An automated Nextera™ DNA Flex library preparation workflow for high-throughput metagenomics

The streamlined, automated workflow generates highly uniform libraries and provides excellent data for species identification, metagenomic profiling, and *de novo* genome assembly.

Introduction

In recent years, characterization of the human microbiome and its role in human health has gained increased attention. Bacterial flora in the gut may influence immune system responses, prevent certain disease states, or alleviate some health conditions.[1] A number of chronic diseases, such as allergies and obesity, have been linked to the composition of the microbiome.[1,2] The ability to detect cultivable and uncultivable bacteria and to examine the presence or absence of many bacterial species in the human microbiome has been greatly improved through the use of next-generation sequencing (NGS) technologies.[1]

Though NGS has provided significant advantages in speed, accuracy, and depth of information to microbiology labs, library preparation can become a bottleneck for high-throughput laboratories (Figure 1). To address this challenge, Illumina collaborated with PerkinElmer to offer the Automated Nextera DNA Flex Library Preparation Workflow for Metagenomics: a comprehensive NGS solution that supports fully automated DNA extraction through DNA analysis (Figure 2). The preparation of Nextera DNA Flex libraries on a liquid handling system offers significant advantages over manual sample preparation. These include higher throughput and scalability, reduction in human touch-points and human error, higher workflow consistency, and increased speed.

The Automated Nextera DNA Flex Library Preparation Workflow for Metagenomics includes automated DNA extraction from stool samples using the chemagic 360 instrument (PerkinElmer) and the chemagic DNA Stool Kit Special (PerkinElmer). DNA extraction is followed by library preparation on the Sciclone G3 NGSx Workstation (PerkinElmer) liquid handler using Nextera DNA Flex Library Preparation Kits (Illumina). Nextera DNA Flex Library Prep Kits feature innovative, on-bead tagmentation chemistry that supports quick and easy library preparation from variety of organisms and specimens.[3] The kit is compatible with a wide DNA input range (100-500 ng), which eliminates the need for accurate quantitation of the initial DNA sample, saving time and costs associated with library input normalization.[3]

The Automated Nextera DNA Flex Library Preparation Workflow for Metagenomics, from DNA extraction to final library pool quantitation, delivers up to 96 ready-to-sequence metagenomic libraries in just under six hours. This application note demonstrates the performance of the Automated Nextera DNA Flex Library Preparation Workflow for Metagenomics in comparison to the standard, manual workflow using stool samples from four human subjects.

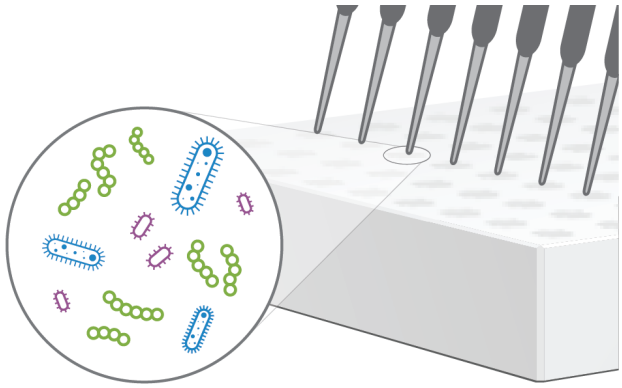

Figure 1: High-throughput metagenomics–Metagenomics labs profile entire microbial communities from complex samples, identify new species, and explore the connection between microbial communities and human health. High-throughput metagenomics labs may process hundreds of samples per week and must be careful to address bottlenecks in their sequencing operations.

Methods

Stool collection

Stool samples were collected from four donors: two adults on a Western diet and two children (twins) on a vegetarian diet. Before DNA isolation, stool samples were stored at 4°C for 20 hours.

Factors such as storage temperature, time, presence or absence of stabilization buffer, and choice of storage tubes can have a significant impact on DNA integrity. In this study, Illumina stored the stool samples in Omega StableGUT Collection Device Tubes (Catalog No. AC7005, Omega Bio-Tek, Inc.), according to manufacturer protocol, at 4°C.

