APPLICATION NOTE

Automated Illumina DNA Prep workflow for high-throughput metagenomics

Highly uniform libraries and excellent data for species identification, metagenomic profiling, and *de novo* genome assembly

In collaboration with

revvity

illumina®

Introduction

In recent years, characterization of the human microbiome and its role in human health has gained increased attention. Variation in bacterial flora in the gut may influence immune system responses, disease states, and other human health conditions.[1] In fact, several chronic diseases, such as allergies and obesity, have been linked to the composition of the microbiome.[1,2] At the same time, the ability to detect the presence or absence of many bacterial species in the human microbiome has been greatly improved by the use of next-generation sequencing (NGS) technologies.[2]

While NGS-based whole-genome sequencing (WGS) has provided significant advantages in speed, accuracy, and depth of information to microbiology labs, library preparation can become a bottleneck for high-throughput laboratories. For researchers processing hundreds of metagenomic samples per week, Illumina collaborated with Revvity (formerly, PerkinElmer) to offer the automated Illumina DNA Prep workflow for metagenomics. This comprehensive NGS solution supports fully automated workflow, from DNA extraction through DNA analysis (Figure 1). Preparing Illumina DNA Prep libraries using a liquid-handling system offers significant advantages over manual sample preparation, including higher throughput, reduction in touchpoints, reduced chances for human error, greater workflow consistency, reduced labor cost, and increased speed.

The automated Illumina DNA Prep workflow for metagenomics, from DNA extraction to final library pool quantification, delivers up to 96 ready-to-sequence metagenomic libraries in just under six hours. This application note demonstrates the performance of the automated Illumina DNA Prep workflow for metagenomics in comparison to the standard, manual workflow using stool samples from four human subjects. Refer to the automation methods web page for more information on qualified library prep automation methods.

Methods

The Automated Illumina DNA Prep workflow for metagenomics includes automated DNA extraction from stool samples using the chemagic 360 instrument (Revvity, Catalog no. 2024-0020) and the chemagic DNA Stool Kit (Revvity, Catalog no. CMG-1076). DNA extraction is followed by library preparation on the Sciclone G3 NGSx Workstation (Revvity, Catalog no. CLS145321) liquid handler using Illumina DNA Prep, (M) Tagmentation (96 Samples, IPB) (Illumina, Catalog no. 20060059). IDT for Illumina DNA/RNA UD Indexes Sets A to D allow users to generate up to 384 metagenomic libraries. Illumina DNA Prep Library Prep Kits feature innovative, on-bead tagmentation chemistry that supports quick and easy library preparation from various organisms and specimens.[3] The Illumina DNA Prep kit is compatible with a

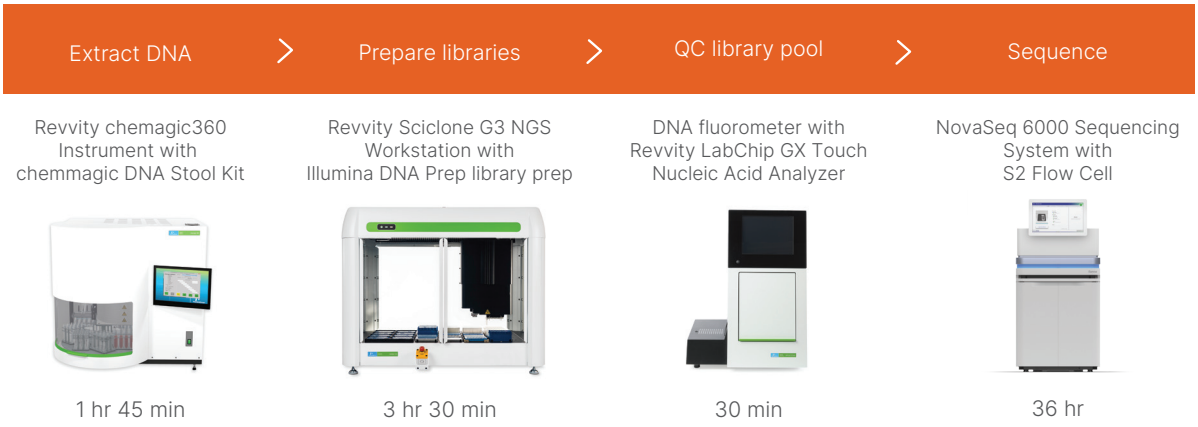

Figure 1: The automated Illumina DNA Prep workflow for metagenomics—Illumina and Revvity have collaborated to create a comprehensive, automated NGS library preparation workflow for high-throughput metagenomics.

