

Ein NGS-Workflow für die Shotgun- Metagenomik zur Beurteilung mikrobieller Populationen in komplexen Proben

Anhand der 600-Zyklus-
Kits für das NextSeq™ 1000
System und das NextSeq
2000 System lassen sich
Spezies genau und flexibel
bestimmen



Klassifizierung komplexer Proben mithilfe von Metagenomik

Bei der metagenomischen Shotgun-Sequenzierung handelt es sich um eine Alternative zu Amplikonsequenzierungsverfahren wie der 16S- und der ITS-rRNA-Sequenzierung (Internal Transcribed Spacer der ribosomalen RNA), die die Beurteilung der mikrobiellen Vielfalt in komplexen Proben ermöglicht. Im Gegensatz zu Amplikonverfahren erfasst die metagenomische Shotgun-Sequenzierung mithilfe von NGS (Next-Generation Sequencing, Sequenzierung der nächsten Generation) umfassende genomische Informationen zu allen in der Probe enthaltenen Organismen. Dank der Möglichkeit zur Erfassung vollständiger Genome lassen sich mit der Shotgun-Metagenomik Spezies bestimmen, die bei der Amplikonsequenzierung¹ unbeachtet bleiben. Außerdem enthalten die Ergebnisdaten Informationen zur Funktion, die Amplikonmethoden nicht liefern.^{2,3}

Der vorliegende Anwendungshinweis zeigt die Vergleichbarkeit der Leistung des NextSeq 1000 System, des NextSeq 2000 System und des MiSeq™ System für Shotgun-Metagenomikstudien mit hohem Durchsatz. Anhand von auf dem renommierten NextSeq 550 System generierten Daten werden zudem die Vorteile von 600-Zyklen-Kits bei Metagenomikanwendungen gegenüber den in der Regel eingesetzten 300-Zyklen-Kits aufgezeigt. Daten aus Proben mit synthetischen und natürlichen Populationen dienen als Beleg für die überlegene Leistung von Kits mit 600 Zyklen beim Nachweis von Gattungen und Spezies.

Methoden

Das Kit NextSeq 1000/2000 P1 Reagents (600 cycles) und das Kit NextSeq 1000/2000 P2 Reagents (600 cycles) erweitern die Kapazität und Sequenzierungsausgabe des NextSeq 1000 System sowie des NextSeq 2000 System. Die erreichten Spezifikationen sind ideal für die metagenomische Shotgun-Sequenzierung. Das NextSeq 1000 System und das NextSeq 2000 System nutzen Load-and-Go-Reagenzien ohne integrierte Fluidik. Dadurch werden die Anzahl der Workflowschritte und das Risiko einer Probenkontamination reduziert. Der Workflow für die Shotgun-Metagenomik umfasst die Bibliotheksvorbereitung, bewährte Illumina-NGS und die Sekundäranalyse der Daten auf Tastendruck. Damit steht eine vollständige Lösung für die Mikrobiombestimmung zur Verfügung (Abbildung 1).

Bibliotheksvorbereitung

Die Proben mit mikrobieller genomischer DNA stammten aus zwei Quellen. Bei der ersten Probe handelte es sich um die im Handel erhältliche American Type culture collection ATCC 20 Strain Staggered Mix Genomic Material (ATCC, Katalog-Nr. MSA-1003). Diese ATCC-Probe enthält eine synthetisch zusammengesetzte mikrobielle Gemeinschaft, die aus vorbereiteter genomischer DNA aus Bakterienstämmen in gestaffelter Verteilung besteht. Ausschlaggebend für die Auswahl waren Attribute wie Gramfärbung, GC-Gehalt und Sporulation. Außerdem wurde, wie bereits erläutert, ein zweiter Satz mit natürlichen Stuhlproben⁴ zur Analyse entnommen.

Die Bibliotheken wurden mit Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (Illumina, Katalog-Nr. 20060060) und IDT for Illumina DNA/RNA UD Indexes



Abbildung 1: NGS-Workflow für die Shotgun-Metagenomik auf dem NextSeq 1000 System und dem NextSeq 2000 System

Set A, Tagmentation (96 Indexes, 96 Samples) (Illumina, Katalog-Nr. 20027213) vorbereitet. Mithilfe der IDT für Illumina DNA/RNA UD Indexes Sets A bis D können Anwender 384 16S-Bibliotheken generieren.

