

Um fluxo de trabalho de NGS com metagenômica shotgun para avaliar populações microbianas em amostras complexas

Os kits de 600 ciclos dos sistemas NextSeq™ 1000 e NextSeq 2000 fornecem precisão e flexibilidade para identificação de espécies

illumina®

Classificação metagenômica de amostras complexas

O sequenciamento metagenômico shotgun é um método alternativo para abordagens de sequenciamento de amplicon, como o sequenciamento de RNA ribossômico (rRNA) 16S e o espaçador interno transcrito (Internal Transcribed Spacer, ITS), para avaliar a diversidade microbiana em amostras complexas. Diferentemente das abordagens de amplicon, o sequenciamento metagenômico shotgun com o uso do sequenciamento de última geração (NGS) captura informações genômicas abrangentes para cada organismo presente em uma amostra. A capacidade de capturar genomas completos significa que a metagenômica do sequenciamento shotgun pode identificar espécies perdidas pelo sequenciamento do amplicon¹ e que os dados resultantes contêm informações funcionais não disponíveis nos métodos do amplicon.^{2,3}

Esta nota de aplicação demonstra as semelhanças de desempenho dos sistemas NextSeq 1000, NextSeq 2000 e MiSeq™ em estudos de sequenciamentos metagenômicos shotgun de alto rendimento. Usando dados gerados no NextSeq 550 System, também demonstramos as vantagens que os kits de 600 ciclos têm em relação aos kits de 300 ciclos que são comumente usados em aplicações de metagenômica. Apresentamos dados da população sintética e amostras reais para demonstrar a identificação superior de gêneros e espécies ao usar kits de 600 ciclos.

Métodos

Os kits NextSeq 1000/2000 P1 Reagents (600 cycles) e NextSeq 1000/2000 P2 Reagents (600 cycles) ampliam a capacidade e a saída de sequenciamento dos sistemas NextSeq 1000 e NextSeq 2000 com especificações ideais para o sequenciamento metagenômico shotgun. Os sistemas NextSeq 1000 e NextSeq 2000 usam reagentes prontos para uso e sem fluidos integrados, o que reduz o número de etapas do fluxo de trabalho e o risco de contaminação da amostra. O fluxo de trabalho do sequenciamento metagenômico shotgun integra a preparação da biblioteca, o comprovado Illumina NGS e um botão de análise de dados secundários para obter uma solução completa de descoberta do microbioma (Figura 1).

Preparação da biblioteca

As amostras microbianas de DNA genômico foram obtidas de duas fontes. A primeira amostra utilizada foi a 20 Strain Staggered Mix Genomic Material da organização American Type Culture Collection (ATCC), disponível comercialmente (ATCC, nº de catálogo MSA-1003). Essa amostra da ATCC é uma comunidade microbiana simulada composta por uma distribuição escalonada de DNA genômico preparado, obtida de cepas bacterianas selecionadas com base em atributos como coloração de Gram, conteúdo GC e atributos de esporulação. Um segundo conjunto de amostras reais de fezes, descrito anteriormente,⁴ também foi obtido para análise.

As bibliotecas foram preparadas com Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (Illumina, nº de catálogo 20060060) e IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)



Figura 1: Fluxo de trabalho do sequenciamento metagenômico shotgun de NGS nos sistemas NextSeq 1000 e NextSeq 2000

(Illumina, nº de catálogo 20027213). Os produtos IDT for Illumina DNA/RNA UD Indexes Sets A-D possibilitam que os usuários gerem bibliotecas 384 16S.

Sequenciamento

As bibliotecas preparadas foram agrupadas e carregadas em uma lâmina de fluxo pré-carregada do NextSeq 1000/2000 P1 Reagent Kit (600 cycles), uma lâmina de fluxo do MiSeq Reagent Kit v3 (600 cycles) (Illumina, nº de catálogo MS-102-3003) ou uma lâmina de fluxo do NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, nº de catálogo 20024908). O sequenciamento foi realizado no NextSeq 2000 System, no MiSeq System ou no NextSeq 550 System, respectivamente. Os dados de sequenciamento representativos de todas as corridas estão disponíveis na página Web [BaseSpace™ Sequence Hub demo data](#) (Dados de demonstração do BaseSpace™ Sequence Hub).