range of DNA input amounts from 100 ng to 500 ng, which eliminates the need for precise quantification of the initial DNA sample, saving time and costs associated with library input normalization.[3]

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 hr.

Extractions were performed on the chemagic 360 Instrument using the chemagic DNA Stool Kit. Each isolation was performed with 150 µl elution volume, which produced a total of 300 ng to 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 Illumina DNA Prep input volume. Integrity of extracted DNA was assessed with the LabChip GX Touch Nucleic Acid Analyzer (Revvity, Catalog no. CLS137031), the HT DNA NGS 3K Reagent Kit (Revvity, Catalog no. CLS960013), and the Genomic DNA Reagent Kit (Revvity, Catalog no. CLS760685). The chemagic method provides optimal isolation of DNA from both Gram-negative and Gram-positive species.

Automated and manual library preparation

Ninety Illumina DNA Prep libraries were prepared from two independent automation runs on the Sciclone G3 NGSx Workstation liquid handler using Illumina DNA Prep, (M) Tagmentation (96 Samples, IPB). The total DNA input range (100–600 ng), overlapped with the recommended DNA input range for Illumina DNA Prep 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 Illumina DNA Prep manual protocol, a subset of 42 libraries from the same DNA isolates was prepared manually according to the standard protocol.

Sequencing

To generate sufficient genomic coverage for in-depth metagenomic analysis, 48 Illumina DNA Prep libraries were pooled by volume (5 µl each). The pooled libraries were assessed with the LabChip GX Touch Nucleic Acid Analyzer using an HT DNA NGS 3K Reagent Kit and yield was measured with the Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Catalog No. 15397463). Libraries were sequenced on the HiSeq™ X or NovaSeq™ 6000 Sequencing System* (S2 flow cell) 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[4,5] and Kraken Metagenomics (now DRAGEN™ Metagenomics)[6] apps using sequencing data sets downsampled to as low as 3 million reads and up to 80 million reads. *De novo* genome assembly quality was evaluated using MEGAHIT v1.1.1.9[7] and QUAST v4.410[8] using data sets downsampled to 40 million and 60 million reads.

Results

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 Illumina DNA Prep chemistry. More uniform insert size and greater library yield enable more uniform genome coverage and improves data accuracy. To assess insert size distribution, eight libraries prepared with the automated Illumina DNA Prep workflow were analyzed using the LabChip GX Touch Nucleic Acid Analyzer. An overlay of the eight LabChip traces demonstrates highly uniform insert sizes (Figure 2).

To further evaluate the consistency of the automated Illumina DNA Prep workflow, a series of libraries was prepared in triplicate with seven different extracted DNA input amounts. To evaluate the yield of the automated library preparation, the percentage of reads passing filter was plotted for each library in the sequenced pool of 21 libraries (Figure 3). High uniformity of index representation indicates uniform library yields and demonstrates that each library is evenly represented on the flow cell. The automated Illumina DNA Prep workflow produced libraries with highly uniform index representation, even with a range of extracted DNA inputs.

* The HiSeq X System is no longer available and has been replaced by the NovaSeq 6000 or NovaSeq X Systems.

