illumina®

Nextera™ crude lysate protocol for metagenomic next-generation sequencing

The Nextera DNA Flex Library Prep Kit with the Nextera Crude Lysate Protocol deliver fast, accurate metagenomic profiling and high-quality *de novo* genome assembly data.

Introduction

Next-generation sequencing (NGS) technology has become an important tool in metagenomics research for organism identification, metagenomic profiling, and *de novo* assembly.[1,2] The Nextera library prep portfolio introduced several workflow advantages that support faster and more efficient library preparation for metagenomic applications (Figure 1). Tagmentation chemistry and the release of Nextera DNA and Nextera XT DNA Library Preparation Kits consolidated DNA fragmentation and adapter tagging steps into a single reaction.[3] With innovative on-bead tagmentation, the Nextera DNA Flex Library Prep Kit further shortened the library prep workflow by integrating DNA extraction, quantitation, fragmentation, and library normalization steps.[4]

Though NGS-based whole-genome sequencing (WGS) has provided significant advantages in speed, accuracy, and depth of information to microbiology labs, DNA extraction and library preparation steps have remained a significant bottleneck in the NGS workflow. Preparation of NGS libraries for metagenomic studies typically begins with time-consuming and labor-intensive genomic DNA extraction steps. To address this challenge in metagenomics, Illumina offers the Nextera DNA Flex Library Preparation Kit with the Nextera Crude Lysate Protocol—a library prep method that supports quick and easy library preparation directly from crude lysate. Sequencing directly from crude lysate eliminates the time and cost associated with DNA extraction steps. In addition to increased speed and efficiency, the Nextera DNA Flex Library Preparation Kit offers exceptional flexibility for sample input type and cell lysis methods, including direct bacterial colonies, blood, and saliva.

This application note demonstrates the performance of the Nextera DNA Flex Library Prep Kit using the Nextera Crude Lysate Protocol using mock microbial communities and real-world human stool samples. This application note also shows how the Nextera DNA Flex Library Prep Kit, with its advanced chemistry and faster workflow, outperforms the original Nextera DNA Library Prep Kit in metagenomic profiling, and *de novo* assembly.*

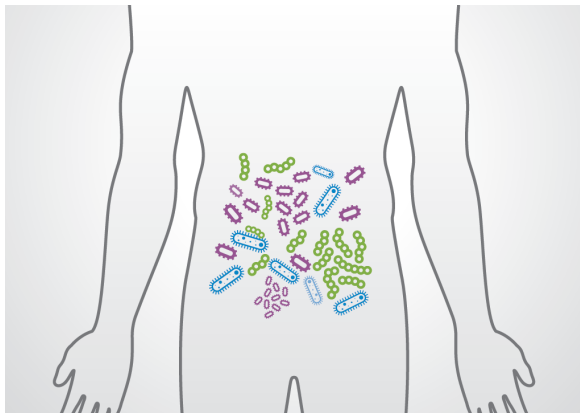

Figure 1: Nextera Crude Lysate Protocol for Metagenomics—The Nextera DNA Flex Kit with the Nextera Crude Lysate Protocol enables library prep directly from crude lysate, resulting in additional time and cost savings while retaining data consistency and quality.

Methods

Sample sources

To represent mock microbial communities, 20 Strain Even Mix Genomic Material (ATCC MSA-1002), 20 Strain Staggered Mix Genomic Material (ATCC MSA-1003), and 20 Strain Even Mix Whole Cell Material (ATCC MSA-2002) were used. The 20 Strain Even/Staggered Mix Genomic Material mixtures are composed of purified DNA extracted from 20 bacterial strains, while the 20 Strain Even Mix Whole Cell Material is a mixture of whole (unlysed) cells. The species in these mixtures were selected for their diverse genome size, range of GC content, and Gram stain profile. Additionally, the microorganisms in the 20 strain mixtures have fully sequenced and characterized genomes. To represent real-world metagenomic samples, stool samples were obtained from both healthy donors and patients undergoing drug treatment.