Análise

As bibliotecas agrupadas foram demultiplexadas na plataforma de computação genômica em nuvem BaseSpace Sequence Hub. O pipeline DRAGEN™ Metagenomics foi usado para processar dados gerados nos sistemas NextSeq 2000, MiSeq e NextSeq 550. Os metagenomas foram montados usando o SPAdes Genome Assembler. As classificações taxonômicas foram conduzidas por meio do pipeline DRAGEN Metagenomics.

Para comparar os dados do NextSeq 2000 (gerados usando um kit de 600 ciclos) com os dados do NextSeq 500 (gerados usando um kit de 300 ciclos), as leituras do NextSeq 2000 foram cortadas usando o DRAGEN FASTQ

Toolkit, disponível no BaseSpace Sequence Hub. Para possibilitar comparações entre amostras, cada amostra foi reduzida ao mesmo número de leituras (30 milhões, 10 milhões, 1 milhão) por meio do DRAGEN FASTQ Toolkit. A redução das amostras é necessária nos casos em que apenas um subconjunto da amostra pode ser processado por um aplicativo (por exemplo, montagem *de novo* com restrições de memória) ou quando o conjunto de dados completo não é necessário para processar uma amostra (por exemplo, para validar uma abordagem em níveis variados de cobertura genômica).

Resultados

Métricas primárias de sequências aprimoradas

O NextSeq 1000/2000 P1 Reagents (600 cycles), no NextSeq 2000 System, mostra uma porcentagem maior de pontuações de qualidade \geq Q30 em comparação com a corrida de sequenciamento do MiSeq Reagent Kit v3 (600 cycles) no MiSeq System. O NextSeq 1000/2000 P1 Reagents (600 cycles) também fornece até 100 milhões de filtros de passagem de leituras tipo single-end ou 200 milhões de filtros de passagem de leituras tipo paired-end. Com aproximadamente 60 Gb, o NextSeq 1000/2000 P1 Reagents (600 cycles) gera quatro vezes mais dados do que o MiSeq Reagent Kit v3 (600 cycles), que gera aproximadamente 15 Gb. Além disso, as corridas de sequenciamento com o kit NextSeq 1000/2000 P1 Reagents (600 cycles) foram concluídas em cerca de 34 horas, o que é aproximadamente 20 horas a menos do que uma corrida de sequenciamento do MiSeq Reagent Kit v3 (600 cycles) (Figura 2).

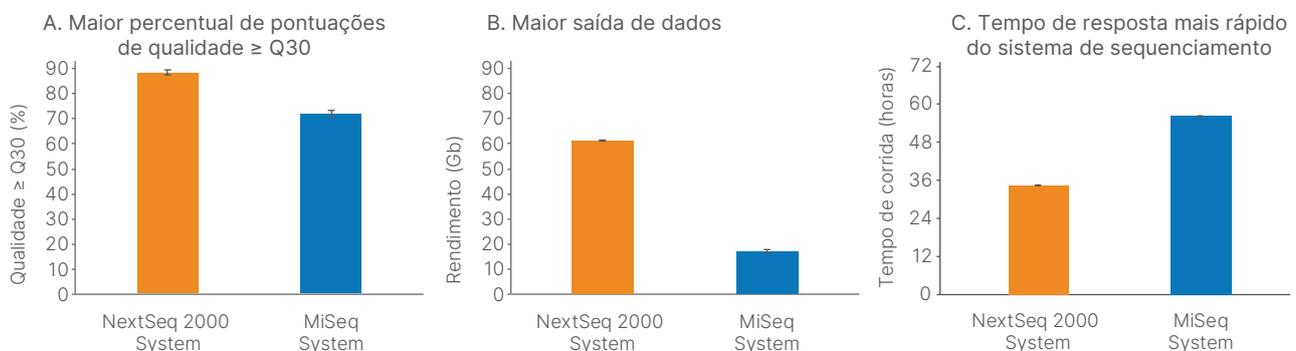


Figura 2: Comparações de métricas de desempenho primário dos sistemas NextSeq 2000 e MiSeq - Em comparação com a corrida do MiSeq Reagent Kit v3 (600 cycles) no MiSeq System, o NextSeq 1000/2000 P1 Reagents (600 cycles) no NextSeq 2000 System oferece: (A) pontuações maiores em relação ao percentual de qualidade \geq Q30, (B) saída de dados 4 vezes maior e (C) redução do tempo de corrida do instrumento em aproximadamente 20 horas, mediante o uso da lâmina de fluxo do NextSeq P1.

O NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, catálogo nº 20024908) é capaz de gerar até 120 Gb de dados de sequenciamento de alta qualidade, com duração da leitura de 2 × 150 bp, em cerca de 29 horas. A especificação de qualidade para este kit é > 75% das bases ≥ Q30 (dados não mostrados).

Comparações de análises metagenômicas

Para comparar o desempenho entre os sistemas, o 20 Strain Staggered Mix Genomic Material foi sequenciado nos sistemas NextSeq 2000, MiSeq e NextSeq 550. O aplicativo DRAGEN Metagenomics no BaseSpace Sequence Hub foi usado para análise posterior, elucidando as classificações taxonômicas. A análise metagenômica NGS identificou todos os membros esperados da comunidade bacteriana simulada e mostrou resultados comparáveis entre os três sistemas de sequenciamento (Figura 3).

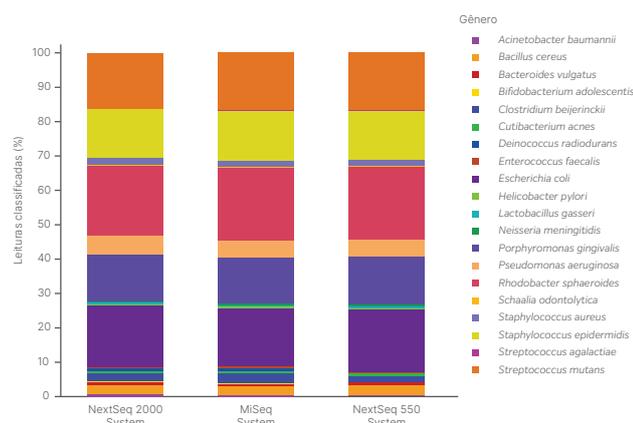


Figura 3: A análise comparativa da composição microbiana revela a composição da amostra 20 Strain Staggered Mix da ATCC, analisada nos sistemas NextSeq 2000, MiSeq e NextSeq 550 - A análise do aplicativo DRAGEN Metagenomics da composição microbiana, demonstra excelente identidade e distribuição de gêneros reproduzíveis.

Embora a amostra da ATCC demonstre desempenho altamente reproduzível, uma amostra simulada pode não ser um indicativo de desempenho com amostras reais. Portanto, testamos o desempenho da detecção de organismos com amostras reais de fezes. Os 20 gêneros mais representados nas amostras de fezes foram comparados entre os sistemas NextSeq 2000, NextSeq 550 e MiSeq (Figura 4). Embora os resultados não sejam tão uniformes com esse tipo de amostra mais complexo, os perfis metagenômicos das comunidades de todas as amostras reais de fezes foram altamente concordantes

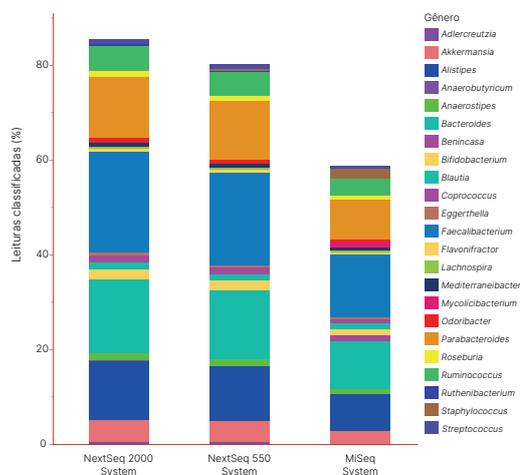


Figura 4: A análise metagenômica revela uma composição comunitária semelhante para uma amostra real de fezes representativa analisada em vários sistemas - Aplicativo DRAGEN Metagenomics para análise da composição microbiana de amostras reais de fezes, limitada aos 20 gêneros mais representados na amostra, sequenciados nos sistemas NextSeq 2000, NextSeq 550 e MiSeq. Os dados demonstram cobertura de gêneros semelhante em todas as plataformas, mesmo em amostras complexas.

entre os sistemas NextSeq 2000, MiSeq e NextSeq 550. Isso indica que os três sistemas são capazes de realizar sequenciamento metagenômico confiável em vários tipos de amostras.

