

Flujo de trabajo de metagenómica indiscriminada mediante NGS para evaluar poblaciones microbianas en muestras complejas

Los kits de 600 ciclos de NextSeq™ 1000 System y NextSeq 2000 System proporcionan precisión y flexibilidad para la identificación de especies



Clasificación metagenómica de muestras complejas

La secuenciación metagenómica indiscriminada es un método alternativo a las estrategias de secuenciación de amplicones, como la secuenciación del ARN ribosómico (ARNr) de la subunidad 16S y el espaciador transcrito interno (ITS, internal transcribed spacer), para evaluar la diversidad microbiana en muestras complejas. A diferencia de las estrategias basadas en amplicones, la metagenómica indiscriminada mediante secuenciación de nueva generación (NGS, next-generation sequencing) captura información genómica completa para cada microorganismo presente en una muestra. La capacidad de capturar genomas completos significa que la metagenómica indiscriminada puede identificar especies que no se detectan mediante la secuenciación de amplicones¹ y que los datos obtenidos contienen información funcional no disponible a partir de los métodos basados en amplicones.^{2,3}

Esta nota de aplicación demuestra las similitudes de rendimiento de NextSeq 1000 System, NextSeq 2000 System y MiSeq™ System para estudios mediante metagenómica indiscriminada de alta productividad. Mediante el uso de los datos generados en el destacado NextSeq 550 System, también demostramos las ventajas que tienen los kits de 600 ciclos sobre los kits de 300 ciclos utilizados habitualmente en aplicaciones de metagenómica. Presentamos datos de poblaciones sintéticas y muestras de la vida real para demostrar una identificación superior de géneros y especies cuando se utilizan kits de 600 ciclos.

Métodos

El kit NextSeq 1000/2000 P1 Reagents (600 cycles) y el kit NextSeq 1000/2000 P2 Reagents (600 cycles) amplían la capacidad y el rendimiento de secuenciación de NextSeq 1000 System y NextSeq 2000 System, con especificaciones ideales para la secuenciación metagenómica indiscriminada. NextSeq 1000 System y NextSeq 2000 System utilizan reactivos con tecnología “load-and-go” sin fluídica integrada, lo que reduce el número de pasos del flujo de trabajo y el riesgo de contaminación de las muestras. El flujo de trabajo de metagenómica indiscriminada integra la preparación de librerías, la NGS probada de Illumina y el análisis de datos secundario con solo pulsar un botón para una solución completa para el descubrimiento del microbioma (figura 1).

Preparación de librerías

Se obtuvieron muestras de ADN genómico microbiano de dos fuentes. La primera muestra fue la muestra comercial ATCC 20 Strain Staggered Mix Genomic Material (n.º de catálogo MSA-1003 de ATCC). Esta muestra de la ATCC es una comunidad microbiana simulada compuesta por una distribución escalonada de ADN genómico preparado a partir de cepas bacterianas seleccionadas en función de atributos como la tinción de Gram, el contenido de GC y atributos de esporulación. También se obtuvo un segundo conjunto de muestras de heces de la vida real, descrito anteriormente⁴, para su análisis.

Las librerías se prepararon con Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (Illumina, n.º de catálogo 20060060) e IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)

