

Un flux de travail de SNG métagénomique aléatoire pour l'évaluation des populations microbiennes dans les échantillons complexes

Les trousse de 600 cycles des systèmes NextSeq^{MC} 1000 et NextSeq 2000 offrent précision et flexibilité pour l'identification des espèces



Classification métagénomique des échantillons complexes

Le séquençage métagénomique aléatoire est une méthode de rechange aux approches de séquençage des amplicons, telles que le séquençage de l'ARN ribosomique (ARNr) 16S et des espaceurs transcrits internes (ITS, Internal Transcribed Spacer), pour évaluer la diversité microbienne dans les échantillons complexes. Contrairement aux approches basées sur les amplicons, le séquençage métagénomique aléatoire utilisant le séquençage de nouvelle génération (SNG) capture des renseignements génomiques complets pour chaque organisme présent dans un échantillon. La capacité à capturer des génomes complets signifie que la métagénomique aléatoire peut identifier les espèces non identifiées par le séquençage des amplicons¹ et que les données obtenues contiennent des renseignements fonctionnels non disponibles avec les méthodes basées sur les amplicons^{2,3}.

Cette note d'application démontre les similitudes de performance des systèmes NextSeq 1000, NextSeq 2000 et MiSeq^{MC} pour les études à haut débit en métagénomique aléatoire. En utilisant les données générées sur le vénérable NextSeq 550 System, nous démontrons également les avantages que les trousse de 600 cycles ont par rapport aux trousse de 300 cycles couramment utilisées dans les applications de métagénomique. Nous présentons des données provenant de la population synthétique et d'échantillons de sujets réels pour démontrer une identification supérieure des genres et des espèces lors de l'utilisation de trousse de 600 cycles.

Méthodes

La trousse NextSeq 1000/2000 P1 Reagents (600 cycles) et la trousse NextSeq 1000/2000 P2 Reagents (600 cycles) augmentent la capacité et le débit de séquençage des systèmes NextSeq 1000 et NextSeq 2000 avec des spécifications idéales pour le séquençage métagénomique aléatoire. Les systèmes NextSeq 1000 et NextSeq 2000 utilisent des réactifs chargement-exécution sans fluide intégrée, ce qui réduit le nombre d'étapes du flux de travail et le risque de contamination des échantillons. Le flux de travail de métagénomique aléatoire intègre la préparation des librairies, le SNG éprouvé d'Illumina et l'analyse secondaire des données par bouton-poussoir pour une solution complète de découverte des microbiomes (figure 1).

Préparation des librairies

Des échantillons d'ADN génomique microbien ont été obtenus à partir de deux sources. Le premier échantillon était « 20 Strain Staggered Mix Genomic Material » d'American Type Culture Collection (ATCC, référence n° MSA-1003); disponible sur le marché. Cet échantillon ATCC est une communauté microbienne simulée composée d'une répartition échelonnée d'ADN génomique préparé à partir de souches bactériennes sélectionnées en fonction d'attributs tels que le colorant de Gram, la teneur en GC et les attributs de sporulation. Une deuxième série d'échantillons de selles de sujets réels, décrite précédemment⁴, a également été obtenue pour analyse.

Les librairies ont été préparées à l'aide d'Illumina DNA Prep, (M) Tagmentation (24 échantillons, IPB) (Illumina, référence n° 20060060) et d'IDT for Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 index, 96 échantillons)



Figure 1 : Flux de travail de SNG métagénomique aléatoire sur les systèmes NextSeq 1000 et NextSeq 2000

(Illumina, référence n° 20027213). Les ensembles A à D d'IDT for Illumina DNA/RNA UD Indexes permettent aux utilisateurs de générer 384 bibliothèques 16S.

Séquençage

Les bibliothèques préparées ont été regroupées et chargées dans une Flow Cell préremplie de NextSeq 1000/2000 P1 Reagents (600 cycles) Kit, une Flow Cell de MiSeq Reagent Kit v3 (600 cycles) (Illumina, référence n° MS-102-3003) ou une Flow Cell de NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, référence n° 20024908). Le séquençage a été effectué sur le NextSeq 2000 System, le MiSeq System ou le NextSeq 550 System, respectivement. Les données de séquençage représentatives pour toutes les analyses sont disponibles sur la page Web des [données de démonstration BaseSpace^{MC} Sequence Hub](#).

