

Flusso di lavoro di metagenomica shotgun NGS per la valutazione delle popolazioni microbiche in campioni complessi

I kit da 600 cicli su
NextSeq™ 1000 System
e NextSeq 2000 System
offrono accuratezza e
flessibilità per l'identificazione
delle specie

illumina®

Classificazione metagenomica di campioni complessi

Il sequenziamento metagenomico shotgun è un metodo alternativo agli approcci di sequenziamento di ampliconi, come il sequenziamento dell'RNA ribosomiale (rRNA, ribosomal RNA) 16S e dello spaziatore trascritto interno (ITS, internal transcribed spacer), per la valutazione della diversità microbica in campioni complessi. A differenza degli approcci basati sugli ampliconi, il sequenziamento metagenomico shotgun che impiega il sequenziamento di nuova generazione (NGS, next-generation sequencing) acquisisce informazioni genomiche complete per ogni organismo presente in un campione. La capacità di catturare genomi completi fa sì che la metagenomica shotgun possa identificare le specie non rilevate dal sequenziamento di ampliconi¹ e che i dati risultanti contengano informazioni funzionali non disponibili con i metodi basati sugli ampliconi.^{2,3}

Questa nota applicativa illustra le somiglianze a livello di prestazioni di NextSeq 1000 System, NextSeq 2000 System e MiSeq™ System per gli studi metagenomici shotgun a elevata produttività. Utilizzando i dati generati dal ben collaudato NextSeq 550 System, dimostriamo anche i vantaggi dei kit da 600 cicli rispetto ai kit da 300 cicli comunemente utilizzati nelle applicazioni di metagenomica. Presentiamo dati basati sulla popolazione sintetica e su campioni reali per dimostrare le migliori prestazioni di identificazione di generi e specie con l'utilizzo di kit da 600 cicli.

Metodi

NextSeq 1000/2000 P1 Reagents (600 cycles) e NextSeq 1000/2000 P2 Reagents (600 cycles) ampliano la capacità e l'output del sequenziamento del NextSeq 1000 System e del NextSeq 2000 System con specifiche ideali per il sequenziamento metagenomico shotgun. Il NextSeq 1000 System e il NextSeq 2000 System utilizzano reagenti "carica e vai" senza fluidica integrata, riducendo il numero di fasi del flusso di lavoro e il rischio di contaminazione del campione. Il flusso di lavoro di metagenomica shotgun integra la preparazione delle librerie, il comprovato NGS Illumina e l'analisi dei dati secondaria con pulsante, per una soluzione completa di studio del microbioma (Figura 1).

Preparazione delle librerie

I campioni di DNA genomico microbico sono stati ottenuti da due fonti. Il primo campione consisteva nella raccolta di colture disponibile in commercio American Type ATCC 20 Strain Staggered Mix Genomic Material (ATCC, n. di catalogo MSA-1003). Questo campione ATCC è una comunità microbica simulata composta da una distribuzione scaglionata di DNA genomico preparato a partire da ceppi batterici selezionati in base ad attributi quali colorazione di Gram, contenuto di GC e attributi di sporulazione. È stata inoltre raccolta per l'analisi una seconda serie di campioni reali di feci, già descritti in precedenza⁴.

Le librerie sono state preparate utilizzando Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (Illumina, n. di catalogo 20060060) e IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples) (Illumina, n. di catalogo 20027213). IDT for Illumina DNA/RNA UD Indexes Sets A to D consente agli utenti di generare 384 librerie 16S.



Figura 1: flusso di lavoro di metagenomica shotgun NGS su NextSeq 1000 System e NextSeq 2000 System

Sequenziamento

Le librerie preparate sono state raggruppate in pool e caricate in una cella a flusso preriempita del kit NextSeq 1000/2000 P1 Reagents (600 cycles), una cella a flusso del MiSeq Reagent Kit v3 (600-cycle) (Illumina, n. di catalogo MS-102-3003) o una cella a flusso del NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, n. di catalogo 20024908). Il sequenziamento è stato eseguito rispettivamente sul NextSeq 2000 System, su un MiSeq System o su un NextSeq 550 System. I dati di sequenziamento rappresentativi di tutte le corse sono disponibili nella pagina Web dei [dati dimostrativi di BaseSpace™ Sequence Hub](#).