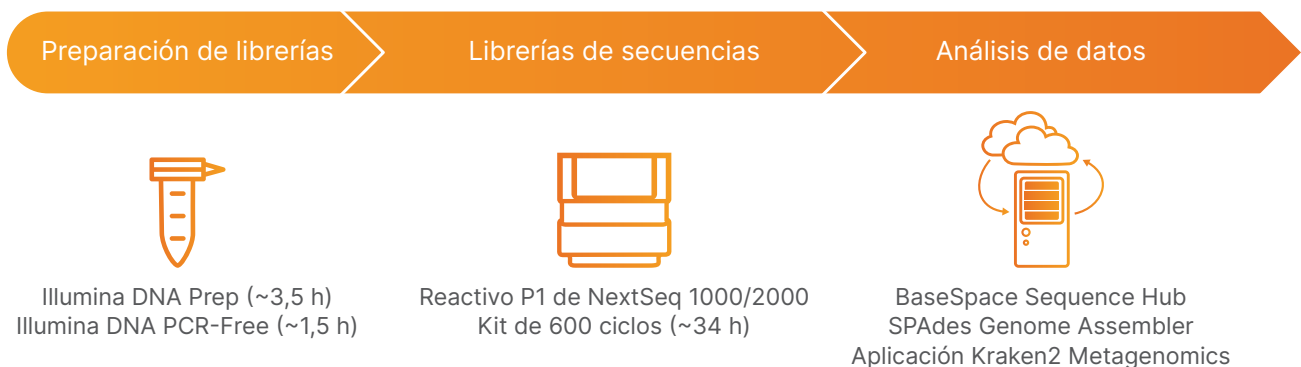


Figura 1: Flujo de trabajo de NGS metagenómica indiscriminada en NextSeq 1000 System y NextSeq 2000 System

(Illumina, n.º de catálogo 20027213). IDT for Illumina DNA/RNA UD Indexes Sets A a D permite a los usuarios generar 384 librerías de 16S.

Secuenciación

Las librerías preparadas se agruparon y cargaron en una celda de flujo precargada del kit NextSeq 1000/2000 P1 Reagents (600 cycles), una celda de flujo de MiSeq Reagent Kit v3 (600 cycles) (Illumina, n.º de catálogo MS-102-3003) o una celda de flujo NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, n.º de catálogo 20024908). La secuenciación se realizó en NextSeq 2000 System, MiSeq System o NextSeq 550 System, respectivamente. Los datos de secuenciación representativos de todos los experimentos están disponibles en la página web de [datos de demostración de BaseSpace™ Sequence Hub](#).

Análisis

Las librerías agrupadas se demultiplexaron en la plataforma de computación de la nube genómica BaseSpace Sequence Hub. Para procesar los datos generados en NextSeq 2000 System, MiSeq System y NextSeq 550 System se utilizó el proceso de metagenómica de DRAGEN™. Los metagenomas se ensamblaron con SPAdes Genome Assembler. Las clasificaciones taxonómicas se llevaron a cabo mediante el proceso DRAGEN Metagenomics.

Para comparar los datos de NextSeq 2000 generados con un kit de 600 ciclos con los datos de NextSeq 500 generados con un kit de 300 ciclos, las lecturas de NextSeq 2000 se recortaron con DRAGEN FASTQ Toolkit, disponible en BaseSpace Sequence Hub.

Para permitir comparaciones entre muestras, cada muestra se redujo al mismo número de lecturas (30 M, 10 M, 1 M) mediante DRAGEN FASTQ Toolkit. Esta reducción es necesaria en los casos en los que solo se puede procesar un subconjunto de la muestra mediante una aplicación (p. ej., un conjunto *de novo* con restricciones de memoria) o cuando no es necesario el conjunto de datos completo para procesar una muestra (p. ej., para validar una estrategia con diferentes niveles de cobertura genómica).

Resultados

Mejora de los parámetros principales de la secuencia

NextSeq 1000/2000 P1 Reagents (600 cycles) en NextSeq 2000 System muestran un mayor porcentaje de puntuaciones de calidad $\geq Q30$ en comparación con MiSeq Reagent Kit v3 (600 cycles) ejecutado en MiSeq System. Además, NextSeq 1000/2000 P1 Reagents (600 cycles) proporcionan hasta 100 M de lecturas individuales que pasan el filtro o 200 M de lecturas “paired-end” que pasan el filtro. A aproximadamente 60 Gb, NextSeq 1000/2000 P1 Reagents (600 cycles) generan cuatro veces más datos que MiSeq Reagent Kit v3 (600 cycles) a aproximadamente 15 Gb. Asimismo, los experimentos de secuenciación con NextSeq 1000/2000 P1 Reagents (600 cycles) se completaron en aproximadamente 34 h, lo que supone aproximadamente 20 h menos que un experimento de secuenciación con MiSeq Reagent Kit v3 (600 cycles) (figura 2).