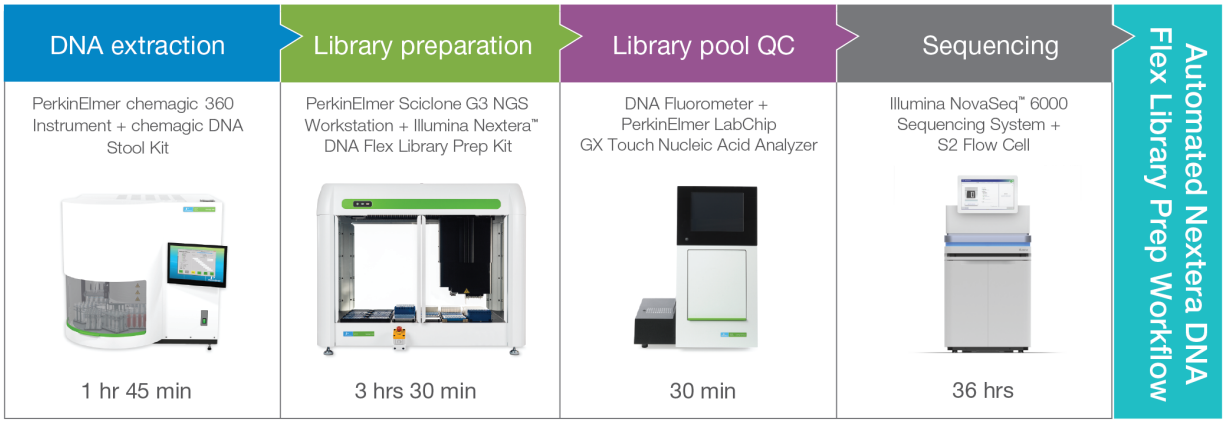

Figure 2: The Automated Nextera DNA Flex Library Preparation Workflow for Metagenomics–Illumina and PerkinElmer have collaborated to create a comprehensive, automated NGS library preparation workflow for high-throughput metagenomics.

DNA extraction

Extractions were performed on the chemagic 360 Instrument (Catalog No. 2024-0020, PerkinElmer)[4] using the chemagic DNA Stool Kit Special (Catalog No. CMG-1076, PerkinElmer). Each isolation was performed with 150 µl elution volume, which produced a total of 300 ng-3 µg of purified DNA. The extraction method was optimized to produce ≥ 100 ng DNA in a total volume of 30 µl, which is the maximum Nextera DNA Flex Library Prep Kit input volume. Integrity of extracted DNA was assessed with the LabChip GX Touch Nucleic Acid Analyzer (Catalog No. CLS137031, PerkinElmer),[5] the HT DNA NGS 3K Reagent Kit (Catalog No. CLS960013, PerkinElmer), and the Genomic DNA Reagent Kit (Catalog No. CLS760685, PerkinElmer). The chemagic method provides optimal isolation of DNA from both Gram negative and Gram positive species.

Automated and manual library preparation

90 Nextera DNA Flex libraries were prepared from two independent automation runs on the Sciclone G3 NGSx Workstation liquid handler (Catalog No. CLS145321, PerkinElmer)[6] and Nextera DNA Flex Library Prep Kits (Catalog No. 20018705, Illumina). The total DNA input range (100-600 ng), overlapped with the recommended DNA input range for Nextera DNA Flex libraries (100-500 ng). For the automated library preps, a fixed volume of 30 µl chemagic purified DNA was used, ensuring ≥ 100 ng DNA input per library. To compare the performance of the Sciclone G3 NGSx script with the Nextera DNA Flex manual protocol, a sub-set of 42 libraries from the same DNA isolates were prepared manually according to the standard protocol.