*The original Nextera DNA Library Prep and Nextera XT DNA Library Prep Kits use tagmentation chemistry, while the Nextera DNA Flex Library Prep Kits use advanced on-bead tagmentation chemistry. On-bead tagmentation chemistry delivers more uniform library yield, consistent insert sizes, and uniform genome coverage compared to the original tagmentation chemistry.

Crude lysate and DNA extraction from mock microbial communities

To prepare crude lysate, 200 µl from the 20 Strain Even Mix Whole Cell sample were processed as described in the Nextera Crude Lysate Protocol with the PureLink Microbiome DNA Purification Kit (Catalog No. A29790, ThermoFisher). Supernatant from step H, which includes the transfer of supernatant from homogenized cells to a clean tube, was used as crude lysate.[5] To prepare extracted DNA, an equivalent number of cells were processed with the PureLink Microbiome DNA Purification Kit using the full protocol. The Nextera Crude Lysate Protocol is compatible with a variety of bacterial DNA extraction kits (Table 1).

Crude lysate and DNA extraction from stool samples

Stool samples from donor patients were used to generate crude lysates and extracted DNA as described in the Nextera Crude Lysate Protocol from Stool Samples with the PureLink Microbiome DNA Purification Kit. 0.05 g of stool samples were processed with the PureLink Microbiome DNA Purification Kit. Supernatant from step G, which includes the transfer of supernatant from homogenized cells to a clean tube, was used as crude lysate.[6] To prepare extracted DNA, an equivalent amount of starting material was processed with the PureLink Microbiome DNA Purification Kit using the full protocol. Stool samples can be highly inhibitory due to the presence of PCR inhibitors such as bile, salts, and polysaccharides. Inhibited library preps show an inability to pellet during clean-up steps or they may deliver low library yield. Decreasing the amount of lysate below the recommended amount may be needed for some highly inhibitory samples. The Nextera Crude Lysate Protocol is compatible with a variety of stool sample purification kits (Table 1).

Table 1: DNA extraction and lysate prep kits[a]

Kits for bacterial samples
PureLink Microbiome DNA Purification Kit (ThermoFisher)
UltraClean Microbial DNA Isolation Kit (MOBIO)
ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher)
Kits for stool samples
PureLink Microbiome DNA purification kit (Invitrogen)
PowerSoil DNA Isolation Kit (MOBIO)
PowerFecal DNA Isolation Kit (MOBIO)
QIAamp DNA Stool mini kit(Qiagen)

a. For detailed information about recommended collection and lysis methods to obtain crude lysate with these kits, and the recommended input amount of crude lysate, download the Nextera Crude Lysate Protocol Guide.

Library preparation and sequencing

To test the accuracy and sensitivity of the Nextera DNA Flex Library Prep Kit with species identification and metagenomic profiling, 10 ng of DNA from the 20 Strain Staggered Mix Genomic Material sample was added directly to the Nextera DNA Flex Library Prep Kit (Catalog No. 20018704, Illumina) and prepared using the standard protocol (extracted DNA).

To compare the performance of crude lysate to extracted DNA, sequencing libraries were prepared with 5 µl lysate or 5 µl extracted DNA from the 20 Strain Even Mix Whole Cell sample or from stool samples. The lysate and extracted DNA samples were used with the Nextera DNA Flex Library Preparation Kit using either the Nextera Crude Lysate Protocol or the standard protocol. Library preparation using the Nextera Crude Lysate Protocol significantly reduced the total amount of time and the number of touch points in the library preparation workflow (Figure 2). By using crude lysate as a direct input into the Nextera DNA Flex Library Preparation Kit, the need for costly, time-consuming DNA quantitation kits are also eliminated.

For a comparison of the Nextera DNA Flex Library Preparation Kit and the original Nextera DNA Library Preparation Kit (Catalog No. FC-121-1030, Illumina), 50 ng DNA from the 20 Strain Even Mix Genomic Material sample and 50 ng DNA from the 20 Strain Staggered Mix Genomic Material sample were added directly to the library prep kits. Libraries were prepared using the standard protocol.

All libraries were sequenced on the NextSeq™ 550 Sequencing System with a run configuration of 2 × 150 bp paired-end reads using the NextSeq 500/550 High Output v2 Kit (Catalog No. FC-404-2004, Illumina).