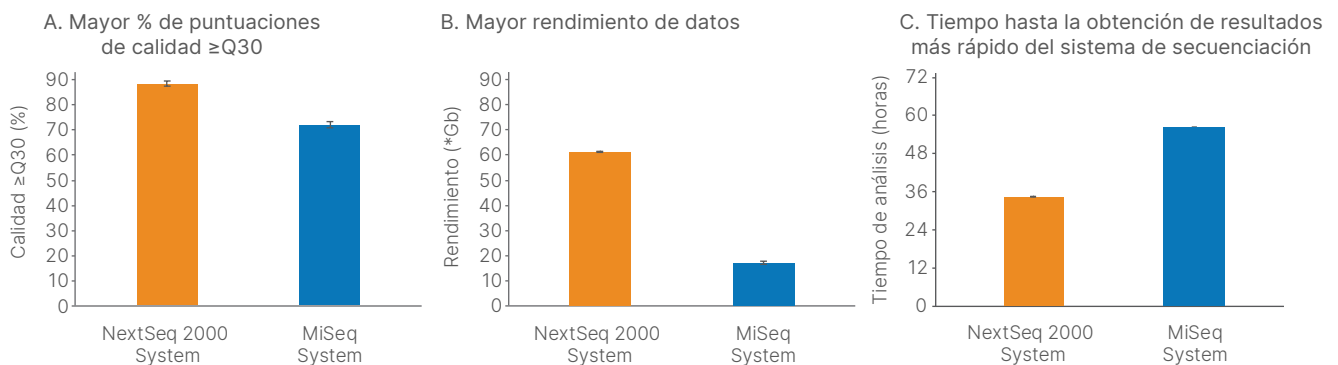


Figura 2: Comparaciones de los parámetros de rendimiento principales para NextSeq 2000 System y MiSeq System. En comparación con MiSeq Reagent Kit v3 (600 ciclos) ejecutado en MiSeq System, NextSeq 1000/2000 P1 Reagents (600 ciclos) en NextSeq 2000 System ofrecen (A) un mayor porcentaje de puntuaciones de calidad $\geq Q30$, (B) cuatro veces más rendimiento de datos y (C) aproximadamente 20 h menos de duración del experimento en el instrumento cuando se utiliza la celda de flujo NextSeq P1.

NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, n.º de catálogo 20024908) puede generar hasta 120 Gb de datos de secuenciación de alta calidad, con una longitud de lectura de 2 × 150 pb, en unas 29 horas. Las especificaciones de calidad de este kit son >75 % de las bases con ≥Q30 (datos no mostrados).

Comparaciones del análisis metagenómico

Para comparar el rendimiento entre sistemas, se secuenció la muestra de 20 Strain Staggered Mix Genomic Material en NextSeq 2000 System, MiSeq System y NextSeq 550 System. Para análisis sucesivos que dilucidan las clasificaciones taxonómicas se utilizó la aplicación DRAGEN Metagenomics en BaseSpace Sequence Hub. El análisis metagenómico de NGS identificó todos los miembros esperados de la comunidad bacteriana simulada y mostró resultados comparables entre los tres sistemas de secuenciación (figura 3).

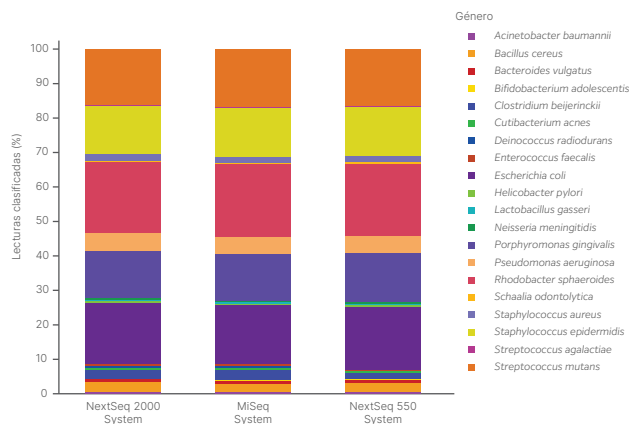


Figura 3: el análisis comparativo de la composición microbiana revela la composición de ATCC 20 Strain Staggered Mix analizada en NextSeq 2000 System, MiSeq System y NextSeq 550 System. El análisis de la composición microbiana mediante la aplicación DRAGEN Metagenomics demuestra una identidad y distribución de géneros excelentes y reproducibles.

