin embargo, la secuenciación del ADN ha avanzado mucho desde Sanger. Las tecnologías de secuenciación de nueva generación (NGS) consiguieron reducir drásticamente el coste del procesado y aumentaron el rendimiento por muestra, pero no el coste de los secuenciadores. En este sentido, la secuenciación genómica de tercera generación Nanopore ha revolucionado el mercado debido a su bajo coste, alta repetibilidad y facilidad de preparación de muestras. En especial, el secuenciador MinION desarrollado por Oxford Nanopore Technologies es un dispositivo económico, portátil y de mecanismo sencillo para la secuenciación de ADN en tiempo real. En este artículo proponemos un modelo que une conceptos experimentales y bioinformáticos utilizando el MinION para determinar el perfil microbiano de diferentes muestras de suelo gracias a las regiones conservadas y variables del gen ARNr 16S.

## 2 Objetivos

Una vez que el alumnado lea con detenimiento este documento será capaz de:

- Desarrollar un protocolo de extracción de ADN.
- Amplificar por reacción en cadena de la polimerasa (PCR) los genes candidatos.
- Construir librerías genómicas.
- Manipular y cargar muestras genómicas en un secuenciador MinION.
- Analizar las secuencias o lecturas brutas generadas por el secuenciador.

## 3 Introducción

La comprensión de la secuencia del genoma es a menudo la clave para responder a importantes cuestiones biológicas. Sin embargo, los elevados costes asociados a esta técnica impedían que se pudiera secuenciar de forma rutinaria en los laboratorios de investigación. Con la aparición de las tecnologías de secuenciación de nueva generación (NGS) se consiguió secuenciar genomas completos con una drástica reducción de tiempo y costes (Imagen 1). La secuenciación genómica de tercera generación es una oportunidad única para que los estudiantes exploren tecnologías NGS y puedan ver en directo el proceso completo de secuenciación de una muestra, desde la extracción del ADN hasta el análisis de las lecturas generadas. Es en este último paso donde reside el problema actual debido a la gran cantidad de secuencias que se generan en tiempo récord y que suponen un desafío para los métodos actuales de análisis de datos.

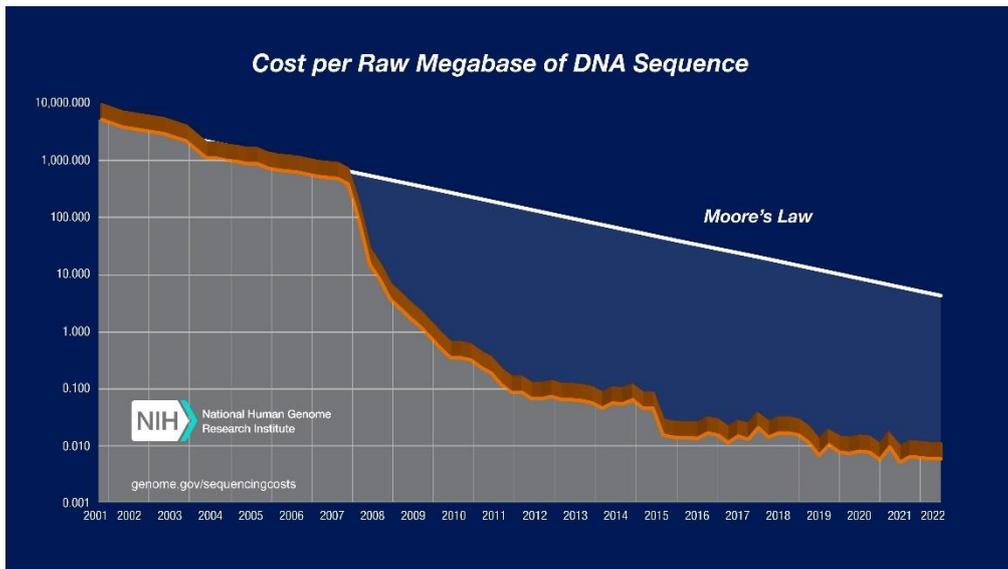


Imagen 1. Serie histórica del coste por megabase (Mb) de secuencia de DNA. Datos del National Human Genome Research Institute (NHGRI, <https://www.genome.gov>).