Analyse

Les bibliothèques regroupées ont été démultiplexées dans la plateforme infonuagique BaseSpace Sequence Hub. Le pipeline DRAGEN^{MC} Metagenomics a été utilisé pour traiter les données générées sur les systèmes NextSeq 2000, MiSeq et NextSeq 550. Les métagénomiques ont été assemblés à l'aide de SPAdes Genome Assembler. Les classifications taxonomiques ont été effectuées par l'entremise du pipeline DRAGEN Metagenomics.

Pour comparer les données de NextSeq 2000 générées à l'aide d'une trousse de 600 cycles aux données de NextSeq 500 générées à l'aide d'une trousse de 300 cycles, les lectures de NextSeq 2000 ont été retranchées à l'aide de DRAGEN FASTQ Toolkit, disponible sur BaseSpace Sequence Hub.

Pour permettre des comparaisons entre les échantillons, chaque échantillon a été sous-échantillonné au même nombre de lectures (30 millions, 10 millions, 1 million) à l'aide de DRAGEN FASTQ Toolkit. Le sous-échantillonnage est requis dans les cas où seul un sous-ensemble de l'échantillon peut être traité par une application (p. ex. un assemblage *de novo* avec des contraintes de mémoire) ou lorsque l'ensemble complet de données n'est pas nécessaire pour traiter un échantillon (p. ex. pour valider une approche à différents niveaux de couverture génomique).

Résultats

Amélioration des principaux indicateurs de séquence

Les réactifs NextSeq 1000/2000 P1 Reagents (600 cycles) sur le NextSeq 2000 System montrent un pourcentage plus élevé de scores de qualité \geq Q30 par rapport à la MiSeq Reagent Kit v3 (600 cycles) analysée sur le MiSeq System. Les NextSeq 1000/2000 P1 Reagents (600 cycles) fournissent également jusqu'à 100 millions de lectures uniques passant le filtre ou 200 millions de lectures appariées passant le filtre. À environ 60 Gb, les NextSeq 1000/2000 P1 Reagents (600 cycles) génèrent quatre fois plus de données que la MiSeq Reagent Kit v3 (600 cycles) qui en génère environ 15 Gb. En outre, les analyses de séquençage avec la trousse NextSeq 1000/2000 P1 Reagents (600 cycles) ont été effectuées en environ 34 heures, ce qui représente environ 20 heures de moins qu'une analyse de séquençage avec MiSeq Reagent Kit v3 (600 cycles) (figure 2).

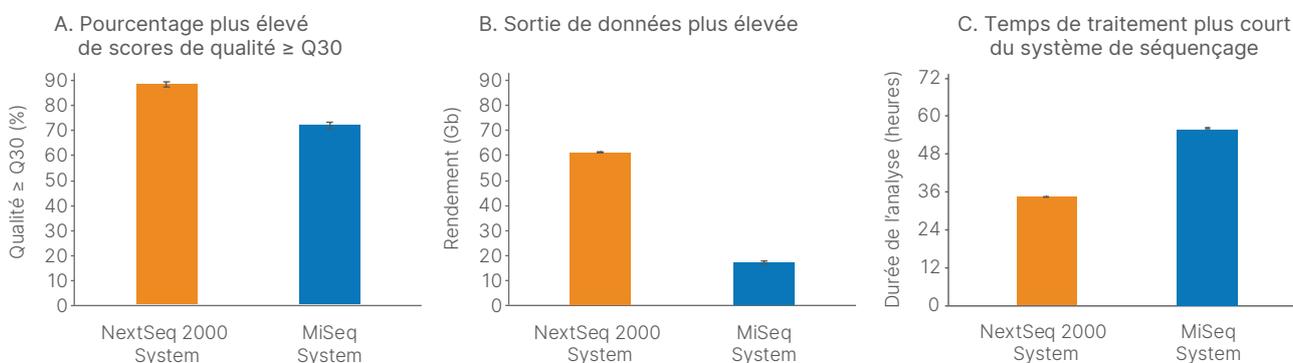


Figure 2 : Comparaisons des principaux indicateurs de performance pour les systèmes NextSeq 2000 et MiSeq : par rapport à la MiSeq Reagent Kit v3 (600 cycles) analysée sur le MiSeq System, les NextSeq 1000/2000 P1 Reagents (600 cycles) du NextSeq 2000 System offrent (A) un pourcentage plus élevé de scores de qualité \geq Q30, (B) une sortie de données quatre fois plus élevée et (C) une durée d'analyse de l'instrument environ 20 heures plus courte lors de l'utilisation de la Flow Cell P1 pour système NextSeq.