Aunque la muestra de la ATCC revela un rendimiento altamente repetible, una muestra simulada puede no ser indicativa del rendimiento con muestras del mundo real. Por lo tanto, se analizó el rendimiento de la detección de microorganismos con muestras de heces del mundo real. Los 20 géneros más representados en las muestras de heces se compararon en NextSeq 2000 System, NextSeq 550 System y MiSeq System (figura 4). Aunque los resultados no son tan uniformes con este tipo de muestra más compleja, los perfiles metagenómicos comunitarios de todas las muestras de heces reales

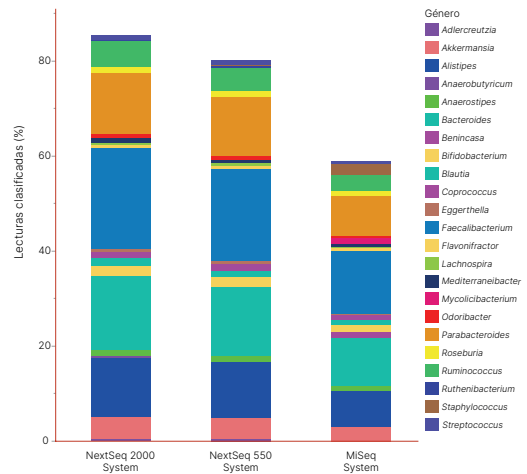


Figura 4: el análisis metagenómico revela una composición comunitaria similar para una muestra de heces representativa de la vida real analizada en múltiples sistemas. Aplicación DRAGEN Metagenomics para analizar la composición microbiana de las muestras de heces de la vida real, limitada a los 20 géneros más representados en la muestra, secuenciados en NextSeq 2000 System, NextSeq 550 System y MiSeq System. Los datos demuestran una cobertura de géneros similar en todas las plataformas, incluso con muestras complejas.

coincidieron en gran medida entre NextSeq 2000 System, MiSeq System y NextSeq 550 System. Esto indica que los tres sistemas son capaces de realizar una secuenciación metagenómica fiable en diversos tipos de muestras.

Una mayor longitud de lectura mejora la caracterización de las muestras

Aunque las celdas de flujo de 300 ciclos pueden proporcionar datos de secuenciación metagenómica significativos, los kits de 600 ciclos ofrecen ventajas en lo que se refiere a la identificación de especies individuales en muestras complejas. Para demostrar el efecto de lecturas más largas en el ensamblaje del metagenoma, las lecturas del sistema NextSeq 2000 se recortaron de 600 ciclos a 300 ciclos utilizando la aplicación DRAGEN FASTQ Toolkit. Se utilizó Kraken2, un clasificador taxonómico basado en k-mer,* para determinar el porcentaje de lecturas clasificadas para cada muestra con 600 ciclos o 300 ciclos (figura 5). Los datos apuntan a que las lecturas más largas proporcionan algunas mejoras para la clasificación taxonómica basada en k-mer con diversas muestras medioambientales.

* La clasificación taxonómica de Kraken2 Metagenomics está disponible a través de la aplicación de BaseSpace DRAGEN Metagenomics de Illumina.