Oxford Nanopore Technologies (ONT) ha desarrollado MinION, un secuenciador de ADN “de bolsillo”, económico y fácil de usar, para la secuenciación en tiempo real de cualquier tipo de muestras. Hasta el momento, las clases prácticas de genómica se centraban exclusivamente en el análisis masivo de datos debido a los elevados costes del procesamiento de las muestras y de los secuenciadores. Sin embargo, este dispositivo de vanguardia permite llevar la secuenciación a una clase práctica, en un laboratorio sin equipamiento especializado, ya que la preparación de las muestras es muy sencilla y el aparato es capaz de conectarse a cualquier ordenador a través de un puerto USB 3.0.

En este artículo se describe un protocolo para analizar el perfil microbiano de muestras de suelo. El perfil microbiano es la composición o conjunto de microorganismos que vive en un ambiente determinado. Para ello, se extraerá el ADN total de la muestra y se amplificará por PCR el gen ARNr 16S utilizando cebadores específicos. El gen ARNr 16S codifica para el ARN ribosomal 16S, componente de la subunidad menor de los ribosomas que interviene en la traducción del ARNm y síntesis de proteínas. Este gen es comúnmente utilizado para llevar a cabo estudios filogenéticos, ya que su secuencia posee regiones altamente conservada entre las distintas especies de bacterias y arqueas intercaladas con regiones variables que permiten identificar los distintos taxones microbianos presentes en una muestra (Imagen 2).

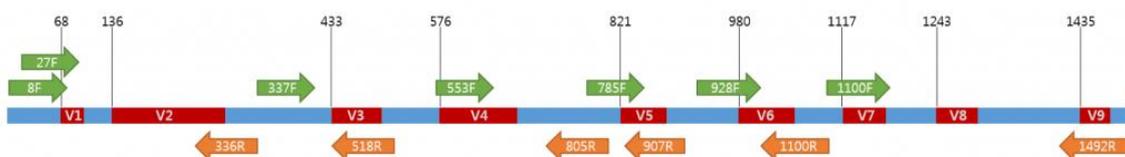


Imagen 2. Representación del gen ARNr 16S. En azul y rojo se muestran las regiones conservadas y variables de la secuencia del gen, respectivamente. En verde y naranja se muestran los cebadores universales directos y reversos diseñados sobre la secuencia del gen, respectivamente. Fuente: <https://help.ezbiocloud.net/>

Una vez generadas las secuencias del gen ARNr 16S por MinION, se requiere de un análisis bioinformático para la clasificación de las mismas. En la actualidad, existe un gran número de softwares específicos de análisis de secuencias, pero a menudo se basan en línea de comandos, lo que dificulta su aprendizaje y manejo. Sin embargo, existen programas que recogen un conjunto amplio de herramientas necesarias para analizar datos de secuenciación en una interfaz gráfica e intuitiva. En concreto, Nanogalaxy es un servidor web (<https://nanopore.usegalaxy.eu>) específico para procesar, analizar y visualizar datos de ONT o datos de secuenciación de lectura larga de tecnologías similares. Puede utilizarse para diversas aplicaciones, incluido el ensamblaje de genomas *de novo* a partir de secuencias metagenómicas para realizar clasificaciones taxonómicas. Más concretamente, tiene integrado un software de clasificación, “Kraken2”, capaz de asignar, eficientemente, taxones a lecturas largas comparándolas con un compendio de secuencias recogidas en una base de datos. NanoGalaxy proporciona flujos de trabajo mejorados, que incluyen desde el post-procesamiento de los datos brutos (eliminación de adaptadores, filtrado de lecturas, etc.), hasta la visualización de los resultados como gráficos circulares interactivos.

## 4 Desarrollo

Este artículo docente combina procesos experimentales con análisis bioinformáticos. Se incluyen metodologías rutinarias de un laboratorio de investigación, como la extracción de ADN y amplificación de un gen candidato, así como el análisis de las secuencias brutas generadas por un secuenciador. De este modo, se consigue trasladar la experiencia completa de un proyecto de secuenciación a una sesión práctica.