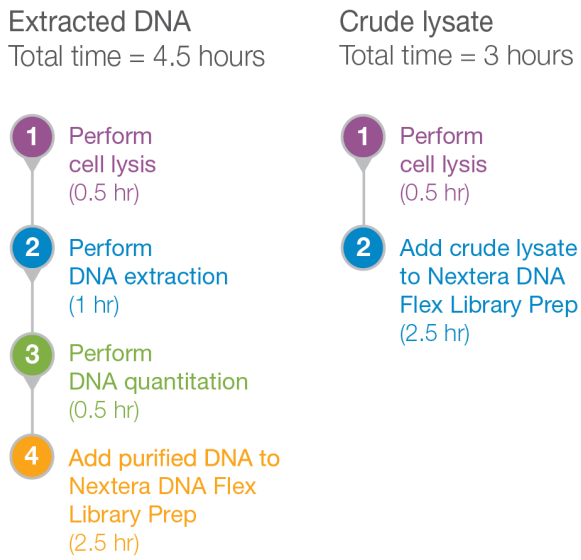

Figure 2: Comparison of extracted DNA and crude lysate workflows—The Nextera Crude Lysate Protocol reduces the amount of time and total number of touch points required in the library preparation protocol. The crude lysate protocol also reduces the time and cost associated with additional purification and quantitation steps. Workflow times were calculated assuming specific methods: DNA extraction (PureLink Microbiome DNA Purification Kit), DNA Quantitation (Qubit). Times may vary depending on equipment used, kits used, sample batch number, automation procedures, and user experience.

Data analysis

Metagenomic profiling metrics, including true positives, relative abundance, false positives, observed abundance, and expected abundance, were calculated on the One Codex[7] platform. Genome assembly bar graphs were generated with two software tools. First MEGAHIT v1.1.1.[8] performed *de novo* alignment with the FASTQ reads to generate contigs, then the contigs were assessed with Quality Assessment Tool for Genome Assemblies (QUAST)[9] with the known reference genomes. Metagenomic profiling stacked bar graphs were compiled with GENIUS Metagenomics software.[10] GENIUS Metagenomics software can be accessed in BaseSpace™ Sequence Hub, the Illumina genomics computing platform.

Results

Metagenomic profiling with mock community mixtures and stool samples

To establish the accuracy and sensitivity of the Nextera DNA Flex Library Prep Kit in basic organism identification, a sequencing library was prepared with the 20 Strain Staggered Mix sample. Metagenomic profiling summary metrics—true positive, relative abundance, and false positives—were calculated on the One Codex platform. The results demonstrate excellent metagenomic profiling performance: all 20 organisms in the 20 Strain Staggered Mix were identified with no false positives (Table 2). The composition statistics with the 20 Strain Staggered Mix were also highly accurate with the observed abundance scores closely matching the expected abundance scores for all 20 species (Table 3). Furthermore, the observed abundance scores spanned four orders of magnitude and were able to detect species present in the sample down to 0.018%, which demonstrates high sensitivity in the Nextera DNA Flex sequencing workflow.

Table 2: 20 Strain Staggered Mix metagenomic profiling summary statistics[a,b]

Library	True positives	Relative abundance	False positives
20 Strain Staggered Mix and Nextera DNA Flex Library Prep Kit	100%	100%	0

a. One Codex defines statistics as follows: **True Positives:** The percentage of organisms present in the control. Organisms are marked as "Present" if they are detected within two logs of the true abundance. **Relative Abundance:** The Pearson correlation coefficient between the known input organism abundances and the detected abundances (based on genome-size adjusted read counts). **False Positives:** 100% less 10 percentage points for each "High" abundance false positive, 5 points for each "Moderate" one, and 1 point for each "Low" one. "Trace" false positives do not count against the score, and the minimum possible score is 0%.
b. Data set down sampled from 20 million reads to 1 million reads.