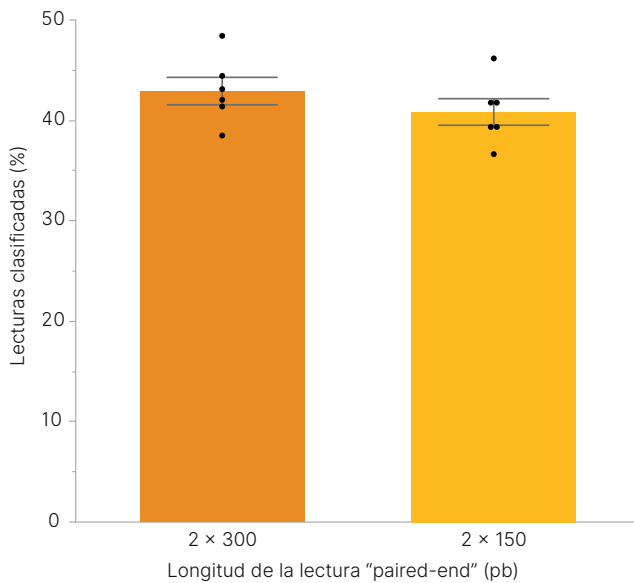


Figura 5: una mayor longitud de lectura mejora la clasificación de las lecturas de muestras de heces de la vida real secuenciadas en NextSeq 2000. Se utilizó la aplicación DRAGEN FASTQ Toolkit para recortar las lecturas obtenidas en NextSeq 2000 System de 600 ciclos (2 x 300 pb) a 300 ciclos (2 x 150 pb). A continuación, se utilizó Kraken2 para determinar el porcentaje de lecturas clasificadas para cada muestra. La profundidad de lectura fue de 30 M de lecturas para el análisis. Las barras de error representan un error estándar de la media.

Utilizando los mismos datos recortados, también se examinaron la longitud de lectura que afecta a la riqueza de la muestra (es decir, el número de especies detectadas en una muestra) y el índice de Shannon (es decir, la representación proporcional de las especies detectadas en la muestra). Estos parámetros indican que la diversidad microbiana observada de las muestras de heces aumenta cuanto mayor es la longitud de lectura, mientras que la proporción de especies detectadas cuantificadas por el índice de Shannon permanece relativamente inalterada, como cabía esperar (figura 6).

Una mayor profundidad de secuenciación mejora la caracterización de las muestras

A continuación, se analizó la importancia de la profundidad de lectura en las medidas de diversidad de la población para las muestras de heces de la vida real. Se redujo el número de lecturas de NextSeq 2000 System a 30 M, 10 M y 1 M de lecturas con DRAGEN FASTQ Toolkit, y se calcularon la riqueza y el índice de Shannon con la aplicación DRAGEN Metagenomics.

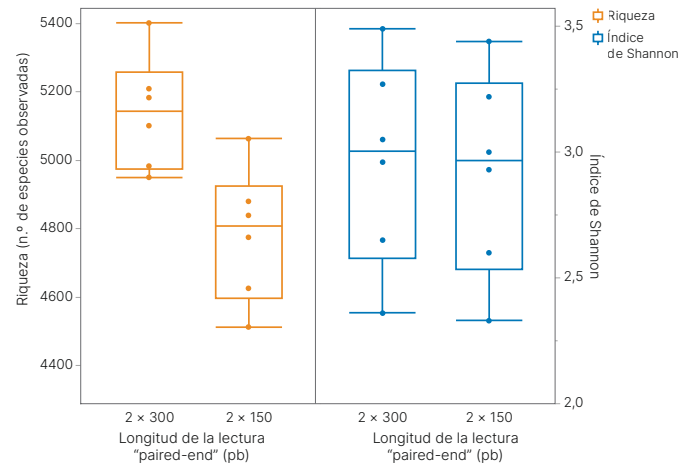


Figura 6: la riqueza microbiana de las muestras de heces aumenta con una mayor longitud de lectura. Se utilizó la aplicación DRAGEN FASTQ Toolkit para recortar las lecturas obtenidas en NextSeq 2000 System de 600 ciclos (2 x 300 pb) a 300 ciclos (2 x 150 pb). A continuación, se utilizó la aplicación DRAGEN Metagenomics para calcular la riqueza y el índice de Shannon para cuantificar el número de especies detectadas y la diversidad proporcional, respectivamente. La riqueza aumenta con lecturas más largas, mientras que el índice de Shannon permanece relativamente sin cambios. La profundidad de lectura fue de 30 M de lecturas para el análisis.