### 4.1 Extracción de ADN y amplificación del gen ARNr 16S por PCR

En primer lugar, se realizará una extracción de ADN con un kit comercial que aísla eficazmente bacterias y hongos de todo tipo de suelos. El objetivo de este apartado es, siguiendo las instrucciones facilitadas, conseguir reproducir un protocolo de extracción de ADN (Imagen 3). Para ello, se recogerán muestras de suelo de lugares que se consideren interesantes a nivel biológico y en los que se encuentre diversidad microbiana. En este primer paso es obligatorio el uso de guantes para evitar la contaminación de las muestras con las bacterias propias y es importante recoger las muestras de suelo sin artefactos (restos vegetales, insectos, etc.), para que éstos no interfieran con el proceso de extracción de ADN.



Imagen 3. Diagrama de flujo del protocolo de extracción de ADN.

Una vez finalizado el protocolo, se comprobará la calidad y cantidad de ADN extraído mediante espectrofotómetro Nanodrop y gel de electroforesis de agarosa al 1% (50 mL TAE + 0,5 g agarosa). Por último, se preparará el “mix” de PCR cómo se detalla en la Tabla 1 y se programará el termociclador con las condiciones adecuadas para la amplificación del gen ARNr 16S (Imagen 4). Las temperaturas y tiempos de un programa de PCR dependen principalmente de la polimerasa utilizada, así como de la longitud del fragmento de ADN que se va a amplificar. Generalmente, a su temperatura óptima de actividad, la polimerasa es capaz de polimerizar mil bases en un minuto. En este caso, como el gen ARNr 16S tiene un tamaño aproximado de 1500 pb, 2 minutos de extensión será suficiente. Se utilizarán adaptadores o “códigos de barras” a modo de etiqueta para identificar cada una de las muestras de suelo de diferentes alumnos, ya que todas ellas se juntarán en una única muestra final que se cargará en la celda de flujo del dispositivo MinION. De este modo, durante el análisis bioinformático, cada una de las lecturas generadas se separará en función del “código de barras” que lleve asociado.

Reactivos	Volumen (µl)
Agua libre de nucleasas	14
DNA (10 ng totales)	10
16S “código de barras”	1
Taq polimerasa mix	25
TOTAL	50

Tabla 1. Reactivos y volumen necesarios para preparar el “mix” de amplificación por PCR.

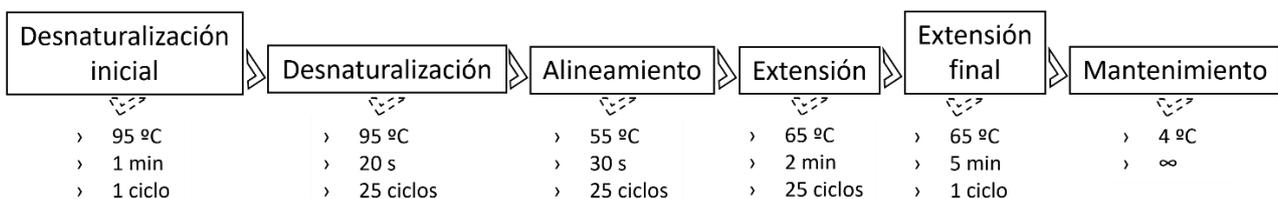


Imagen 4. Diagrama de flujo de un programa de PCR.

## 4.2 Preparación de la librería genómica y carga del MinION

En este apartado, los alumnos se familiarizarán con el secuenciador en tiempo real MinION de Oxford Nanopore Technologies. Aquí conocerán los diferentes puertos de una celda de flujo y comprenderán qué papel juegan en la carga de las muestras (Imagen 5). El objetivo principal es entender la mecánica del instrumento, así como comprender cómo se desarrolla el proceso de secuenciación. Las celdas de flujo están compuestas por cientos de

canales de nanoporos y pueden generar entre 10 y 20 Gb de datos de secuencia de ADN. MinION es capaz de detectar los cambios en la corriente eléctrica que produce cada uno de los nucleótidos al pasar por los canales. De este modo, basándose en la carga única que posee cada nucleótido, va generando lecturas a medida que las cadenas de ADN atraviesan los nanoporos.