Analisi

Si è quindi provveduto al demultiplexing delle librerie raggruppate in pool sulla piattaforma di cloud computing genomico BaseSpace Sequence Hub. Per elaborare i dati generati su NextSeq 2000 System, MiSeq System e NextSeq 550 System, è stata utilizzata la pipeline DRAGEN™ Metagenomics. I metagenomi sono stati assemblati utilizzando lo SPAdes Genome Assembler. Per le classificazioni tassonomiche è stata impiegata la pipeline DRAGEN Metagenomics.

Per confrontare i dati del NextSeq 2000 generati utilizzando un kit da 600 cicli con i dati del NextSeq 500 generati utilizzando un kit da 300 cicli, le letture del NextSeq 2000 sono state tagliate impiegando il DRAGEN FASTQ Toolkit, disponibile su BaseSpace Sequence Hub. Per consentire i confronti tra i campioni,

ciascun campione è stato sottocampionato allo stesso numero di letture (30 milioni, 10 milioni, 1 milione) con il DRAGEN FASTQ Toolkit. Il sottocampionamento è richiesto nei casi in cui solo un sottoinsieme del campione può essere elaborato da un'applicazione (ad es., assemblaggio *de novo* con vincoli di memoria) o quando per elaborare un campione non è necessario l'intero set di dati (ad es., per convalidare un approccio a livelli variabili di copertura genomica).

Risultati

Miglioramento dei parametri primari della sequenza

NextSeq 1000/2000 P1 Reagents (600 cycles) sul NextSeq 2000 System mostra una percentuale maggiore di punteggi qualitativi superiori o uguali a Q30 rispetto alla corsa con il MiSeq Reagent Kit v3 (600-cycle) sul MiSeq System. NextSeq 1000/2000 P1 Reagents (600 cycles) fornisce inoltre fino a 100 milioni di letture single-end che superano i filtri o 200 milioni di letture paired-end che superano i filtri. A circa 60 Gb, NextSeq 1000/2000 P1 Reagents (600 cycles) genera quattro volte più dati rispetto al MiSeq Reagent Kit v3 (600-cycle) a circa 15 Gb. Inoltre, le corse di sequenziamento con NextSeq 1000/2000 P1 Reagents (600 Cycles) Kit vengono completate in circa 34 ore, ovvero approssimativamente 20 ore in meno rispetto a una corsa di sequenziamento con il MiSeq Reagent Kit v3 (600-cycle) (Figura 2).

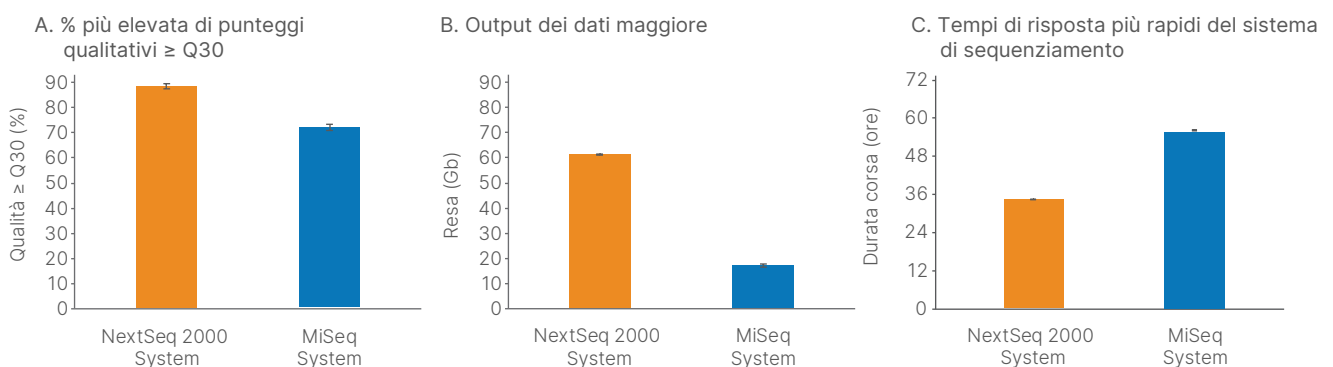


Figura 2: principali parametri prestazionali di NextSeq 2000 System e MiSeq System a confronto. Rispetto al MiSeq Reagent Kit v3 (600-cycle) eseguito sul MiSeq System, il NextSeq 1000/2000 P1 Reagents (600 cycles) sul NextSeq 2000 System offre (A) una percentuale più elevata di punteggi qualitativi maggiori o uguali a Q30, (B) un output di dati quattro volte superiore a (C) durata della corsa dello strumento più breve di circa 20 ore quando si utilizza la cella a flusso NextSeq P1.