Los parámetros de diversidad demostraron que la diversidad microbiana de las muestras de heces aumenta cuanto mayor es la profundidad de secuenciación, mientras que la proporción de especies indicadas por el índice de Shannon permanece relativamente sin cambios (figura 7).

Las lecturas más largas mejoran la identificación taxonómica

Uno de los desafíos actuales con la creación de perfiles de diversas poblaciones microbianas medioambientales es la falta de genomas de referencia completos para muchas especies raras y no cultivables. Por lo general, el número de contigs ensamblados para poblaciones microbianas muy diversas es mayor para longitudes de lectura más largas, como se muestra en la longitud total de conjunto más grande (figura 8). La secuenciación metagenómica indiscriminada con longitudes de lectura de secuenciación de 2 x 301 pb de Illumina generalmente mejora el ensamblaje *de novo* de metagenomas a partir de muestras ambientales, lo que contribuye significativamente a la integridad general de cada metagenoma ensamblado.

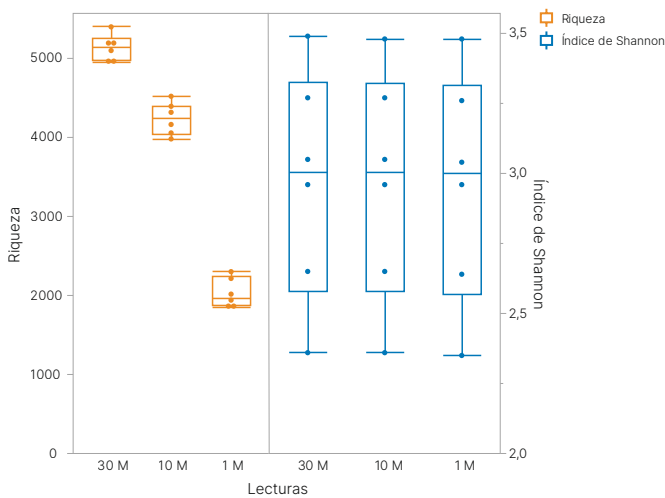


Figura 7: la riqueza microbiana de las muestras de heces aumenta a medida que aumenta la profundidad de la secuenciación. Se calcularon la riqueza y el índice de Shannon utilizando la aplicación DRAGEN Metagenomics para medir el número de especies detectadas y la diversidad proporcional observada con la secuenciación a 30 M, 10 M y 1 M de lecturas. Como cabía esperar, la riqueza disminuye en los datos de muestras reducidas, mientras que el índice de Shannon permanece aproximadamente igual.

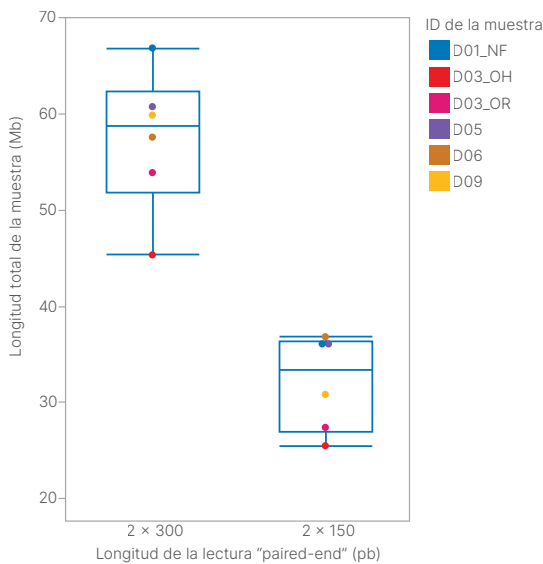


Figura 8: una mayor longitud de lectura admite un mayor número total de bases ensambladas para muestras de heces de la vida real secuenciadas en NextSeq 2000 System. Los datos de secuenciación metagenómica en NextSeq 2000 System se recortaron a 1 M de lecturas y a longitudes de lectura de 2 x 300 pb a 2 x 150 pb utilizando la aplicación DRAGEN FASTQ Toolkit. Para las comparaciones, se utilizó SPAdes Genome Assembler para generar longitudes de contigios totales a partir de lecturas de secuenciación establecidas en 2 x 300 pb y 2 x 150 pb.