NextSeq 500/550 High Output Kit v2.5 (300 cycles) (Illumina, référence n° 20024908) est capable de générer jusqu'à 120 Gb de données de séquençage de haute qualité, à une longueur de lecture de 2 × 150 pb, en environ 29 heures. La spécification de qualité pour cette trousse est supérieure à 75 % des bases ≥ Q30 (données non présentées).

Comparaisons des analyses métagénomiques

Pour comparer les performances entre les systèmes, l'échantillon 20 Strain Staggered Mix Genomic Material a été séquençé sur les systèmes NextSeq 2000, MiSeq et NextSeq 550. L'application DRAGEN Metagenomics sur BaseSpace Sequence Hub a été utilisée pour l'analyse en aval permettant de déterminer les classifications taxonomiques. L'analyse du SNG métagénomique a identifié tous les éléments attendus de la communauté bactérienne simulée et a montré des résultats comparables pour les trois systèmes de séquençage (figure 3).

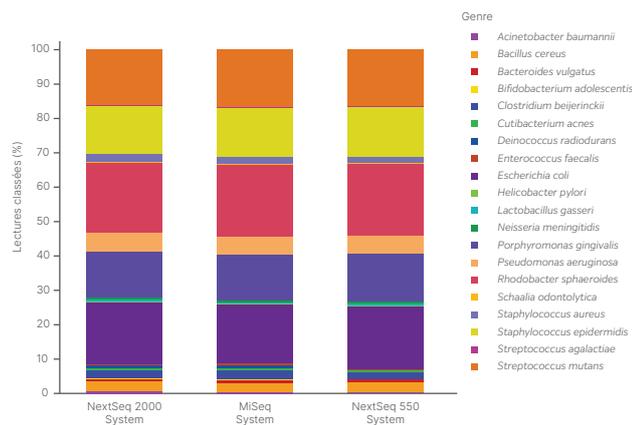


Figure 3 : L'analyse comparative de la composition microbienne révèle la composition du ATCC 20 Strain Staggered Mix analysé avec les systèmes NextSeq 2000, MiSeq et NextSeq 550 : l'analyse de la composition microbienne par l'application DRAGEN Metagenomics démontre une identité et une répartition des genres excellentes et reproductibles.

Bien que l'échantillon ATCC montre des performances hautement reproductibles, un échantillon simulé ne reflète pas nécessairement les performances des échantillons réels. Par conséquent, nous avons testé les performances de détection des organismes avec des échantillons de selles de sujets réels. Les 20 genres les plus représentés dans les échantillons de selles ont été comparés dans les systèmes NextSeq 2000, NextSeq 550 et MiSeq (figure 4). Bien que les résultats ne soient pas aussi uniformes avec ce type d'échantillon plus complexe, les profils de communauté métagénomique de tous les échantillons de selles de

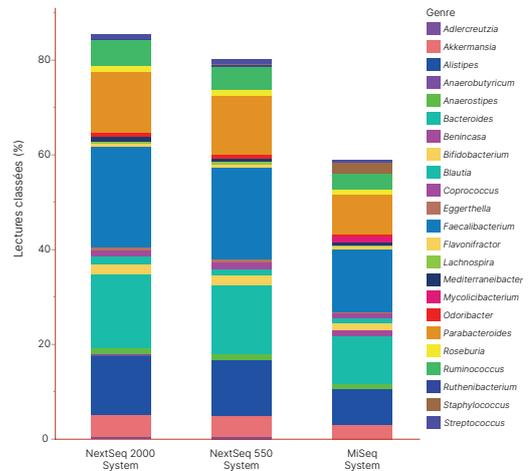


Figure 4 : L'analyse métagénomique révèle une composition similaire de la communauté pour un échantillon représentatif de selles de sujets réels analysé sur plusieurs systèmes : l'application DRAGEN Metagenomics analyse la composition microbienne d'échantillons de selles de sujets réels, limitée aux 20 genres les plus représentés dans l'échantillon, séquençés sur les systèmes NextSeq 2000, NextSeq 550 et MiSeq. Les données démontrent une couverture des genres similaire entre les plateformes, même avec des échantillons complexes.

sujets réels étaient hautement concordants entre les systèmes NextSeq 2000, MiSeq et NextSeq 550. Cela indique que les trois systèmes sont capables d'effectuer un séquençage métagénomique fiable de divers types d'échantillons.