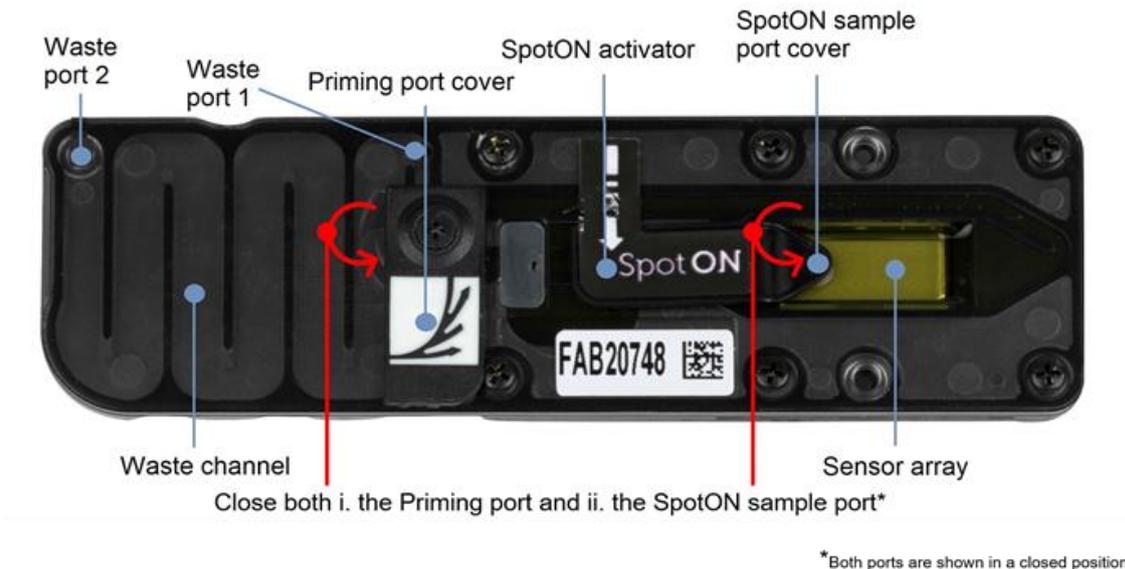


Imagen 5. Puertos de carga de una celda de flujo de un secuenciador MinION  
 (<https://www.nanoporetech.com>).

Previamente a la carga de la celda de flujo, se precipitará y purificará el producto de PCR mediante un protocolo muy sencillo en el que se emplean esferas magnéticas (Imagen 6). Este método está basado en la afinidad de las esferas magnéticas por los ácidos nucleicos en determinadas condiciones. Con la ayuda de un campo magnético de carga opuesta, las esferas quedarán pegadas a una de las paredes del tubo formando un pellet o conglomerado, permitiendo así el lavado de las muestras. Cambiando las condiciones, los ácidos nucleicos se liberan de las esferas magnéticas, logrando un rápido aislamiento y purificación de las muestras. Tras la purificación es importante volver a comprobar la concentración de las muestras mediante Nanodrop para diluir todas ellas a la misma concentración. La muestra final de carga debe tener una concentración equimolar para cada uno de los “códigos de barras”. Hay que tener en cuenta que para la preparación de las librerías genómicas existen requerimientos especiales según el secuenciador a utilizar. En el caso de MinION, se utiliza un kit comercial con reactivos estandarizados que se deben mezclar junto con el producto de PCR purificado siguiendo las instrucciones que se proporcionan.

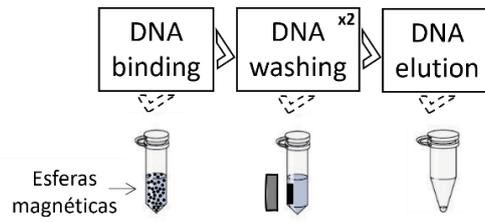


Imagen 6. Diagrama de flujo de la precipitación y purificación del producto de PCR mediante el uso de esferas magnéticas.

### 4.3 Análisis de los datos de secuenciación

Por último, cada estudiante analizará las secuencias generadas con su "código de barras". El objetivo de este apartado es dilucidar la composición de la muestra recogida. Para ello, se seguirá un flujo de trabajo integrado en el programa NanoGalaxy que incluye paso a paso desde el post-procesamiento de los datos brutos hasta la visualización gráfica de los resultados (Imagen 7).