A duração de leitura mais longa melhora a caracterização da amostra

Embora as lâminas de fluxo de 300 ciclos possam fornecer dados significativos de sequenciamento metagenômico, os kits de 600 ciclos oferecem vantagens na identificação de espécies individuais em amostras complexas. Para demonstrar o efeito de leituras mais longas sobre a montagem do metagenoma, as leituras do NextSeq 2000 System foram cortadas de 600 ciclos para 300 ciclos usando o aplicativo DRAGEN FASTQ Toolkit. O Kraken2, um classificador taxonômico baseado em k-mer,* foi usado para determinar a porcentagem de leituras classificadas para cada amostra em 600 ciclos ou 300 ciclos (Figura 5). Os dados sugerem que leituras mais longas fornecem algumas melhorias para a classificação taxonômica baseada em k-mer com diversas amostras ambientais.

* A classificação taxonômica do Kraken2 Metagenomics está disponível no aplicativo Illumina DRAGEN Metagenomics BaseSpace.

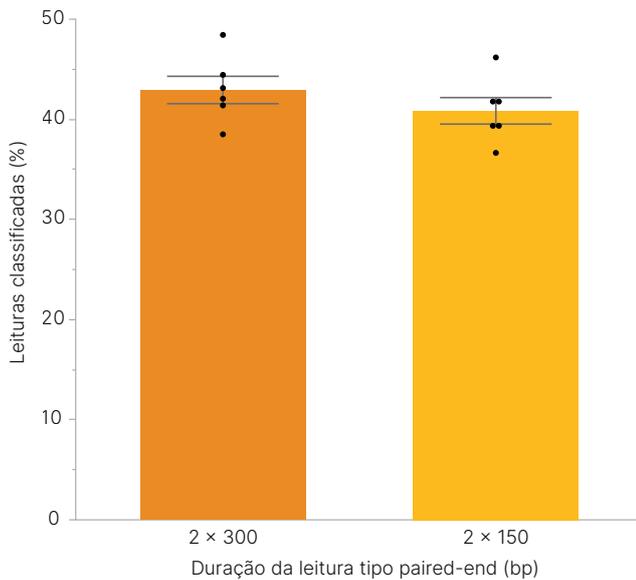


Figura 5: Uma maior duração da leitura melhora a classificação das leituras para amostras reais de fezes sequenciadas no NextSeq 2000 - O aplicativo DRAGEN FASTQ Toolkit foi usado para cortar leituras do NextSeq 2000 System de 600 ciclos (2 x 300 bp) para 300 ciclos (2 x 150 bp). Em seguida, o Kraken2 foi usado para determinar a porcentagem de leituras classificadas em cada amostra. A profundidade de leitura foi de 30 milhões de leituras para a análise. As barras de erro representam um erro padrão da média.

Usando os mesmos dados cortados, os impactos da duração da leitura na riqueza da amostra (ou seja, o número de espécies detectadas em uma amostra) e o índice de Shannon (ou seja, a representação proporcional das espécies detectadas na amostra) também foram examinados. Essas métricas indicam que a diversidade microbiana observada nas amostras de fezes aumenta quando a duração da leitura aumenta, enquanto a proporção de espécies detectadas quantificadas pelo índice de Shannon permanece relativamente inalterada, conforme esperado (Figura 6).

A profundidade superior de sequenciamento melhora a caracterização da amostra

Em seguida, examinamos a importância da profundidade de leitura em medições da diversidade populacional para as amostras reais de fezes. Usando o DRAGEN FASTQ Toolkit, o número de leituras do NextSeq 2000 System foi reduzido para 30 milhões, 10 milhões e 1 milhão. A riqueza e o índice de Shannon foram calculados usando o aplicativo DRAGEN Metagenomics.