Sequenzierung

Die vorbereiteten Bibliotheken wurden gepoolt und in eine vorgefüllte Fließzelle aus dem Kit NextSeq 1000/2000 P1 Reagents (600 cycles), eine Fließzelle aus dem MiSeq Reagent Kit v3 (600-cycle) (Illumina, Katalog-Nr. MS-102-3003) oder eine Fließzelle aus dem NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, Katalog-Nr. 20024908) geladen. Die Sequenzierung erfolgte auf einem NextSeq 2000 System, einem MiSeq System bzw. einem NextSeq 550 System. Für sämtliche Läufe sind auf der [BaseSpace™ Sequence Hub-Webseite](#) mit Demodaten repräsentative Sequenzierungsdaten verfügbar.

Analyse

Die gepoolten Bibliotheken wurden in der Genomik-Cloud-Computing-Plattform BaseSpace Sequence Hub demultiplexiert. Die auf dem NextSeq 2000 System, dem MiSeq System und dem NextSeq 550 System generierten Daten wurden mithilfe der DRAGEN™ Metagenomics-Pipeline verarbeitet. Die Metagenome wurden mit SPAdes Genome Assembler assembliert. Die taxonomischen Klassifizierungen erfolgten durch die DRAGEN Metagenomics-Pipeline.

Zum Vergleich der anhand eines 600-Zyklen-Kits auf dem NextSeq 2000 System generierten Daten mit den anhand eines 300-Zyklen-Kits auf dem NextSeq 500 System generierten Daten wurden die NextSeq 2000-

Reads mit dem DRAGEN FASTQ Toolkit gekürzt, das auf BaseSpace Sequence Hub verfügbar ist. Die Proben wurden mithilfe des DRAGEN FASTQ Toolkit einem Downsampling auf die gleiche Anzahl von Reads (30, 10 und 1 Mio.) unterzogen, um diese vergleichbar zu machen. Das Downsampling ist erforderlich, wenn von einer Anwendung nur Probenteile verarbeitet werden können (z. B. bei der *De-novo*-Assemblierung mit beschränkter Speichermenge) oder wenn zur Verarbeitung einer Probe nicht der vollständige Datensatz benötigt wird (z. B. für die Validierung eines Verfahrens mit unterschiedlichen Werten für die Genom-Coverage).

Ergebnisse

Verbesserte Sequenz-Primärmetriken

Bei NextSeq 1000/2000 P1 Reagents (600 cycles) auf dem NextSeq 2000 System zeigt sich gegenüber dem Lauf mit dem MiSeq Reagent Kit v3 (600-cycle) auf dem MiSeq System ein höherer Prozentsatz an Qualitäts-Scores \geq Q30. Außerdem bietet NextSeq 1000/2000 P1 Reagents (600 cycles) bis zu 100 Mio Single-End-Reads nach Filterung oder 200 Mio. Paired-End-Reads nach Filterung. Mit ca. 60 Gb generiert NextSeq 1000/2000 P1 Reagents (600 cycles) viermal mehr Daten als das MiSeq Reagent Kit v3 (600-cycle) mit ca. 15 Gb. Hinzu kommt, dass Sequenzierumläufe mit dem Kit NextSeq 1000/2000 P1 Reagents (600 cycles) nur ca. 34 Stunden benötigen und damit ca. 20 Stunden weniger als bei einem Sequenzierlauf mit dem MiSeq Reagent Kit v3 (600-cycle) ([Abbildung 2](#)).