Il NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, n. di catalogo 20024908) è in grado di generare fino a 120 Gb di dati di sequenziamento di alta qualità, con una lunghezza della lettura di 2 × 150 bp, in circa 29 ore. La specificità di qualità per questo kit è superiore al 75% delle basi con un punteggio maggiore o uguale a Q30 (dati non mostrati).

Analisi metagenomiche a confronto

Per confrontare le prestazioni dei diversi sistemi, il 20 Strain Staggered Mix Genomic Material è stato sequenziato sul NextSeq 2000 System, sul MiSeq System e sul NextSeq 550 System. Per l'analisi a valle finalizzata al chiarimento delle classificazioni tassonomiche, è stata utilizzata l'app DRAGEN Metagenomics su BaseSpace Sequence Hub. L'analisi metagenomica NGS ha identificato tutti i membri previsti della comunità batterica simulata e ha mostrato risultati comparabili su tutti e tre i sistemi di sequenziamento (Figura 3).

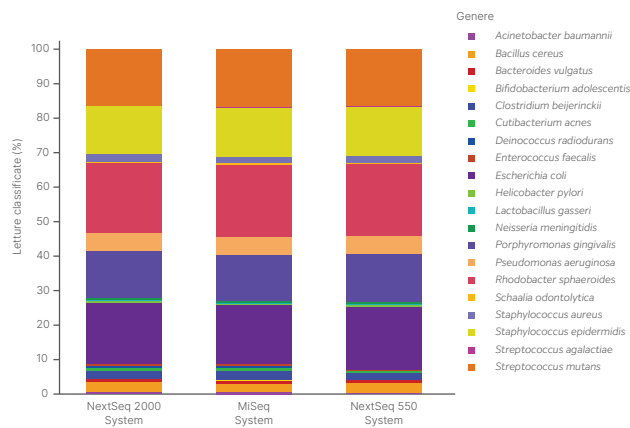


Figura 3: l'analisi comparativa della composizione microbica rivela la composizione del campione ATCC 20 Strain Staggered Mix analizzato sui tre sistemi, NextSeq 2000 System, MiSeq System e NextSeq 550 System. L'analisi della composizione microbica mediante l'app DRAGEN Metagenomics dimostra un'identità e una distribuzione eccellenti e riproducibili dei generi.

Sebbene il campione ATCC mostri prestazioni altamente ripetibili, un campione simulato potrebbe non essere indicativo delle prestazioni con campioni reali. Pertanto, abbiamo testato le prestazioni di rilevamento degli organismi con campioni fecali reali. Per i 20 generi più rappresentati nei campioni fecali, è stato effettuato un confronto fra i tre sistemi, NextSeq 2000 System, NextSeq 550 System e MiSeq System (Figura 4). Sebbene i risultati non siano altrettanto uniformi con questo tipo di campione più complesso, i profili di metagenomica della comunità di tutti i campioni

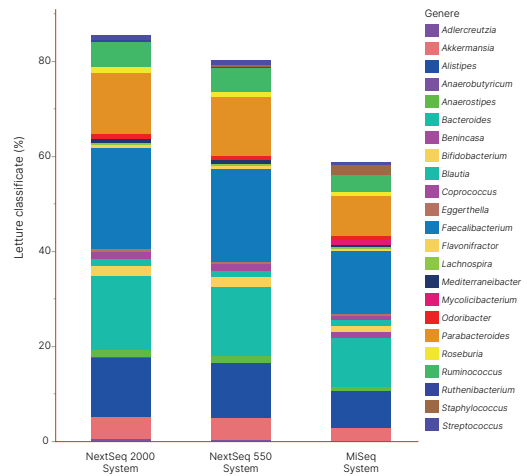


Figura 4: l'analisi metagenomica rivela una composizione comunitaria simile per un campione fecale reale rappresentativo analizzato su più sistemi. Con l'app DRAGEN Metagenomics è stata analizzata la composizione microbica di campioni fecali reali, limitatamente ai 20 generi maggiormente rappresentati nel campione, sequenziati su NextSeq 2000 System, NextSeq 550 System e MiSeq System. I dati dimostrano una copertura dei generi simile su tutte le piattaforme, anche con campioni complessi.

fecali reali sono risultati altamente concordanti con i tre sistemi, NextSeq 2000 System, MiSeq System e NextSeq 550 System. Ciò indica che tutti e tre i sistemi sono in grado di eseguire un sequenziamento metagenomico affidabile in diversi tipi di campioni.