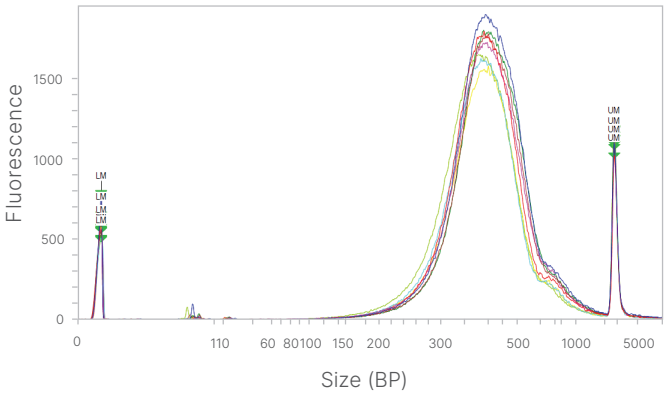

Figure 2: Insert size distribution of eight libraries prepared with automated workflow—Overlay of eight traces from the LabChip GX Touch Nucleic Acid Analyzer representing eight different libraries prepared with the automated Illumina DNA Prep workflow. The libraries represent two DNA isolation replicates from the adult, donor 1, sample and two DNA isolation replicates from the child, donor 3, sample. Two replicates were generated from each DNA isolate producing a total of eight libraries.

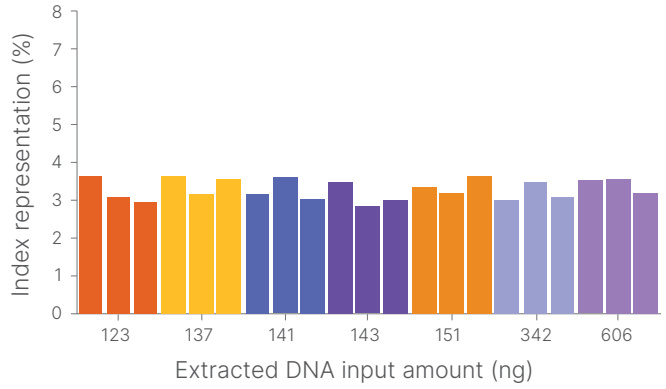

Figure 3: Index representation of libraries prepared from various amounts of extracted DNA with automated workflow—Libraries were prepared with seven different extracted DNA input amounts, pooled together by volume, and sequenced in triplicate. The graph illustrates the % reads identified passing filter (PF) for each library in the sequenced pool of 21 libraries.

Comparable metagenomic profiling results

To assess the performance of automated and manually prepared Illumina DNA Prep libraries in metagenomic profiling, automated libraries and manually prepared libraries were sequenced and analyzed with Kraken and CosmosID Metagenomics Apps (Figure 4, Figure 5).

The libraries prepared with the automated Illumina DNA Prep workflow show the same distribution of bacterial phyla and species as the manually prepared libraries. Furthermore, the automated Illumina DNA Prep workflow enabled identification of over 100 species in the donor 1 sample (Figure 6).

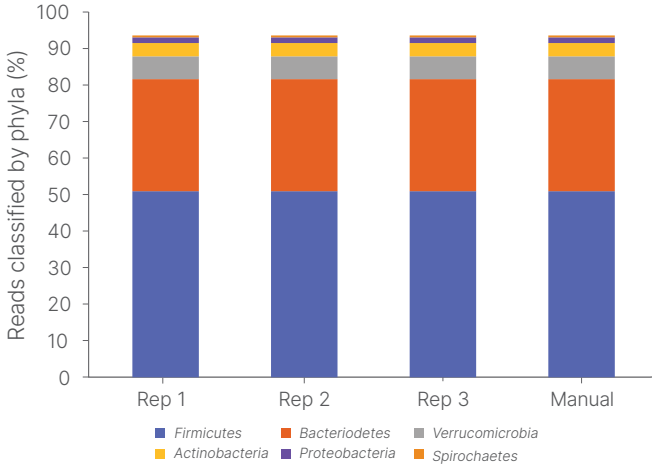

Figure 4: Comparison of automated and manually prepared libraries phyla distribution—Analysis of automated sequencing library replicates and one manually prepared library produced from the adult sample (donor 1). Bacterial phyla distribution assessed by DRAGEN Metagenomics App using 10 million paired-end reads.