Sequencing

To generate sufficient genomic coverage for in-depth metagenomic analysis, 48 Nextera DNA Flex libraries were pooled by volume (5 µl each). The pooled libraries were assessed with the LabChip GX Touch Nucleic Acid Analyzer using a HT DNA NGS 3K Reagent Kit and yield was measured with a DNA fluorometer. Libraries were sequenced on the NovaSeq™ 6000 Sequencing System (Catalog No. 20012850) using a NovaSeq 6000 S2 Reagent Kit (Catalog No. 20012860, Illumina) with a run configuration of 2 × 150 bp.

Data analysis

Index representation plots were generated in BaseSpace™ Sequence Hub, the Illumina genomics computing platform. Metagenomic profiling stacked bar graphs were compiled with CosmosID Metagenomics[7] and Kraken Metagenomics[8] Apps using sequencing data sets down-sampled to as low as 3 million reads and up to 80 million reads. CosmosID Metagenomics and Kraken Metagenomics Apps can be freely accessed in BaseSpace Sequence Hub (Figure 3). *De novo* genome assembly quality was evaluated using MEGAHIT v1.1.1.[9] and QUAST v4.4[10] using data sets down sampled to 40 and 60 paired-end reads.

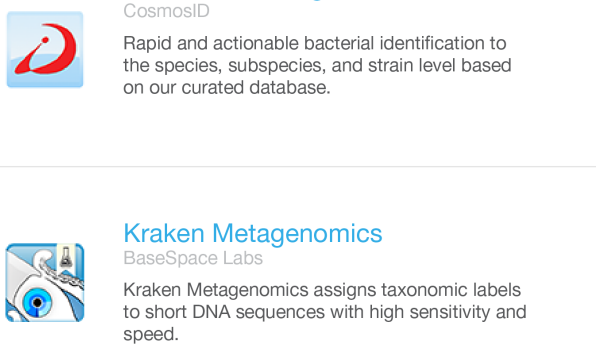

Figure 3: Metagenomics Apps in BaseSpace Sequence Hub–A wide range of metagenomics analysis apps are available in BaseSpace Sequence Hub, including CosmosID, Kraken, Prokka, QIIME, and more.

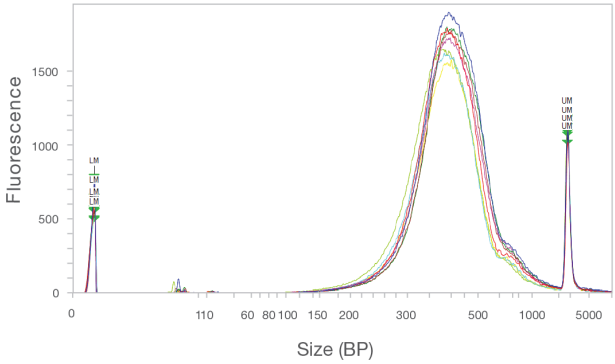

Figure 4: Insert size distribution of eight libraries prepared with automated workflow–The Nextera DNA Flex workflow delivers a highly uniform insert size distribution, resulting in more uniform genome coverage and data accuracy. The graph illustrates an overlay of eight traces from the LabChip GX Touch Nucleic Acid Analyzer representing eight different libraries prepared with the Automated Nextera DNA Flex Library Prep Workflow.

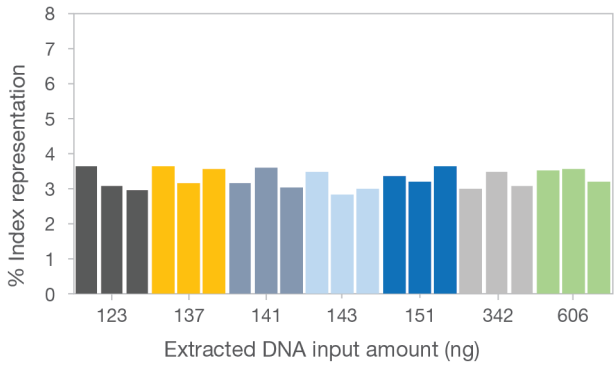

Figure 5: Index representation of libraries prepared from various amounts of extracted DNA with automated workflow–The Nextera DNA Flex workflow produced consistent and uniform index representation across a broad range of DNA inputs. Libraries were prepared with seven different extracted DNA input amounts, pooled together by volume, and sequenced in triplicate.