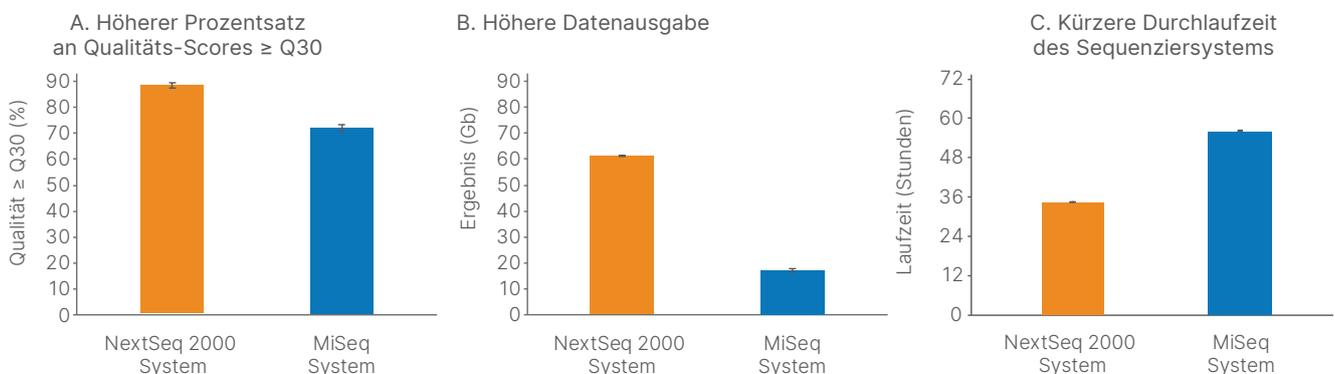


Abbildung 2: Vergleich der primären Leistungsmetriken für das NextSeq 2000 System und das MiSeq System: Im Vergleich zum Lauf mit dem MiSeq Reagent Kit v3 (600-cycle) auf dem MiSeq System bietet NextSeq 1000/2000 P1 Reagents (600 cycles) auf dem NextSeq 2000 System bei Verwendung der NextSeq P1-Fließzelle (A) einen höheren Prozentsatz an Qualitäts-Scores \geq Q30, (B) eine viermal höhere Datenausgabe und (C) eine ca. 20 Stunden kürzere Gerätelaufzeit.

Das NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, Katalog-Nr. 20024908) generiert in ca. 29 Stunden bis zu 120 Gb Sequenzierungsdaten mit hoher Qualität und einer Read-Länge von 2 × 150 bp. Die Qualitätsspezifikation für dieses Kit entspricht > 75 % der Basen ≥ Q30 (Daten nicht dargestellt).

Vergleich metagenomischer Analysen

Für den systemübergreifenden Leistungsvergleich wurde 20 Strain Staggered Mix Genomic Material auf dem NextSeq 2000 System, dem MiSeq System und dem NextSeq 550 System sequenziert. Die nachgeschaltete Analyse zur taxonomischen Klassifizierung erfolgte in der DRAGEN Metagenomics-App auf BaseSpace Sequence Hub. Bei der metagenomischen NGS-Analyse wurden alle erwarteten Mitglieder der synthetischen mikrobiellen Gemeinschaft bestimmt. Alle drei Sequenziersysteme lieferten vergleichbare Ergebnisse (Abbildung 3).

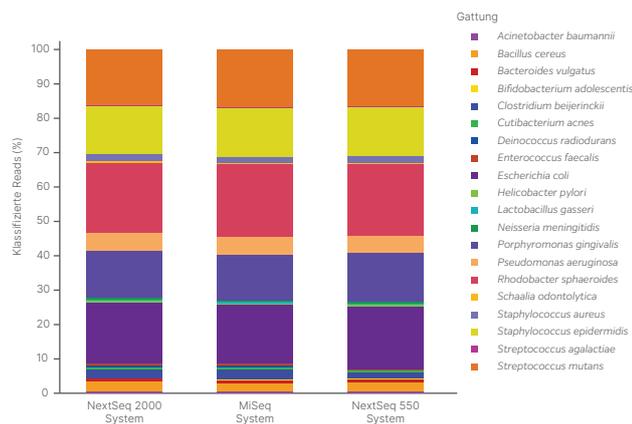


Abbildung 3: Die vergleichende Analyse der mikrobiellen Zusammensetzung zeigt die für ATCC 20 Strain Staggered Mix auf dem NextSeq 2000 System, dem MiSeq System und dem NextSeq 550 System analysierte Zusammensetzung: Die Analyse der mikrobiellen Zusammensetzung mithilfe der DRAGEN Metagenomics-App ergibt herausragende und reproduzierbare Werte für den Nachweis und die Verteilung der Gattungen.