Table 3: 20 Strain Staggered Mix metagenomic profiling composition statistics

Name of organism	Observed abundance	Expected abundance
Rhodobacter sphaeroides	19.07%	18.00%
Escherichia coli	18.54%	18.00%
Staphylococcus epidermidis	18.26%	18.00%
Porphyromonas gingivalis	17.17%	18.00%
Streptococcus mutans	17.61%	18.00%
Pseudomonas aeruginosa	1.93%	1.80%
Clostridium beijerinckii	1.80%	1.80%
Bacillus cereus	1.15%	1.80%
Staphylococcus aureus	1.64%	1.80%
Streptococcus agalactiae	1.77%	1.80%
Acinetobacter baumannii	0.18%	0.18%
Propionibacterium acnes	0.21%	0.18%
Neisseria meningitidis	0.21%	0.18%
Lactobacillus gasseri	0.19%	0.18%
Helicobacter pylori	0.19%	0.18%
Bacteroides vulgatus	0.015%	0.018%
Enterococcus faecalis	0.021%	0.018%
Deinococcus radiodurans	0.017%	0.018%
Bifidobacterium adolescentis	0.013%	0.018%
Actinomyces odontolyticus	0.007%	0.018%

The Nextera DNA Flex Crude Lysate Protocol vs. the standard protocol with mock microbial communities

To compare the performance of the Nextera Crude Lysate Protocol to the standard protocol, sequencing libraries were prepared with crude lysate and extracted DNA from the 20 Strain Even Mix Whole Cell Material sample. Libraries made with crude lysate and extracted DNA generated equivalent metagenomic profiling summary statistics (Table 4).

Table 4: Comparison of crude lysate and extracted DNA metagenomic profiling summary data

Libraries	True positives	Relative abundance	False positives
Crude Lysate	100%	33%	0
Extracted DNA	100%	32%	0

a. Data set down sampled from 20 million reads to 5 million reads.

Using the same crude lysate and extracted DNA data set, the *de novo* genome assembly quality was also compared and the percentage of genome fraction assembled was calculated with QUAST. In general, a higher fraction of genome assembled is indicative of a higher-quality genome assembly. For all 20 organisms analyzed, the crude lysate and extracted DNA libraries generated very similar genome assembly results (Figure 3).

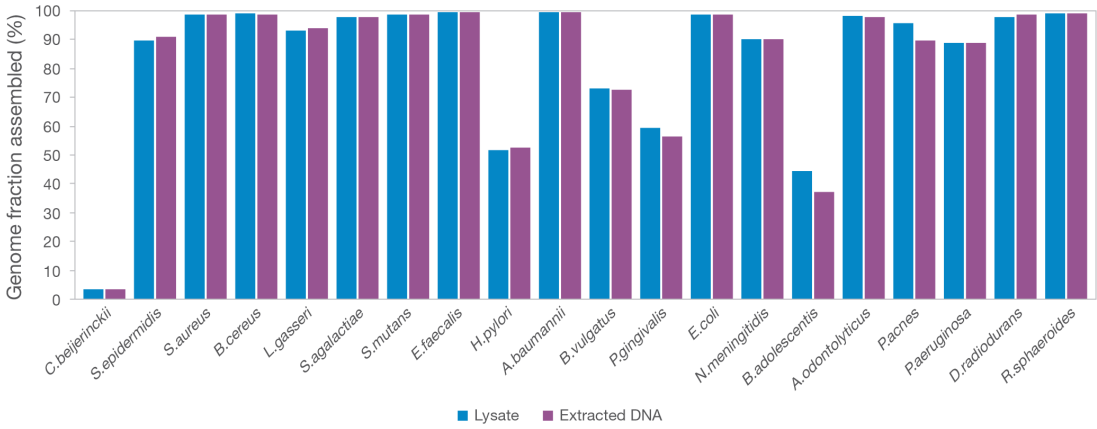

Figure 3: Comparison of crude lysate and extracted DNA genome assembly—Assembly was performed with MEGAHIT using 5 million, paired-end, 150 bp reads. Bar graph illustrates genome assembly performance for 20 organisms as reported by QUAST.