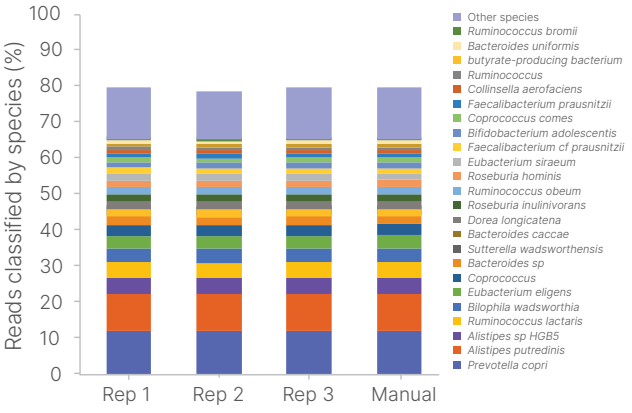

Figure 5: Comparison of automated and manually prepared libraries species distribution—Analysis of automated sequencing library replicates and one manually prepared library produced from the adult sample (donor 1). Bacterial species distribution assessed by CosmosID Metagenomics using 10 million paired-end reads.

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 percentage of the 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, the automated and manually prepared libraries generated nearly identical genome assembly results for all 12 organisms analyzed (Figure 7).

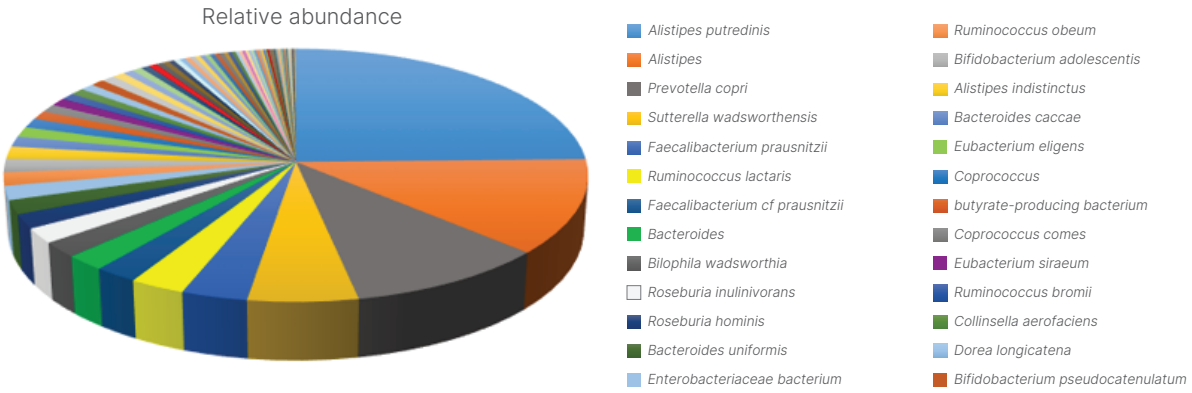

Figure 6: Automated libraries produce rich metagenomic profiles—Libraries were prepared from the adult sample (donor 2) using the automated Illumina DNA Prep library prep workflow. The CosmosID Metagenomics App was used with 40 million reads to assemble the relative abundance pie chart and identify over 100 species (26 of > 100 identified species are included in the figure legend).