El MinION almacena a tiempo real cada una de las lecturas generadas en carpetas separadas por el "código de barras". Para empezar con el análisis, se deben cargar los archivos en formato FASTQ, los cuales incluyen la calidad de la secuencia. A continuación, se debe identificar y eliminar los adaptadores, que son secuencias cortas que previamente hemos añadido a los extremos, comunes a todas las lecturas y que pueden interferir en el análisis. Para ello, se utilizará la herramienta "Porechop", específica para secuencias ONT. También se realizará un filtrado de las lecturas por calidad (Phred > 7) y longitud (1200-1800 pb, ya que el ARNr 16S tiene una longitud aproximada de unas 1500 pb). Una vez procesadas las lecturas, se utilizará el software "Kraken2" para clasificar las lecturas en los diferentes taxones microbianos. Este software realiza un estudio por similitud de las lecturas de entrada con un conjunto de secuencias recogidas en una base de datos. Por último, se utilizarán las herramientas "Convert Kraken" y "Krona pie chart" para visualizar gráficamente los resultados. Los resultados se muestran en un gráfico circular interactivo que subdivide los taxones en sectores separados en base a una jerarquía (como la abundancia taxonómica en un metagenoma). A partir de estos gráficos, se estudiará tanto la presencia como el porcentaje de taxones microbianos encontrados en la muestra.

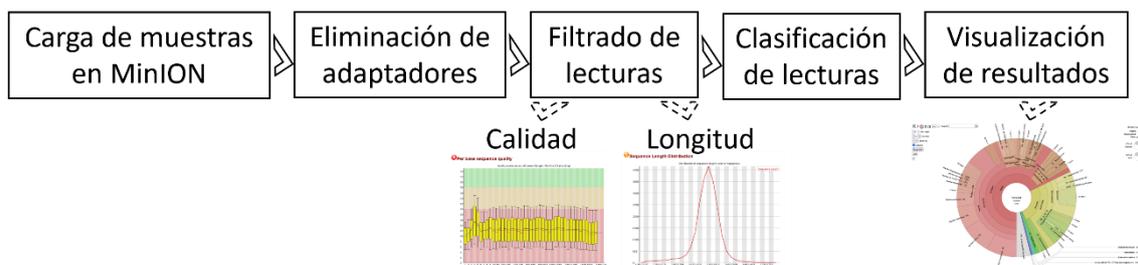


Imagen 7. Procedimiento de análisis de las lecturas obtenidas del secuenciador MinION para identificar la composición de la muestra.



## 5 Cierre

A lo largo de este objeto de aprendizaje el alumnado utilizará y asimilará procedimientos básicos de laboratorio como la extracción de ADN o la amplificación por PCR. También se familiarizarán con los fundamentos de bioinformática necesarios para el análisis masivo de datos. Además, habrán utilizado una tecnología de secuenciación de vanguardia, lo que será positivo para sus carreras profesionales. Debido al bajo coste y a la portabilidad de la tecnología MinION, esta sesión práctica podría generalizarse como práctica fundamental en asignaturas de genómica utilizando muestras de diferentes orígenes.

## 6 Bibliografía

de Koning, W.; Miladi, M.; Hiltmann, S.; Heikema, A.; Hays, J. P.; Flemming, S., et al.: "NanoGalaxy: Nanopore long-read sequencing data analysis in Galaxy", *GigaScience*, 2020, 9(10), g100105.

Hu, Z. L.; Ying, Y. L.; Huo, M. Z.; Kong, X. F.; Yu, X. D.; Zhang, J. R.; et al.: "A Course of Hands-On Nanopore Experiments for Undergraduates: Single-Molecule Detection with Portable Electrochemical Instruments", *Journal of Chemical Education*, 2020, 97(12), pág. 4345–4354.

Janda, J. M.; Abbott, S. L.: "16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls", *Journal of clinical microbiology*, 2007, 45(9), pág. 2761–2764.

Salazar, A. N.; Nobrega, F. L.; Anyansi, C.; Aparicio-Maldonado, C.; Costa, A. R.; Haagsma, A. C.; et al.: "An educational guide for nanopore sequencing in the classroom", *PLoS computational biology*, 2020, 16(1), e1007314.

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., et al.: "DNA sequencing at 40: past, present and future", *Nature*, 2017, 550(7676), pág. 345–353.

Zeng, Y.; Martin, C. H.: "Oxford Nanopore sequencing in a research-based undergraduate course", *BioRxiv*, 2017, 227439.