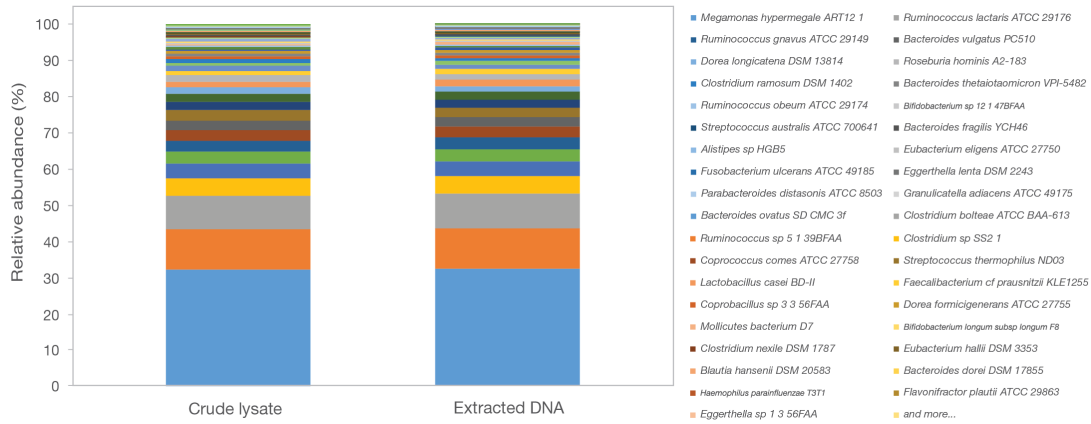

Figure 4: Comparison of crude lysate and extracted DNA metagenomic profiles—–Bar graph illustrates the abundance of each organism identified by GENIUS metagenomics analysis (using 3 million, paired-end, 150 bp reads) for the > 60 organisms identified in both stool libraries.

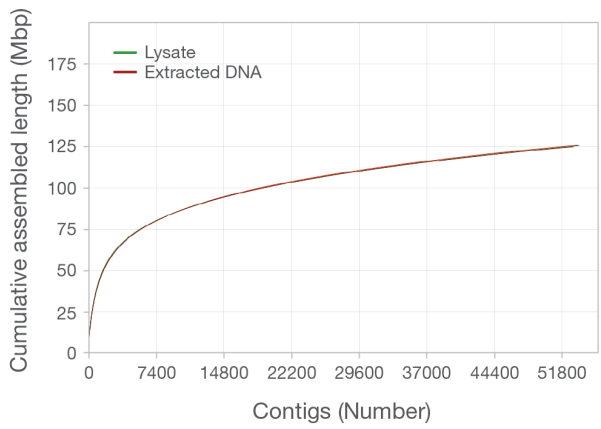

Figure 5: Comparison of crude lysate and extracted DNA genome assembly with unknown species—Graph illustrates cumulative assembled length (Mbp) with the contigs ordered from largest to smallest. 20 million reads were used for assembly.

The Nextera DNA Flex crude lysate protocol vs. the standard protocol with stool samples

The metagenomic profiles resulting from the Nextera Crude Lysate Protocol and the standard Nextera DNA Flex Library Prep protocol are very similar even with highly inhibitory, challenging stool samples (Figure 4). Unlike the mock microbial community samples, the stool samples contain a mixture of unknown species and an unknown species composition (percentage of each species). For the *de novo* assembly of metagenomics samples with unknown species, cumulative assembled length can be used to assess the quality of genome assembly. Assembly using QUAST revealed that the cumulative assembled lengths are highly similar between libraries prepared with the Nextera Crude Lysate Protocol from Stool Samples and the standard Nextera DNA Flex Library Prep protocol (Figure 5).