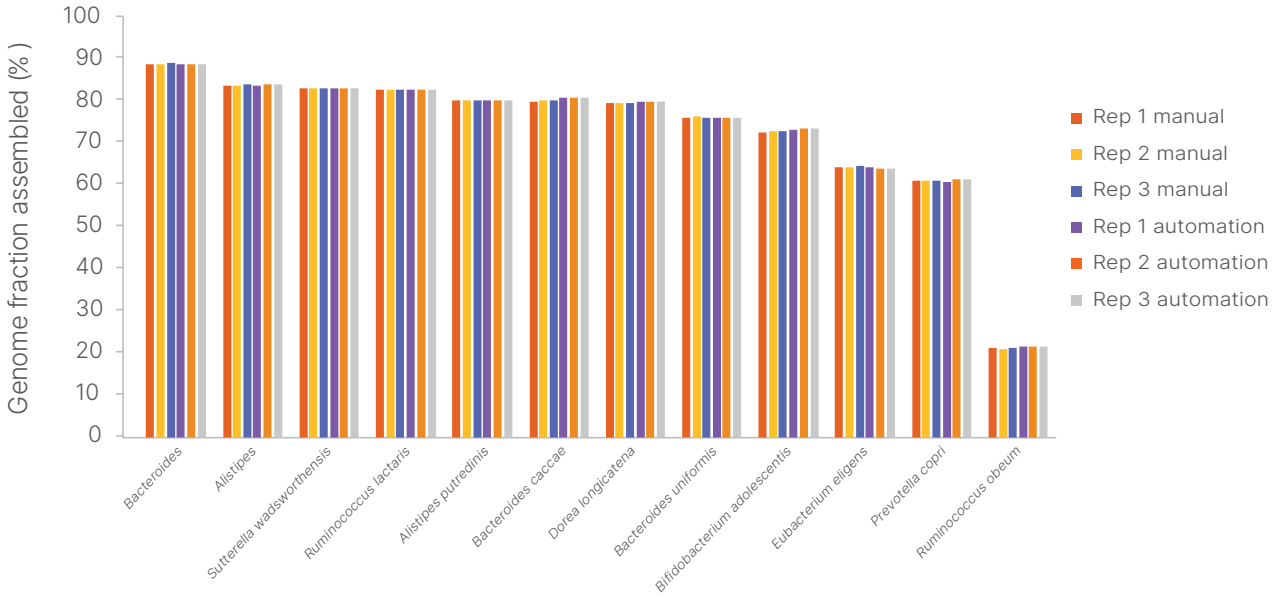

Figure 7: Comparison of automated and manually prepared libraries with genome assembly—*De novo* genome assembly of 12 microorganisms was performed with QUAST using 60 million reads. Libraries were prepared from donor 1 sample in triplicate using the automated and manual protocols.

Summary

The automated Illumina DNA Prep 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 and prepare up to 96 libraries with the Sciclone G3 NGSx Workstation and the Illumina DNA Prep kit. The automated workflow generates highly uniform libraries and provides excellent data for species identification and metagenomic profiling of complex samples, including challenging stool samples. With significant advantages that include higher library consistency, fewer touchpoints, and greater throughput, the automated Illumina DNA Prep library prep workflow is an ideal library prep solution for labs seeking to scale up their next-generation sequencing capacity.

Learn More

Illumina DNA Prep

Microbial whole-genome sequencing

Ordering information

Product	Catalog no.
Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB)	20060060
Illumina DNA Prep, (M) Tagmentation (96 Samples, IPB)	20060059
Flex Lysis Reagent Kit	20018706
Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)	20091654
Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 Indexes, 96 Samples)	20091656
Illumina DNA/RNA UD Indexes Set C, Tagmentation (96 Indexes, 96 Samples)	20091658
Illumina DNA/RNA UD Indexes Set D, Tagmentation (96 Indexes, 96 Samples)	20091660

References

1. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap Adv Gastroenterol.* 2013;6(4):295-308. doi:10.1177/1756283X13482996
2. Clarke SF, Murphy EF, Nilaweera K, et al. The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes*. 2012;3(3):186-202. doi:10.4161/gmic.20168
3. Illumina. Illumina DNA Prep data sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-dna-prep-data-sheet-m-gl-01373/illumina-dna-prep-data-sheet-m-gl-01373.pdf. Published 2023. Accessed April 22, 2024.
4. Illumina. CosmosID Metagenomics App. illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/cosmosid-cosmosid-metagenomics.html. Accessed April 22, 2024.
5. CosmosID. CosmosID homepage. cosmosid.com. Published 2023. Accessed May 7, 2024.
6. Illumina. DRAGEN Metagenomics Pipeline. illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/dragen-metagenomics-pipeline.html. Accessed April 22, 2024.
7. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674-1676. doi:10.1093/bioinformatics/btv033
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi:10.1093/bioinformatics/btt086

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02085 v1.0