Une longueur de lecture plus longue améliore la caractérisation des échantillons

Bien que les Flow Cell à 300 cycles soient en mesure de fournir des données de séquençage métagénomique significatives, les trousse de 600 cycles offrent des avantages lorsqu'il s'agit d'identifier des espèces individuelles dans des échantillons complexes. Pour démontrer l'effet de lectures plus longues sur l'assemblage du métagénome, les lectures du NextSeq 2000 System ont été retranchées de 600 à 300 cycles à l'aide de l'application DRAGEN FASTQ Toolkit. Kraken2, un classificateur taxonomique basé sur les k-mers*, a été utilisé pour déterminer le pourcentage de lectures classées pour chaque échantillon à 600 cycles ou 300 cycles (figure 5). Les données suggèrent que des lectures plus longues apportent des améliorations à la classification taxonomique basée sur les k-mers avec divers échantillons prélevés dans l'environnement.

* La classification taxonomique métagénomique de Kraken2 est disponible via l'application Illumina DRAGEN Metagenomics sur BaseSpace.

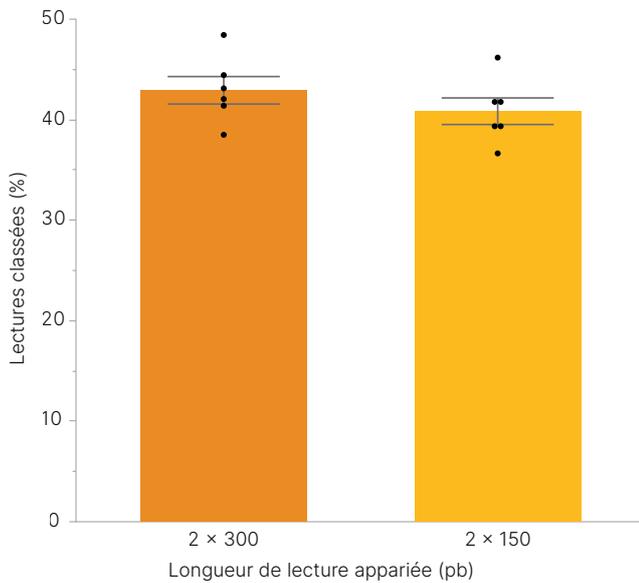


Figure 5 : Une longueur de lecture plus longue améliore la classification des lectures pour les échantillons de selles de sujets réels séquencés sur le NextSeq 2000 : l'application DRAGEN FASTQ Toolkit a été utilisée pour retrancher les lectures du NextSeq 2000 System de 600 cycles (2 x 300 pb) à 300 cycles (2 x 150 pb). Ensuite, Kraken2 a été utilisé pour déterminer le pourcentage de lectures classées pour chaque échantillon. La profondeur de lecture était de 30 millions de lectures pour l'analyse. Les barres d'erreur représentent une erreur type de la moyenne.

À l'aide des mêmes données retranchées, les effets des lectures longues sur la richesse de l'échantillon (c.-à-d. le nombre d'espèces détectées dans un échantillon) et l'indice de Shannon (c.-à-d. la représentation proportionnelle des espèces détectées dans l'échantillon) ont également été examinés. Ces indicateurs montrent que la diversité microbienne observée des échantillons de selles augmente lorsque la longueur de lecture augmente, tandis que la proportion d'espèces détectées quantifiées par l'indice de Shannon reste, comme prévu, relativement inchangée (figure 6).

Une plus grande profondeur de séquençage améliore la caractérisation des échantillons

Ensuite, nous avons examiné l'importance de la profondeur de lecture sur les mesures de la diversité de la population pour les échantillons de selles de sujets réels. Le nombre de lectures du NextSeq 2000 System a été sous-échantillonné à 30 millions, 10 millions et 1 million de lectures à l'aide de DRAGEN FASTQ Toolkit; et la richesse et l'indice de Shannon ont été calculés à l'aide de l'application DRAGEN Metagenomics.