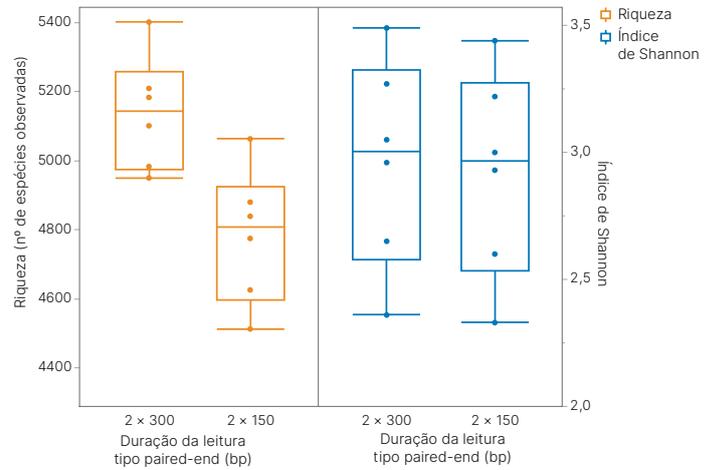


Figura 6: A riqueza microbiana das amostras de fezes aumenta com a duração de leitura mais longa - O aplicativo DRAGEN FASTQ Toolkit foi usado para cortar leituras do NextSeq 2000 System de 600 ciclos (2 x 300 bp) para 300 ciclos (2 x 150 bp). Em seguida, o aplicativo DRAGEN Metagenomics foi usado para calcular a riqueza e o índice de Shannon para quantificar o número de espécies detectadas e a diversidade proporcional, respectivamente. A riqueza aumenta com leituras mais longas, enquanto o índice de Shannon permanece relativamente inalterado. A profundidade de leitura foi de 30 milhões de leituras para a análise.

As métricas de diversidade demonstraram que a diversidade microbiana das amostras de fezes aumenta quando a profundidade do sequenciamento aumenta, enquanto a proporção das espécies indicadas pelo índice de Shannon permanece relativamente inalterada (Figura 7).

Leituras mais longas melhoram a identificação taxonômica

Em relação à determinação do perfil de diversas populações microbianas ambientais, um dos desafios atuais é a falta de genomas de referência completos para muitas espécies raras e não cultiváveis. Geralmente, o número de contigs montados para populações microbianas altamente diversas é maior para durações de leitura mais longas, como mostrado pelo maior comprimento total da montagem. (Figura 8). O sequenciamento metagenômico shotgun com durações de leitura de sequenciamento Illumina 2 x 301 bp geralmente melhora a montagem *de novo* de metagenomas de amostras ambientais, contribuindo significativamente para a integridade geral de cada metagenoma montado.

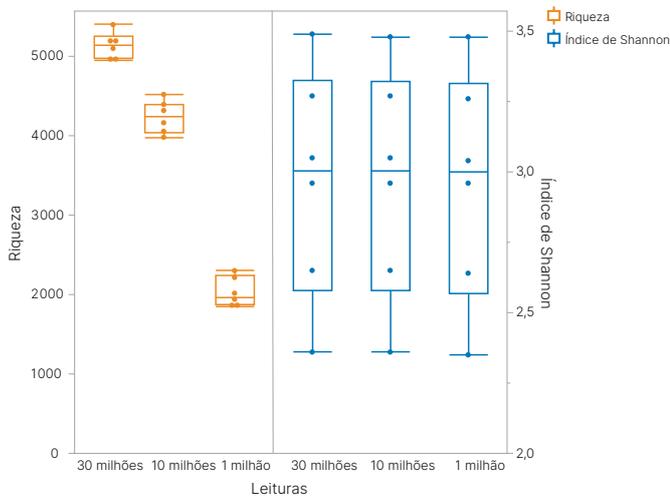


Figura 7: A riqueza microbiana das amostras de fezes aumenta à medida que a profundidade do sequenciamento aumenta. Os índices de riqueza e de Shannon foram calculados com o aplicativo DRAGEN Metagenomics para medir o número de espécies detectadas e a diversidade proporcional observada com o sequenciamento de 30 milhões, 10 milhões e 1 milhão de leituras. Como esperado, a riqueza diminui nos dados com redução de amostras, enquanto o índice de Shannon permanece aproximadamente o mesmo.