Bei der ATCC-Probe zeigt sich zwar eine hochgradig reproduzierbare Leistung, die Ergebnisse sind jedoch nicht zuverlässig auf natürliche Proben übertragbar. Aus diesem Grund wurde die Leistung beim Nachweis von Organismen anhand natürlicher Stuhlproben getestet. Die Ergebnisse des NextSeq 2000 System, des NextSeq 550 System und des MiSeq System für die 20 häufigsten Gattungen in den Stuhlproben wurden verglichen (Abbildung 4). Die bei diesem komplexeren Probentyp erzielten Ergebnisse sind zwar weniger einheitlich, jedoch zeigte sich bei den Metagenomik-Profilen der mikrobiellen Gemeinschaft sämtlicher natürlicher

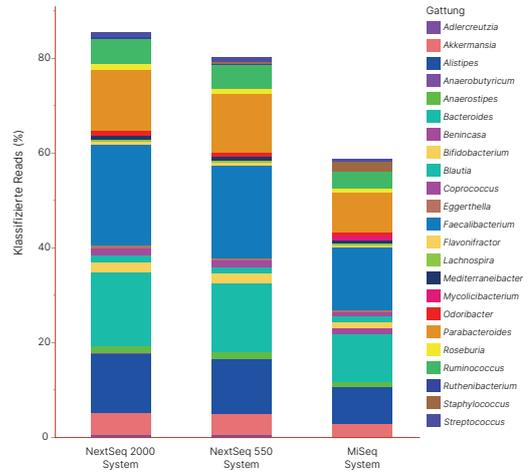


Abbildung 4: Die metagenomische Analyse zeigt bei der Analyse auf mehreren Systemen vergleichbare Ergebnisse für die Zusammensetzung der mikrobiellen Gemeinschaft: Analyse der mikrobiellen Zusammensetzung natürlicher Stuhlproben mit der DRAGEN Metagenomics-App, begrenzt auf die 20 am häufigsten in der Probe vorhandenen Gattungen, Sequenzierung auf dem NextSeq 2000 System, dem NextSeq 550 System und dem MiSeq System. Die Daten zeigen selbst bei komplexen Proben plattformübergreifend eine vergleichbare Coverage der Gattungen.

Stuhlproben eine starke Übereinstimmung zwischen dem NextSeq 2000 System, dem MiSeq System und dem NextSeq 550 System. Folglich sind alle drei Systeme in der Lage, eine zuverlässige Metagenomiksequenzierung unterschiedlicher Probentypen durchzuführen.

Größere Read-Länge verbessert die Probencharakterisierung

Obwohl Fließzellen für 300 Zyklen bei der Metagenomiksequenzierung aussagekräftige Daten liefern können, bieten 600-Zyklus-Kits Vorteile bei der Bestimmung einzelner, in komplexen Proben enthaltener Spezies. Zur Veranschaulichung der Auswirkungen längerer Reads auf die Metagenom-Assemblierung wurden die Reads des NextSeq 2000 System mit der DRAGEN FASTQ Toolkit-App von 600 auf 300 Zyklen gekürzt. Der Prozentsatz klassifizierter Reads für die einzelnen Proben bei entweder 600 oder 300 Zyklen wurde mit Kraken2, einer k-mer-basierten Anwendung zur taxonomischen Klassifikation* bestimmt (Abbildung 5). Mithilfe unterschiedlicher Umweltproben generierte Daten deuten darauf hin, dass längere Reads Vorteile für die k-mer-basierte taxonomische Klassifizierung bieten.

* Die taxonomische Klassifizierung mit Kraken2 Metagenomics steht über die Illumina DRAGEN Metagenomics BaseSpace-App zur Verfügung.