Results

Automated library preparation delivers uniform insert size distribution and index representation

The ability to use a wide DNA input range while maintaining consistent, uniform insert size and library yield is one of the main advantages of Nextera DNA Flex chemistry. Greater uniformity in insert size distribution and library yield assures more uniform genome coverage and data accuracy. To assess insert size distribution, eight libraries prepared with the Automated Nextera DNA Flex Library Preparation Workflow were analyzed using the LabChip GX Touch Nucleic Acid Analyzer. The libraries represent two DNA isolation replicates from donor 1 (adult) and two DNA isolation replicates from donor 3 (child). Two sequencing library replicates were generated from each DNA isolate producing a total of eight libraries. An overlay of the eight LabChip traces demonstrates highly uniform insert sizes (Figure 4).

To further evaluate the consistency of the Automated Nextera DNA Flex Workflow, a series of libraries were prepared with seven different extracted DNA input amounts in triplicate. To evaluate the yield of the automated library preparation, the percentage of reads identified (passing filter) were plotted for each library in the sequenced pool of 21 libraries (Figure 5). High uniformity of index representation indicates uniform library yields and also demonstrates that each library is evenly represented on the flow cell. The Automated Nextera DNA Flex Library Preparation Workflow produced libraries with highly uniform index representation, even with a range of extracted DNA inputs.

Automated and manual Nextera DNA Flex protocols generate comparable metagenomic profiling results

To assess the performance of automated and manually prepared Nextera DNA Flex libraries in metagenomic profiling, manually prepared libraries and automated libraries were sequenced and

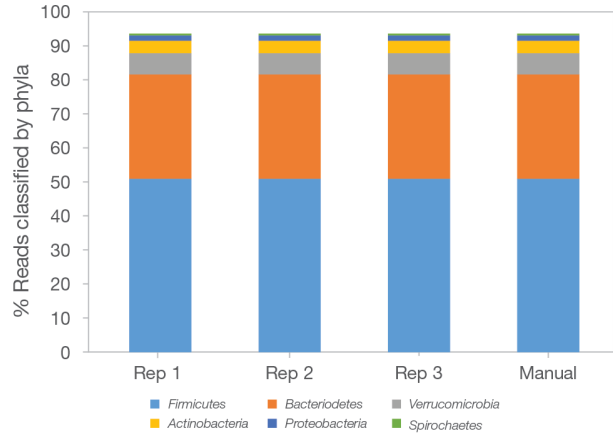

Figure 6: Comparison of automated and manually prepared libraries phyla distribution–Analysis of automated sequencing libraries compared to a manually prepared library with the Kraken Metagenomics App resulted in highly concordant bacterial phyla identification and distribution. Library replicates were produced from the donor 1 sample.

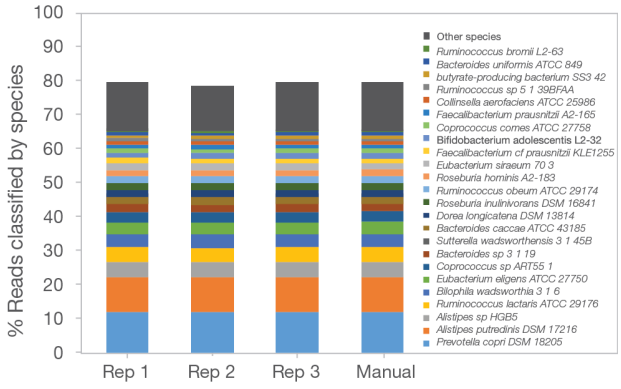

Figure 7: Comparison of automated and manually prepared libraries species distribution–Analysis of automated sequencing libraries compared to a manually prepared library with the CosmosID Metagenomics App resulted in highly concordant bacterial species identification and distribution.