Resumen

El objetivo de esta nota de aplicación es demostrar el rendimiento similar de los kits de 600 ciclos en NextSeq 2000 System y MiSeq System. NextSeq 550 System también es capaz de realizar análisis metagenómicos precisos, pero no cuenta con una opción de 600 ciclos. Los tres sistemas lograron ensamblajes *de novo* precisos independientemente del instrumento que se utilice. En los tres sistemas se lograron perfiles metagenómicos concordantes, especialmente en la diversidad proporcional observada de géneros.

En general, NextSeq 2000 System ofrece ventajas con respecto a MiSeq System y NextSeq 550 System en términos de resolución de detalles de riqueza de muestras a partir de diversas muestras sin cultivo gracias a capacidades adicionales de longitud de lectura y profundidad de lectura.[†] Estos avances son especialmente importantes para muestras metagenómicas complejas, como las heces o las muestras medioambientales.

En resumen, NextSeq 1000/2000 P1 Reagents (600 cycles) y NextSeq 1000/2000 P2 Reagents (600 cycles) en NextSeq 1000 System y NextSeq 2000 System ofrecen secuenciación de alta calidad con un tiempo de respuesta más rápido que MiSeq Reagent Kit v3 (600 ciclos) en MiSeq System. Los kits de 600 ciclos en NextSeq 1000 System y NextSeq 2000 System mantienen puntuaciones Q30 altas y tasas de error bajas, como se demuestra en esta nota de aplicación. Además, NextSeq 1000/2000 P1 (600 cycles) y NextSeq 1000/2000 P2 Reagents (600 cycles) proporcionan un rendimiento de datos flexible que es ideal para experimentos de secuenciación del genoma a pequeña y media escala. Por último, los kits de 600 ciclos para NextSeq 1000 System y NextSeq 2000 System permiten la expansión de aplicaciones y la simplicidad operativa, a la vez que mantienen la calidad de datos establecida en el reconocido MiSeq System.

[†] NextSeq 1000 System es funcionalmente similar a NextSeq 2000 System y ambos deben alcanzar resultados similares utilizando los mismos kits de 600 ciclos.

Información adicional

Plataformas de secuenciación de Illumina

NextSeq 1000 System y NextSeq 2000 System

Reactivos de NextSeq 1000/2000

Illumina DNA Prep

Protocolo de lisado en crudo de Illumina DNA Prep para NGS

Bibliografía

1. Peterson D, Bonham KS, Rowland S, Pattanayak CW; RESONANCE Consortium, Klepac-Ceraj V. [Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes](#). *Front Microbiol.* 2021;12:670336. Publicado el 15 de julio de 2021. doi:10.3389/fmicb.2021.670336
2. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. [Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota](#). *Sci Rep.* 2021;11(1):3030. Publicado el 4 de febrero de 2021. doi:10.1038/s41598-021-82726-y
3. Stothart MR, McLoughlin PD, Poissant J. [Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights](#). *Mol Ecol Resour.* 2023;23(3):549-564. doi:10.1111/1755-0998.13713
4. Illumina. 16S rRNA sequencing on NextSeq 1000 and NextSeq 2000 Systems. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-600c-16s-rrna-application-note-m-gl-01146/nextseq-600c-16s-rrna-application-note-m-gl-01146.pdf. Año de publicación: 2023. Fecha de consulta: 6 de febrero de 2024.



1 800 809 4566 (llamada gratuita, EE. UU.) | tel.: +1 858 202 4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, consulte www.illumina.com/company/legal.html.
M-GL-01147 ESP v1.0