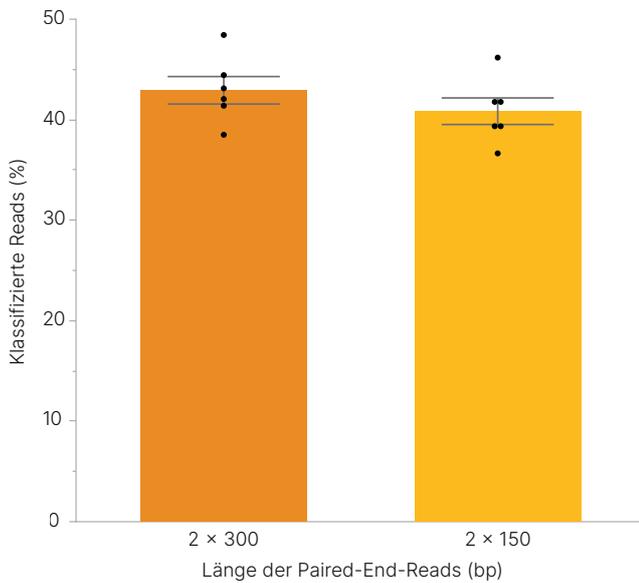


Abbildung 5: Eine höhere Read-Länge verbessert die Klassifizierung der Reads für natürliche Stuhlproben, die auf dem NextSeq 2000 System sequenziert wurden: Die Reads des NextSeq 2000 System wurden mithilfe der DRAGEN FASTQ Toolkit-App von 600 Zyklen (2×300 bp) auf 300 Zyklen (2×150 bp) gekürzt. Anschließend wurde mit Kraken2 der Prozentsatz klassifizierter Reads für die einzelnen Proben bestimmt. Die Read-Tiefe für die Analyse betrug 30 Mio. Reads. Fehlerbalken stellen einen Standardfehler des Mittelwerts dar.

Zudem wurden unter Verwendung derselben gekürzten Daten die Auswirkungen der Read-Länge auf die Probenvielfalt (d. h. die Anzahl der in einer Probe nachgewiesenen Spezies) und den Shannon-Index (d. h. den proportionalen Anteil der in der Probe nachgewiesenen Spezies) untersucht. Diese Metriken deuten darauf hin, dass mit steigender Read-Länge auch die beobachtete mikrobielle Vielfalt der Stuhlproben zunimmt. Der mit dem Shannon-Index angegebene Anteil der nachgewiesenen Spezies bleibt dabei, wie erwartet, relativ unverändert (Abbildung 6).

Höhere Sequenzierungstiefe verbessert die Probencharakterisierung

Im nächsten Schritt wurde die Bedeutung der Read-Tiefe für Werte zur Populationsvielfalt in natürlichen Stuhlproben untersucht. Die Anzahl der Reads des NextSeq 2000 System wurde mit dem DRAGEN FASTQ Toolkit einem Downsampling auf 30, 10 und 1 Mio. Reads unterzogen. Die Vielfalt und der Shannon-Index wurden mit der DRAGEN Metagenomics-App berechnet.

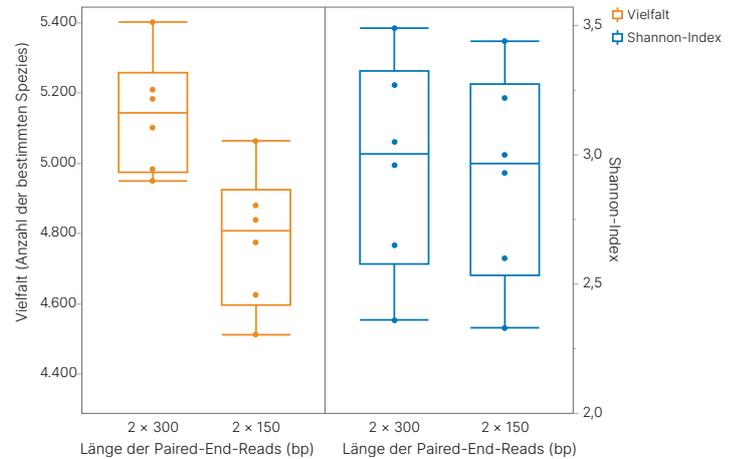


Abbildung 6: Die mikrobielle Vielfalt der Stuhlproben nimmt mit höherer Read-Länge zu: Die Reads des NextSeq 2000 System wurden mit der DRAGEN FASTQ Toolkit-App von 600 Zyklen (2×300 bp) auf 300 Zyklen (2×150 bp) gekürzt. Im nächsten Schritt wurden mithilfe der DRAGEN Metagenomics-App die Vielfalt und der Shannon-Index berechnet, um die Anzahl der nachgewiesenen Spezies bzw. die proportionale Vielfalt zu quantifizieren. Die ermittelte Vielfalt wächst mit der Read-Länge, während der Shannon-Index relativ unverändert bleibt. Die Read-Tiefe für die Analyse betrug 30 Mio. Reads.