Una maggiore lunghezza della lettura migliora la caratterizzazione del campione

Sebbene le celle a flusso da 300 cicli siano in grado di fornire dati di sequenziamento metagenomico significativi, i kit da 600 cicli offrono vantaggi quando si lavora per identificare singole specie in campioni complessi. Per dimostrare l'effetto di letture più lunghe sull'assemblaggio di metagenomi, le letture del NextSeq 2000 System sono state ridotte da 600 cicli a 300 cicli utilizzando l'app DRAGEN FASTQ Toolkit. Per determinare la percentuale di letture classificate per ciascun campione a 600 cicli o 300 cicli, è stato utilizzato Kraken2, un classificatore tassonomico basato su k-mer* (Figura 5). I dati suggeriscono che letture più lunghe migliorano in certa misura la classificazione tassonomica basata su k-mer con diversi campioni ambientali.

* La classificazione tassonomica di Kraken2 Metagenomics è disponibile tramite l'app Illumina DRAGEN Metagenomics BaseSpace.

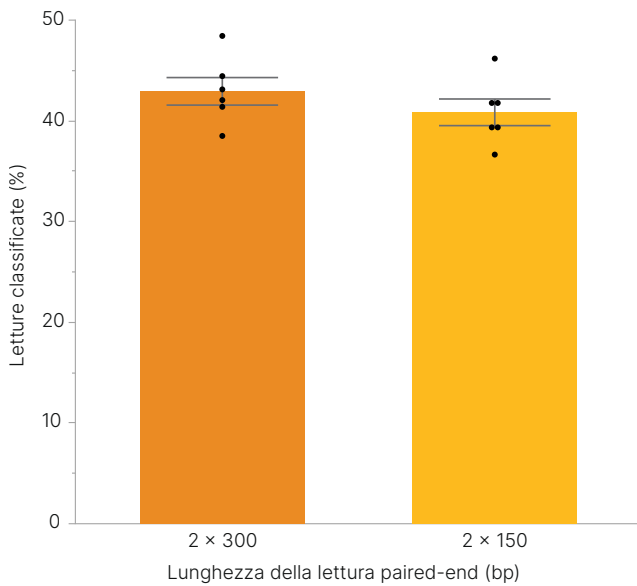


Figura 5: la maggiore lunghezza delle letture migliora la classificazione delle letture relative ai campioni fecali reali sequenziati sul NextSeq 2000. È stata utilizzata l'app DRAGEN FASTQ Toolkit per tagliare le letture eseguite dal NextSeq 2000 System da 600 cicli (2 x 300 bp) a 300 cicli (2 x 150 bp). Successivamente, è stato utilizzato Kraken2 per determinare la percentuale di letture classificate per ciascun campione. La profondità di lettura relativa all'analisi è stata di 30 milioni di letture. Le barre di errore rappresentano una deviazione standard rispetto alla media.

Utilizzando gli stessi dati tagliati, sono stati esaminati anche gli impatti della lunghezza della lettura sulla ricchezza del campione (ovvero, il numero di specie rilevate in un campione) e sull'indice di Shannon (ovvero, la rappresentazione proporzionale delle specie rilevate nel campione). Questi parametri indicano che la diversità microbica osservata dei campioni di feci aumenta di pari passo con la lunghezza della lettura, mentre la proporzione delle specie rilevate quantificate dall'indice di Shannon, come previsto, rimane relativamente invariata (Figura 6).

Una maggiore profondità di sequenziamento migliora la caratterizzazione del campione

Successivamente, abbiamo esaminato l'importanza della profondità di lettura sulle misure della diversità della popolazione per i campioni fecali reali. Il numero di letture del NextSeq 2000 System è stato sottocampionato a 30 milioni, 10 milioni e 1 milione di letture utilizzando il DRAGEN FASTQ Toolkit e la ricchezza e l'indice di Shannon sono stati calcolati con l'app DRAGEN Metagenomics.