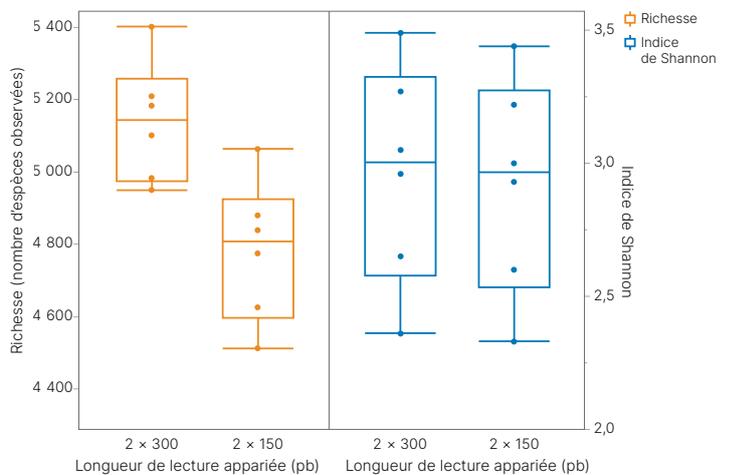


Figure 6 : La richesse microbienne des échantillons de selles augmente avec une longueur de lecture plus longue : l'application DRAGEN FASTQ Toolkit a été utilisée pour retrancher les lectures du NextSeq 2000 System de 600 cycles (2 x 300 pb) à 300 cycles (2 x 150 pb). Ensuite, l'application DRAGEN Metagenomics a été utilisée pour calculer la richesse et l'indice de Shannon afin de quantifier le nombre d'espèces détectées et la diversité proportionnelle, respectivement. La richesse augmente avec des lectures plus longues tandis que l'indice de Shannon reste relativement inchangé. La profondeur de lecture était de 30 millions de lectures pour l'analyse.

Les indicateurs de diversité ont démontré que la diversité microbienne des échantillons de selles augmente lorsque la profondeur de séquençage augmente, tandis que la proportion des espèces indiquée par l'indice de Shannon reste relativement inchangée (figure 7).

Des lectures plus longues améliorent l'identification taxonomique

L'un des défis actuels du profilage de diverses populations microbiennes environnementales est l'absence de génomes de référence complets pour de nombreuses espèces rares et non cultivables. En général, le nombre de contigs assemblés pour des populations microbiennes très diversifiées est plus élevé pour les longueurs de lecture plus longues, comme le montre la longueur totale d'assemblage plus grande (figure 8). Le séquençage métagénomique aléatoire avec des longueurs de lecture de séquençage de 2 x 301 pb d'Illumina améliore généralement l'assemblage *de novo* des métagénomiques à partir d'échantillons prélevés dans l'environnement, contribuant ainsi de manière significative à l'exhaustivité globale de chaque métagénome assemblé.

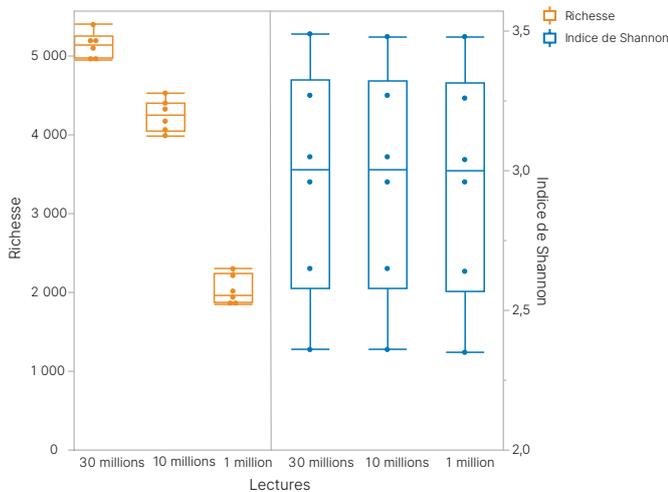


Figure 7 : La richesse microbienne des échantillons de selles augmente à mesure que la profondeur du séquençage augmente : la richesse et l'indice de Shannon ont été calculés à l'aide de l'application DRAGEN Metagenomics pour mesurer le nombre d'espèces détectées et la diversité proportionnelle observée avec le séquençage à 30 millions, 10 millions et 1 million de lectures. Comme prévu, la richesse diminue dans les données sous-échantillonnées, tandis que l'indice de Shannon reste relativement inchangé.