Die Vielfaltsmetriken zeigten, dass die erfasste mikrobielle Vielfalt der Stuhlproben mit wachsender Sequenzierungstiefe zunimmt, während der mit den Shannon-Index angegebene Anteil der Spezies relativ unverändert bleibt (Abbildung 7).

Längere Reads verbessern die taxonomische Bestimmung

Eine der derzeitigen Herausforderungen beim Profiling vielfältiger mikrobieller Populationen in Umweltproben ist der Mangel an vollständigen Referenzgenomen für viele seltene und nicht kultivierbare Spezies. Im Allgemeinen ist die Anzahl der assemblierten Contigs für hochgradig vielfältige mikrobielle Populationen bei höheren Read-Längen größer. Das zeigt sich an der höheren Gesamtlänge der Assemblierung (Abbildung 8). Die metagenomische Shotgun-Sequenzierung mit Illumina-Sequenzierungs-Read-Längen von 2×301 bp verbessert im Allgemeinen die *De-novo*-Assemblierung von Metagenomen aus Umweltproben und trägt signifikant zur Gesamtvollständigkeit der einzelnen assemblierten Metagenome bei.

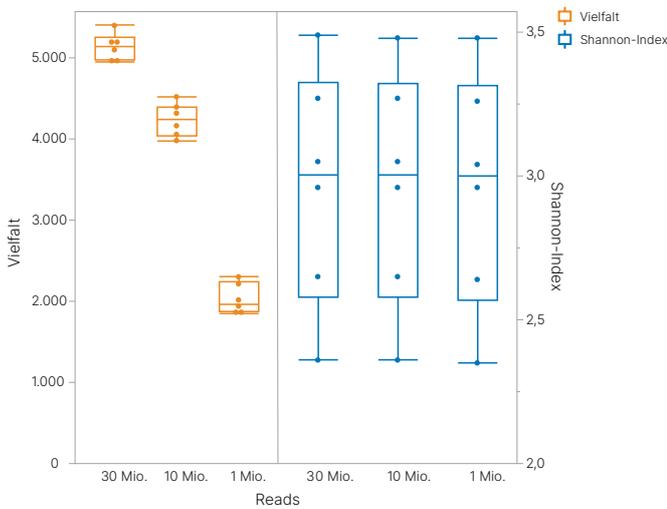


Abbildung 7: Die ermittelte mikrobielle Vielfalt der Stuhlproben nimmt mit der Sequenzierungstiefe zu: Die Vielfalt und der Shannon-Index wurden mithilfe der DRAGEN Metagenomics-App berechnet. Damit wurden die bei der Sequenzierung bei 30, 10 und 1 Mio. Reads beobachtete Anzahl der nachgewiesenen Spezies sowie die entsprechende proportionale Vielfalt bestimmt. Wie erwartet, nimmt die Vielfalt bei einem Downsampling unterzogenen Daten ab, während der Shannon-Index ungefähr gleich bleibt.