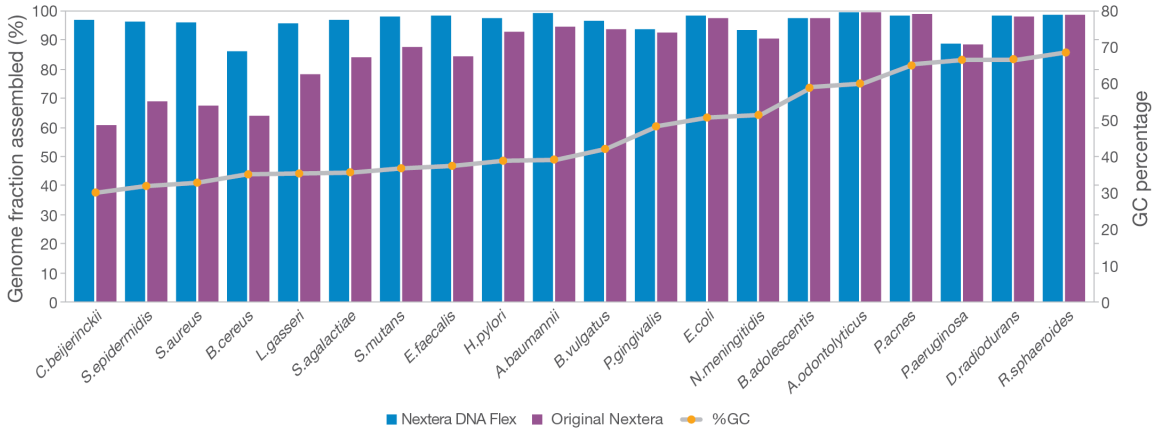

Figure 6: Comparison of Nextera DNA Flex and original Nextera genome assembly metrics—Bar graph illustrates genome fraction assembled with MEGAHIT (using 2 million, paired-end, 150 bp reads) for all 20 organisms. Organisms are listed in the order of increasing GC% from left to right.

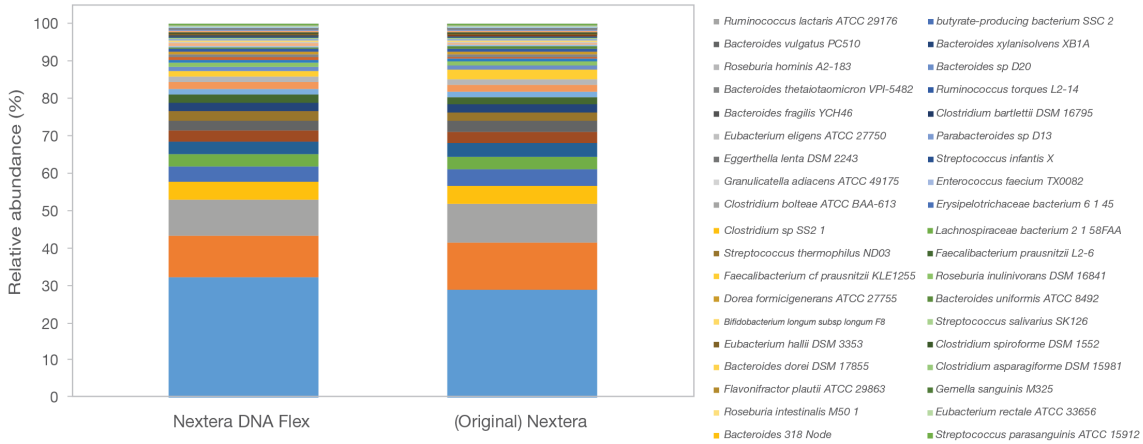

Figure 7: Comparison of Nextera DNA Flex and original Nextera DNA metagenomic profiles—Bar graph illustrates the abundance of each organism identified by GENIUS metagenomics analysis (3 million, paired-end, 150 bp reads were used for analysis) for the > 60 organisms identified in both libraries.

Nextera DNA Flex outperforms the original Nextera DNA Library Prep Kit for metagenomic profiling

The profiling performance of the Nextera DNA Flex Library Prep Kit was compared with the original Nextera DNA Library Prep Kit with libraries prepared from the 20 Strain Even Mix and the 20 Strain Staggered Mix Genomic Material samples. Summary metrics were calculated on the One Codex platform and the results demonstrate that the Nextera DNA Flex Library Prep Kit delivers superior metagenomic profiling results compared to the original Nextera DNA Library Prep Kit (Table 5). Unlike Nextera DNA and Nextera XT Library Prep Kits, Nextera DNA Flex Library Prep Kits use advanced on-bead tagmentation, which delivers uniform library yield, consistent insert sizes, and uniform genome coverage. These advantages work together to produce improved metagenomic profiling.