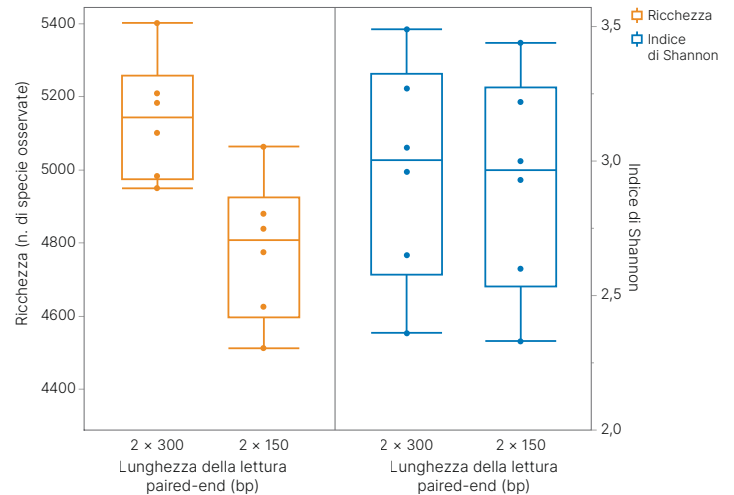


Figura 6: la ricchezza microbica dei campioni di feci aumenta di pari passo con la lunghezza della lettura. È stata utilizzata l'app DRAGEN FASTQ Toolkit per tagliare le letture eseguite dal NextSeq 2000 System da 600 cicli (2 x 300 bp) a 300 cicli (2 x 150 bp). Successivamente, è stata utilizzata l'app DRAGEN Metagenomics per calcolare la ricchezza e l'indice di Shannon al fine di quantificare, rispettivamente, il numero di specie rilevate e la diversità proporzionale. La ricchezza aumenta con letture più lunghe, mentre l'indice di Shannon rimane relativamente invariato. La profondità di lettura relativa all'analisi è stata di 30 milioni di letture.

I parametri di diversità hanno dimostrato che la diversità microbica dei campioni di feci aumenta di pari passo con la profondità di sequenziamento, mentre la proporzione delle specie indicata dall'indice di Shannon rimane relativamente invariata (Figura 7).

Letture più lunghe migliorano l'identificazione tassonomica

Una delle attuali complessità legate alla profilazione di popolazioni microbiche ambientali diversificate è la mancanza di genomi di riferimento completi per molte specie rare e non colture. In generale, il numero di contig assemblati per popolazioni microbiche altamente diversificate risulta maggiore con lunghezze di lettura superiori, come mostrato nella lunghezza totale di assemblaggio maggiore (Figura 8). Il sequenziamento metagenomico shotgun con le lunghezze di lettura del sequenziamento Illumina pari a 2 x 301 bp migliora generalmente l'assemblaggio *de novo* dei metagenomi dei campioni ambientali, contribuendo in modo significativo alla completezza totale di ciascun metagenoma assemblato.

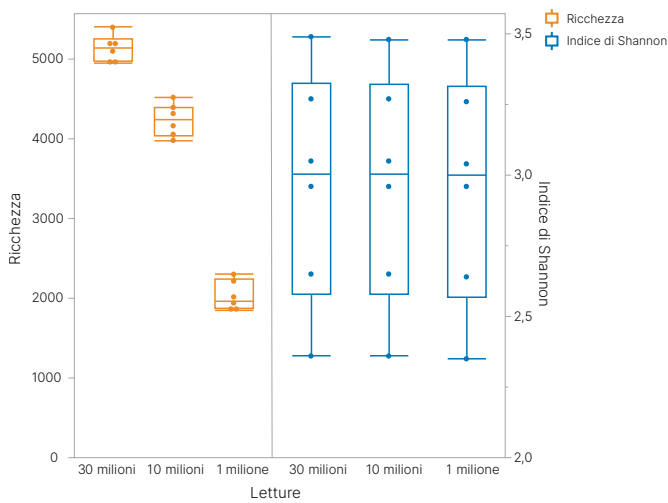


Figura 7: la ricchezza microbica dei campioni di feci aumenta con la profondità di sequenziamento. La ricchezza e l'indice di Shannon sono stati calcolati utilizzando l'app DRAGEN Metagenomics per misurare il numero di specie rilevate e la diversità proporzionale osservata con il sequenziamento a 30 milioni, 10 milioni e 1 milione di letture. Come previsto, la ricchezza diminuisce nei dati sottocampionati, mentre l'indice di Shannon rimane più o meno il medesimo.