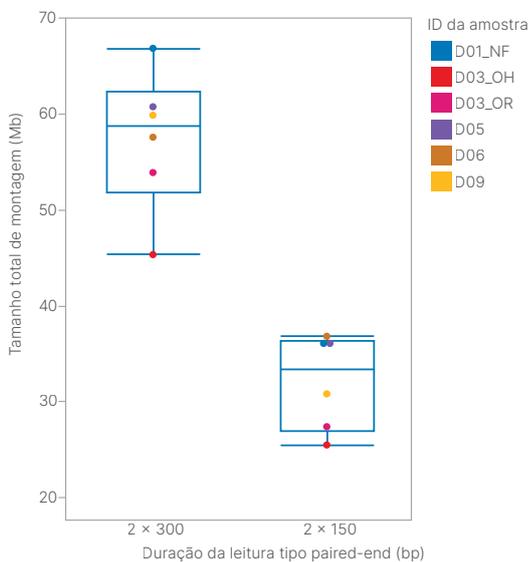


Figura 8: A duração da leitura mais longa sustenta o aumento do número geral de bases montadas para amostras reais de fezes sequenciadas no NextSeq 2000 System - Os dados de sequenciamento metagenômico do NextSeq 2000 System foram cortados para 1 milhão de leituras, e as durações de leituras de 2 x 300 bp para 2 x 150 bp, usando o aplicativo DRAGEN FASTQ Toolkit. Para fins de comparação, o SPAdes Genome Assembler foi usado para gerar comprimentos totais de contigs a partir de leituras de sequenciamento definidas em 2 x 300 bp e 2 x 150 bp.

Resumo

O objetivo desta nota de aplicativo é demonstrar o desempenho semelhante dos kits de 600 ciclos no NextSeq 2000 System e no MiSeq System. O NextSeq 550 System também é capaz de realizar análises metagenômicas precisas, mas não tem uma opção de 600 ciclos. Os três sistemas obtiveram montagens *de novo* precisas, independentemente do instrumento usado. Os perfis metagenômicos concordantes, particularmente a diversidade proporcional observada de gêneros, foram alcançados nos três sistemas.

Em geral, o NextSeq 2000 System oferece vantagens em relação ao MiSeq System e NextSeq 550 System em termos de resolução de detalhes de riqueza da amostra em amostras diversas e livres de cultura, devido ao comprimento adicional de leitura e às capacidades de profundidade de leitura.[†] Essas vantagens são especialmente importantes para amostras metagenômicas complexas, como amostras de fezes ou ambientais.

Em resumo, o NextSeq 1000/2000 P1 Reagents (600 cycles) e o NextSeq 1000/2000 P2 Reagents (600 cycles) nos sistemas NextSeq 1000 e NextSeq 2000 oferecem sequenciamento de alta qualidade com um tempo de resposta mais rápido do que o MiSeq Reagent Kit v3 (600 cycles) no MiSeq System. Os kits de 600 ciclos nos sistemas NextSeq 1000 e NextSeq 2000 mantêm altas pontuações Q30 e baixas taxas de erro, conforme demonstrado nesta nota de aplicação. Além disso, os produtos NextSeq 1000/2000 P1 Reagent (600 cycles) e NextSeq 1000/2000 P2 Reagent (600 cycles) fornecem saída de dados flexível ideal para experimentos de sequenciamento de genoma de pequena e média escala. Por fim, os kits de 600 ciclos para os sistemas NextSeq 1000 e NextSeq 2000 possibilitam a expansão de aplicativos e a simplicidade operacional, mantendo a qualidade dos dados determinada pelo comprovado MiSeq System.

[†] O NextSeq 1000 System é funcionalmente semelhante ao NextSeq 2000 System e deve obter resultados semelhantes ao instrumento usando os mesmos kits de 600 ciclos.

Saiba mais

[Plataformas de sequenciamento da Illumina](#)

[Sistemas NextSeq 1000 e NextSeq 2000](#)

[NextSeq 1000/2000 Reagents](#)

[Illumina DNA Prep](#)

[Protocolo de lisado bruto Illumina DNA Prep para NGS](#)

Referências

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 em 15 de julho 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 em 4 de fevereiro 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. Publicado em 2023. Acessado em 6 de fevereiro de 2024.



1 800-809-4566, ligação gratuita (EUA) | tel. +1 (858) 202-4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Todos os direitos reservados. Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.
M-GL-01147 PTB v1.0