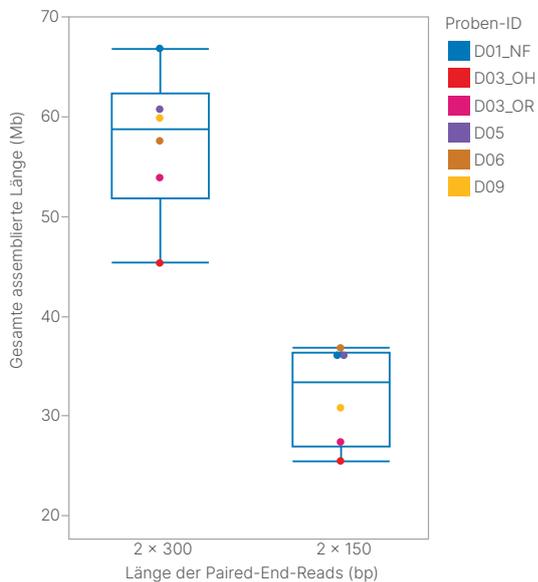


Abbildung 8: Eine höhere Read-Länge hat bei natürlichen, mit dem NextSeq 2000 System sequenzierten Stuhlproben eine insgesamt höhere Anzahl assemblierter Basen zur Folge: Die metagenomischen Sequenzierungsdaten des NextSeq 2000 System wurden mit der DRAGEN FASTQ Toolkit-App auf 1 Mio. Reads gekürzt. Die Read-Längen wurden von 2 x 300 bp auf 2 x 150 bp gekürzt. Die Vergleiche erfolgten mit SPAdes Genome Assembler. Die Anwendung generiert die Contig-Gesamtlängen aus Sequenzierungs-Reads, die auf 2 x 300 bp und 2 x 150 bp festgelegt wurden.

Zusammenfassung

Der vorliegende Anwendungshinweis zeigt die vergleichbare Leistung der 600-Zyklus-Kits auf dem NextSeq 2000 System und dem MiSeq System. Das NextSeq 550 System eignet sich ebenfalls für genaue metagenomische Analysen, verfügt jedoch nicht über eine 600-Zyklus-Option. Mit allen drei Systemen lässt sich eine genaue *De-novo*-Assemblierung erreichen. Die Metagenomikprofile stimmten insbesondere hinsichtlich der beobachteten proportionalen Vielfalt der Gattungen auf allen drei Systemen überein.

Insgesamt bietet das NextSeq 2000 System dank der höheren Read-Länge und -tiefe Vorteile gegenüber dem MiSeq System und dem NextSeq 550 System beim Detailgrad der Bestimmung vielfältiger, kulturfreier Proben.[†] Diese Vorteile sind besonders wichtig bei komplexen Metagenomikproben wie Stuhl- oder Umweltproben.

Zusammenfassend lässt sich sagen, dass NextSeq 1000/2000 P1 Reagents (600 cycles) und NextSeq 1000/2000 P2 Reagents (600 cycles) auf dem NextSeq 1000 System und dem NextSeq 2000 System eine hochwertige Sequenzierung ermöglichen. Die Durchlaufzeit ist kürzer als beim MiSeq Reagent Kit v3 (600-cycle) auf dem MiSeq System. Die 600-Zyklus-Kits auf dem NextSeq 1000 System und dem NextSeq 2000 System zeichnen sich, wie im vorliegenden Anwendungshinweis dargestellt, durch hohe Q30-Scores und niedrige Fehlerraten aus. Darüber hinaus bieten NextSeq 1000/2000 P1 (600 cycles) und NextSeq 1000/2000 P2 Reagents (600 cycles) eine flexible Datenausgabe, die sich ideal für kleine und mittlere Genomsequenzierungsversuche eignet. Letztlich ermöglichen die 600-Zyklus-Kits für das NextSeq 1000 System und das NextSeq 2000 System eine Ausdehnung des Einsatzbereichs und sind bei derselben Datenqualität wie auf dem bewährten MiSeq System besonders anwenderfreundlich.

[†] Das NextSeq 1000 System ist in funktioneller Hinsicht baugleich mit dem NextSeq 2000 System und sollte mit den gleichen 600-Zyklus-Kits vergleichbare Ergebnisse liefern wie dieses Gerät.

Weitere Informationen

[Sequenzierungsplattformen von Illumina](#)

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

[NextSeq 1000/2000 Reagents](#)

[Illumina DNA Prep](#)

[Illumina DNA Prep-Rohlysat-Protokoll für NGS](#)

Quellen

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. Veröffentlicht am 15. Juli 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. Veröffentlicht am 4. Februar 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. Veröffentlicht 2023. Aufgerufen am 6. Februar 2024.



1 800 8094566 (USA, gebührenfrei) | +1 858 2024566 (Tel. außerhalb der USA)
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Inhaber.
Spezifische Informationen zu Marken finden Sie unter www.illumina.com/company/legal.html.
M-GL-01147 DEU v1.0