Nextera DNA Flex enables superior genome assembly compared to the original Nextera DNA Library Prep Kit

Beyond assessments of organism detection and metagenomic profiling, the sequencing data were analyzed to compare genome assembly metrics. To evaluate genome assembly quality, the genome fraction assembly metrics were calculated by QUAST. The QUAST data show that Nextera DNA Flex outperformed the original Nextera DNA Kit for most of the 20 organisms in the 20 Strain Even Mix Whole Cell Material sample, especially for the AT rich organisms (Figure 6).

Table 5: Comparison of Nextera DNA Flex Library Prep and the original Nextera DNA Library Prep metagenomic profiling summary data

	True positives	Relative abundance	False positives
20 Strain Even Mix Genomic Material			
Nextera DNA Flex Library Prep Kit (50 ng)	100%	97%	0
Nextera DNA Library Prep Kit (50 ng)	100%	83%	0
20 Strain Staggered Mix Genomic Material			
Nextera DNA Flex Library Prep Kit (50 ng)	100%	100%	0
Nextera DNA Library Prep Kit (50 ng)	100%	96%	0

To test the performance of the Nextera DNA Flex Library Prep Kit compared to the (original) Nextera DNA Library Prep Kit with real-world metagenomic samples, crude lysate and extracted DNA libraries were prepared from stool samples. The metagenomics profiles resulting from the two library prep kits are very similar, indicating that both provide excellent and comparable profile results (Figure 7).

Summary

The Nextera DNA Flex Library Prep Kit with the Nextera Crude Lysate Protocol provide a fast and easy workflow that eliminates costly, time-consuming DNA extraction steps and delivers exceptional data quality. The Nextera Crude Lysate Protocol offers high sensitivity and accuracy for organism detection and metagenomic profiling, as well as high-quality *de novo* genome assembly comparable to extracted DNA. With the ability to support complex metagenomic mixtures such as stool or mock microbial community samples with the Nextera Crude Lysate Protocol, as well as saliva, blood, and direct bacterial colonies with additional demonstrated protocols, the Nextera DNA Flex Library Prep Kit delivers a flexible, cost-effective approach for NGS-based metagenomic research.

Learn more

To learn more about the Nextera DNA Flex Library Prep Kit, visit www.illumina.com/nextera-dna-flex

For more on direct bacterial colony sequencing with Nextera DNA Flex, download www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/nextera-dna-flex-direct-colony-app-note-770-2017-036.pdf

For more on microbial genome sequencing with the Nextera DNA Flex, visit www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/nextera-dna-flex-small-genomes-application-note-770-2017-019.pdf

Ordering Information

Product	Catalog No.
Nextera DNA Flex Library Prep Kit (24 samples)	20018704
Nextera DNA Flex Library Prep Kit (96 samples)	20018705
Flex Lysis Reagent Kit	20018706
Nextera DNA CD Indexes (24 indexes, 24 samples)	20018707
Nextera DNA CD Indexes (96 indexes, 96 samples)	20018708

CD Indexes: Combinatorial Dual Indexes. 24 dual indexes provided to support up to 24 samples or 96 dual indexes provided to support up to 96 samples.
Single Indexes: 24 single indexes provided to support up to 96 samples.

References

1. Ranjan R, Asha RA, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*. 2016;469:967–977.
2. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550:61-66.
3. Illumina (2016). Nextera DNA Library Preparation Kit Data Sheet. Accessed April 10, 2018.
4. Illumina (2017). Nextera DNA Flex Library Preparation Kit Data Sheet. Accessed April 10, 2018.
5. Invitrogen (2015). PureLink Microbiome DNA Purification Kit User Guide for Microbial Culture. Accessed April 12, 2018.
6. Invitrogen (2015). PureLink Microbiome DNA Purification Kit User Guide for Stool Samples. Accessed April 12, 2018.
7. One Codex platform. www.onecodex.com. Accessed April 12, 2018.
8. MEGAHIT v1.1.1. github.com/voutcn/MEGAHIT. Accessed April 12, 2018.
9. QUAST. quast.sourceforge.net/quast.html. Accessed April 13, 2018.
10. GENIUS Metagenomics. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/cosmosid-genius-metagenomics-know-now.html. Accessed April 12, 2018.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com