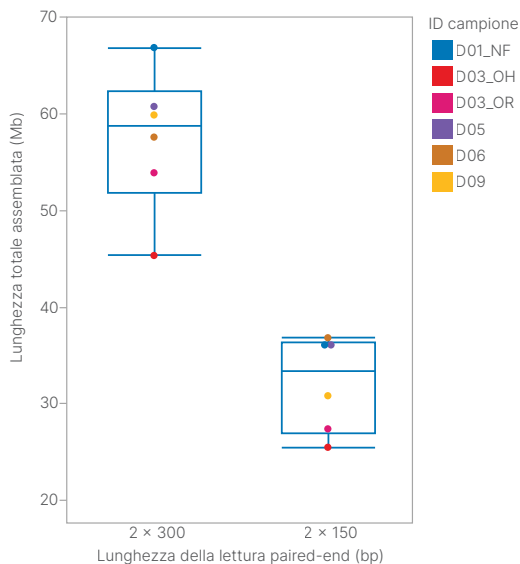


Figura 8: una lunghezza di lettura maggiore comporta un numero complessivo più elevato di basi assemblate per i campioni fecali reali sequenziati sul NextSeq 2000 System. I dati del sequenziamento metagenomico del NextSeq 2000 System sono stati tagliati a 1 milione di letture e le lunghezze di lettura da 2 x 300 bp a 2 x 150 bp utilizzando l'app DRAGEN FASTQ Toolkit. Per i confronti, è stato utilizzato lo SPAdes Genome Assembler allo scopo di generare lunghezze contig totali dalle letture di sequenziamento impostate a 2 x 300 bp e 2 x 150 bp.

Riepilogo

Lo scopo di questa nota applicativa è dimostrare le prestazioni simili dei kit da 600 cicli sul NextSeq 2000 System e sul MiSeq System. Anche il NextSeq 550 System è in grado di eseguire un'analisi metagenomica accurata, ma non dispone di un'opzione a 600 cicli. Tutti e tre i sistemi hanno prodotto assemblaggi *de novo* accurati indipendentemente dallo strumento utilizzato. In tutti e tre i sistemi sono stati prodotti profili metagenomici coerenti, in particolare a livello di diversità proporzionale osservata dei generi.

Nel complesso, il NextSeq 2000 System offre vantaggi superiori rispetto al MiSeq System e al NextSeq 550 System in termini di risoluzione dei dettagli di ricchezza per campioni diversificati e privi di colture grazie a capacità aggiuntive a livello di lunghezza e profondità di lettura.[†] Questi vantaggi sono particolarmente importanti per i campioni di metagenomica complessi, come i campioni fecali o ambientali.

In sintesi, NextSeq 1000/2000 P1 Reagents (600 cycles) e NextSeq 1000/2000 P2 Reagents (600 cycles) sul NextSeq 1000 System e sul NextSeq 2000 System offrono un sequenziamento di alta qualità con tempi di risposta più rapidi rispetto al MiSeq Reagent Kit v3 (600-cycle) sul MiSeq System. I kit da 600 cicli sul NextSeq 1000 System e sul NextSeq 2000 System assicurano punteggi Q30 elevati e tassi di errore ridotti, come dimostrato in questa nota applicativa. Inoltre, l'output dei dati di NextSeq 1000/2000 P1 Reagents (600 cycles) e di NextSeq 1000/2000 P2 (600 cycles) risulta flessibile, ideale per esperimenti di sequenziamento del genoma su piccola e media scala. Infine, i kit da 600 cicli per NextSeq 1000 System e NextSeq 2000 System offrono un maggior numero di applicazioni e la semplicità operativa, assicurando al contempo la qualità dei dati stabilita sul ben collaudato MiSeq System.

[†] Il NextSeq 1000 System funziona similmente al NextSeq 2000 System e dovrebbe dare risultati simili a quest'ultimo utilizzando gli stessi kit da 600 cicli.

Maggiori informazioni

[Piattaforme di sequenziamento Illumina](#)

[NextSeq 1000 System e NextSeq 2000 System](#)

[NextSeq 1000/2000 Reagents](#)

[Illumina DNA Prep](#)

[Protocollo lisato grezzo per NGS Illumina DNA Prep](#)

Bibliografia

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. Pubblicato il 15 luglio 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. Pubblicato il 4 febbraio 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](https://www.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). Pubblicato nel 2023. Consultato il 6 febbraio 2024.



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina Web www.illumina.com/company/legal.html.
M-GL-01147 ITA v1.0