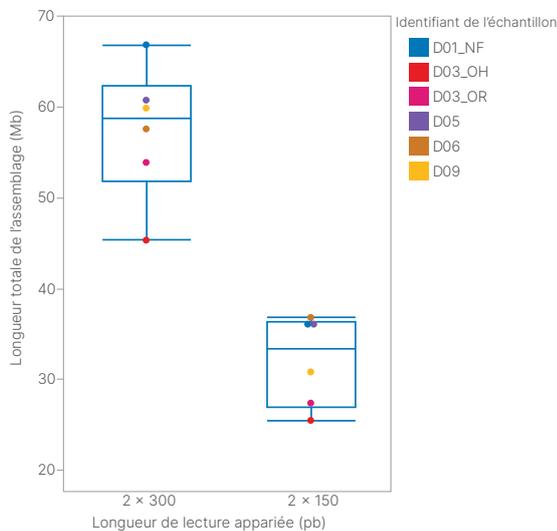


Figure 8 : Une longueur de lecture plus longue permet l'augmentation du nombre global de bases assemblées pour les échantillons de selles de sujets réels séquencés sur le NextSeq 2000 System : les données de séquençage métagénomique du NextSeq 2000 System ont été retranchées à 1 million de lectures et les longueurs de lecture de 2 x 300 pb à 2 x 150 pb à l'aide de l'application DRAGEN FASTQ Toolkit. Pour les comparaisons, SPAdes Genome Assembler a été utilisé pour générer des longueurs de contig totales à partir de lectures de séquençage définies à 2 x 300 pb et 2 x 150 pb.

Résumé

L'objectif de cette note d'application est de démontrer les performances similaires des trousse de 600 cycles sur le NextSeq 2000 System et le MiSeq System. Le NextSeq 550 System est également capable d'effectuer une analyse métagénomique précise, mais ne dispose pas d'une option de 600 cycles. Les trois systèmes ont généré des assemblages *de novo* précis, quel que soit l'instrument utilisé. Des profils métagénomiques concordants, en particulier la diversité proportionnelle observée des genres, ont été obtenus avec les trois systèmes.

Dans l'ensemble, le NextSeq 2000 System offre des avantages par rapport aux systèmes MiSeq et NextSeq 550 en matière de résolution des difficultés associées à la richesse des échantillons provenant d'échantillons diversifiés sans culture en raison de la longueur de lecture et des capacités de profondeur de lecture supplémentaires[†]. Ces avantages sont particulièrement importants pour les échantillons métagénomiques complexes, tels que les échantillons de selles ou prélevés dans l'environnement.

En résumé, les réactifs NextSeq 1000/2000 P1 Reagents (600 cycles) et NextSeq 1000/2000 P2 Reagents (600 cycles) sur les systèmes NextSeq 1000 et NextSeq 2000 offrent un séquençage de haute qualité avec un temps de traitement plus court que la MiSeq Reagent Kit v3 (600 cycles) sur le MiSeq System. La présente note d'application démontre que les trousse de 600 cycles sur les systèmes NextSeq 1000 et NextSeq 2000 conservent des scores Q30 élevés et des taux d'erreur faibles. De plus, les réactifs NextSeq 1000/2000 P1 Reagents (600 cycles) et NextSeq 1000/2000 P2 Reagents (600 cycles) fournissent une sortie de données flexible, idéale pour les expériences de séquençage génomique à petite et moyenne échelle. Enfin, les trousse de 600 cycles pour les systèmes NextSeq 1000 et NextSeq 2000 permettent le développement des applications et la simplicité opérationnelle tout en maintenant la qualité des données établie sur le MiSeq System éprouvé.

[†] Le NextSeq 1000 System est fonctionnellement similaire au NextSeq 2000 System et devrait fournir des résultats similaires à cet instrument en utilisant les mêmes trousse de 600 cycles.

En savoir plus

[Plateformes de séquençage d'Illumina](#)

[Systèmes NextSeq 1000 et NextSeq 2000](#)

[NextSeq 1000/2000 Reagents](#)

[Illumina DNA Prep](#)

[Protocole de lysat brut d'Illumina DNA Prep pour le SNG](#)

Références

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. Publié le 15 juillet 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. Publié le 4 février 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. Publié en 2023. Consulté le 6 février 2024.



Numéro sans frais aux États-Unis : + (1) 800 809-4566 | Téléphone : + (1) 858 202-4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d'Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez la page www.illumina.com/company/legal.html.
M-GL-01147 FRA v1.0