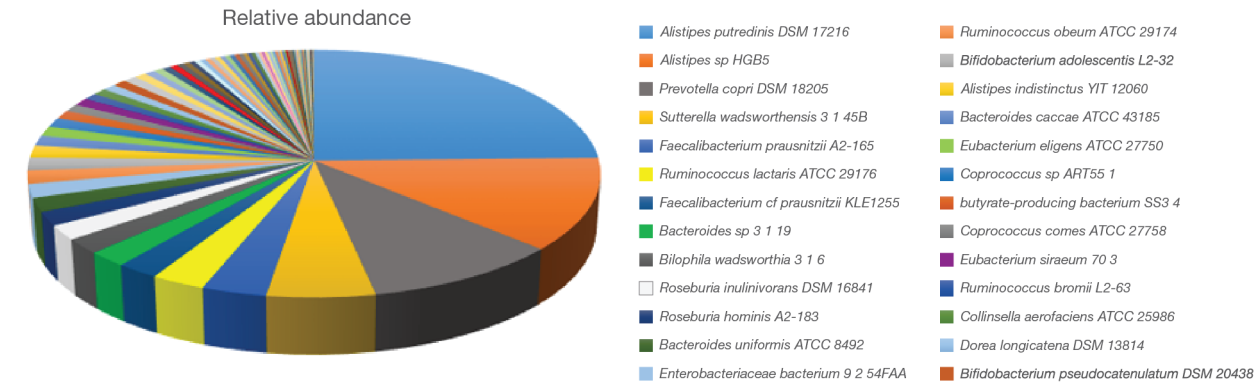

Figure 8: Automated libraries produce rich metagenomic profiles–Libraries were prepared from the donor 2 (adult) sample using the Automated Nextera DNA Flex Library Prep Workflow. The CosmosID Metagenomics App was used with 40 million reads to assemble the relative abundance pie chart and identify over 100 species (only 26 of the > 100 identified species were included in the figure legend).

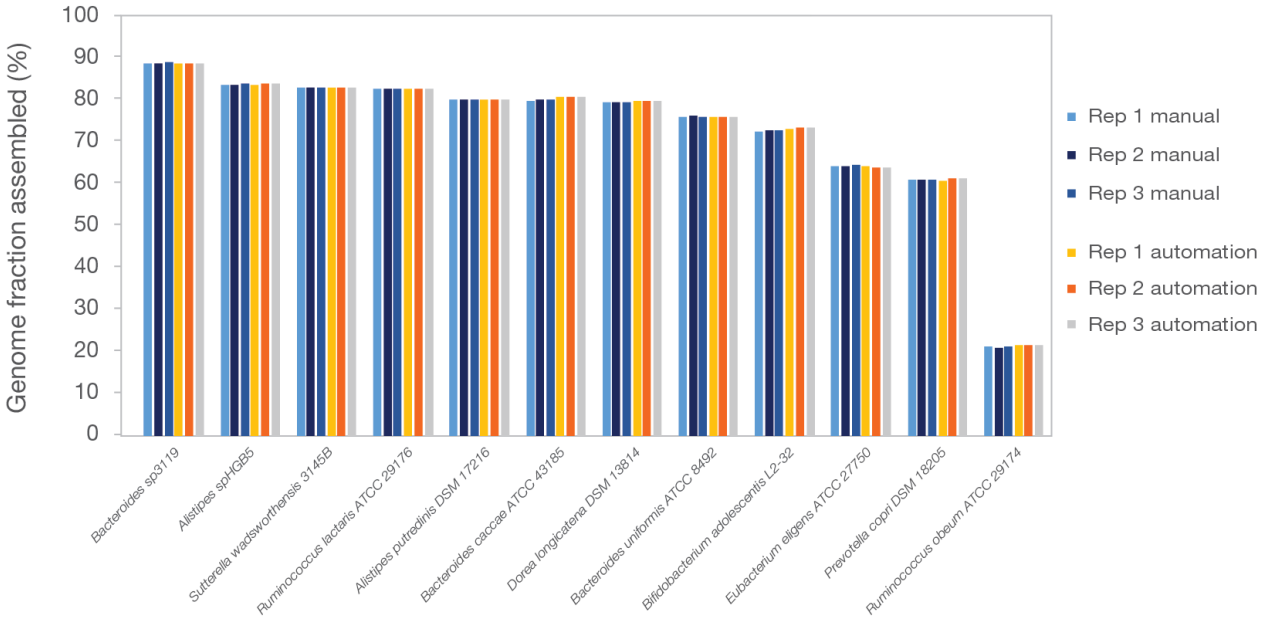

Figure 9: Comparison of automated and manually prepared libraries with genome assembly–*De novo* genome assembly of 12 microorganisms was performed with QUAST using 60 million reads. Libraries were prepared from donor 1 sample in triplicate using the automated and manual protocols.

analyzed with Kraken and CosmosID Metagenomics Apps (Figure 6, Figure 7). The libraries prepared with the Automated Nextera DNA Flex Library Preparation Workflow share the same distribution of bacterial phyla and species as the manually prepared libraries. Furthermore, the automated workflow enabled the identification of over 100 species in the donor 1 sample (Figure 8).

Automated and manual Nextera DNA Flex protocols generate comparable, high-quality genome assemblies

Using the same data set from the automated and manually prepared libraries, the percentage of genome fraction assembled was calculated with QUAST. In general, a higher fraction of genome assembled indicates a higher quality genome assembly. However, the percent of genome assembled also depends on the degree of similarity between the genome of a particular species in the sample and the available reference genome; in some cases, the available reference genome may not be an exact match. In this study, for all 12 organisms analyzed, the automated and manually prepared libraries generated nearly identical genome assembly results (Figure 8).

Summary

The Automated Nextera DNA Flex Library Prep Workflow is an excellent solution for high-throughput metagenomics laboratories. In less than six hours, the automated workflow can perform up to 96 DNA extractions using the chemagic 360 instrument (PerkinElmer) and up to 96 libraries can be prepared with the Sciclone G3 NGSx Workstation (PerkinElmer) and the Nextera DNA Flex Library Prep Kit (Illumina). The automated workflow generates highly uniform libraries and provides excellent data for species identification and metagenomic profiling of complex microbial mixtures–even from challenging stool samples. With significant advantages including higher library consistency, fewer human touch points, and greater scalability, the Automated Nextera DNA Flex Library Prep Workflow is an ideal library prep solution for labs seeking to scale up their operations and harness the power of next-generation sequencing.

Learn More

To learn more about the Nextera DNA Flex Library Prep Kit, visit the Nextera DNA Flex Library Prep page

For information on microbial genome sequencing with the Nextera DNA Flex Library Prep Kit read the Microbial WGS with Nextera DNA Flex Application Note

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.

References

1. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Ther Adv Gastroenterol*. 2013;6:295-308.
2. Clarke SF, Murphy EF, Nilaweera K, et. al. The gut microbiota and its relationship to diet and obesity. *Gut Microbes*. 2012;3:186-202.
3. Illumina (2017). Nextera DNA Flex Library Preparation Kit Data Sheet. Accessed April 10, 2018.
4. Chemagic 360 instrument, PerkinElmer (2016). Compact, High Volume, High Throughput Nucleic Acid Isolation. Accessed April 17, 2018.
5. LabChip GX Touch, PerkinElmer (2016). Automated, High Performance Electrophoresis for Genomics. Accessed April 17, 2018.
6. Sciclone G3 NGSx Workstation, PerkinElmer (2017). Sciclone G3 NGSx Workstation for High Throughput Sequencing Sample Prep Applications. Accessed April 17, 2018.
7. CosmosID Metagenomics. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/cosmosid-CosmosID-metagenomics-know-now.html. Accessed April 12, 2018.
8. Kraken Metagenomics. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/kraken-metagenomics.html. Accessed April 17, 2018.
9. MEGAHIT v1.1.1. github.com/voutcn/MEGAHIT. Accessed April 12, 2018.
10. QUAST v4.4. quast.sourceforge.net/quast.html. Accessed April 13, 2018.

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