BECKMAN COULTER
Life Sciences

Automation of Illumina DNA Prep Kit on the Biomek NGeniuS Next Generation Library Prep System

Abstract

Next-Generation Sequencing (NGS) is rising in popularity as the costs of applying this technology to genome sequencing fall. Laboratories are looking for highly reproducible NGS sample prep methods that limit potential for error. In this paper, we detail an automated process for the Illumina DNA Prep library preparation kit that offers the laboratory optional settings to optimize a demonstrated application that will process between 4 and 24 samples from start to finish with minimal interaction from the user.

Introduction

Over the past several decades, DNA sequencing methods have advanced from Frederick Sanger's discovery of the chain termination technique in the 1970s to the massively parallel sequencing techniques that were introduced beginning in 2005. These modern sequencing techniques have become common laboratory workflows dubbed next-generation sequencing (NGS). One key step in NGS workflows involves preparation of a library of DNA fragments ready for sequencing. Unfortunately, this process has been a bottleneck in NGS workflows. The creation of libraries for NGS is a tedious process that can take anywhere from 2.5 hours to several days to complete depending on the type of library created. Great care must be taken to keep accurate records of sample and adapter pairs. Pipetting each adapter by hand can lead to errors in the creation of libraries with the correct adapters. Many of the processes are timed and do not have safe stopping points, leading to a very long day. Due to these factors, many labs have found automation of these critical NGS steps to be highly desirable.[1]

Illumina's DNA Prep kit (previously called Nextera DNA Flex kit) is a library preparation method that is suitable for a broad range of applications and multiple sample types. The kit supports input masses ranging from 1-500 ng. The Illumina DNA Prep kit uses Illumina's bead-linked transposon chemistry to create sequencing-ready fragments. Sample DNA is incubated with and bound to the transposons linked to beads. Transposons cut the DNA and insert nucleotide adapters on the end of each piece. This simultaneously fragments genomic DNA samples while adding Illumina sequencing primers. The bead-based "tagmentation" selects for fragments that are approximately 300-350 base pairs by limiting the size of fragments that can bind onto the bead. Fragments are amplified, and a double-sided cleanup is done post-PCR to remove any unwanted fragments. The size distribution of the final library is then assessed using a fragment analyzer system such as the Agilent Bioanalyzer or Agilent TapeStation, while

Figure 1. Workfl ow for Illumina DNA Prep protocol on the Biomek NGeniuS system. The blue arrows indicate the steps that are done on the instrument; the red arrows indicate the steps that are not.

a fluorometric assay, such as Qubit, is used to determine the library yield. The average size of each library at the end of the protocol is approximately 600bp; size selection is optional during a double-sided cleanup step.[2,3]

In this application note, we have demonstrated the automated preparation on the Biomek NGeniuS system at 1 ng, 50 ng, and 500 ng and have also analyzed data obtained from processing using the instrument. The hands-on time required to run this assay is reduced, and the interactions with the instrument are limited.

Metric	NGeniuS 8 samples	NGeniuS 16 samples	NGeniuS 24 samples	Manual 16 samples
Hands-On Time	20 min	20 min	20 min	1 hour, 15 min
Total Sample Processing Time (includes reagent aliquoting)	5 hr, 6 min	7 hr, 35 min	10 hr, 10 min	3 hr
Customer Interactions	1	1	1	Multiple

Table 1. Application run time and customer interactions for the Illumina DNA Prep protocol on the Biomek NGeniuS system. Processing time for runs includes nucleic acid normalization.

1. Materials and Methods

1.1. Run Setup

Genomic DNA samples (*Homo sapiens* NA12878 from the Coriell Institute and *Bacillus cereus* strain 971 from ATCC) were quantified using the Qubit DNA BR kit (Thermo Fisher Scientific) and diluted to an initial starting concentration suitable for the Biomek NGeniuS system.

Normalization of input nucleic acid is performed on the instrument by diluting an aliquot of the sample to the input volume required by the library preparation kit to arrive at the correct starting concentration. In order to reduce manual pipetting errors, the concentration of input nucleic acid must be within 100X of the concentration required by the library preparation kit so that the operator is not attempting to manually pipette small volumes of highly concentrated input nucleic acid.

Sample	Vendor	Part Number
Homo sapiens gDNA – CEPH/Utah pedigree NA12878	Coriell Institute for Medical Research	NA12878
Bacillus cereus gDNA – strain 971	American Type Culture Collection	14579D-5

Table 2. Sample types used in preparations of samples for Illumina DNA Prep.

When samples are ready, the run is set up in the Biomek NGeniuS customer portal. The first step is to select the **+create** button to create a batch to be run on the system (Figure 2).

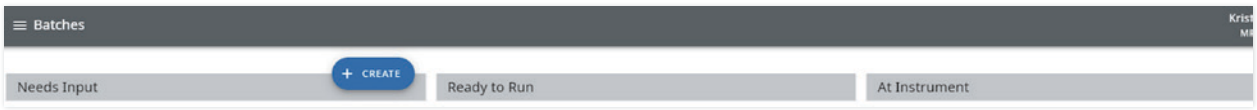

Figure 2. The **+create** button in the above figure is used to begin a new batch setup.

Next, the Illumina DNA Prep App is selected to process samples. The setup is broken up into four sections:

1.1.1. Batch info

This section records the name of the batch and the number of samples to be run (**Figure 3**). The batch name is a unique run name for the samples being processed. The number of samples is any number between 4 and 24 for this application, as indicated by the light grey numbers below the input box.

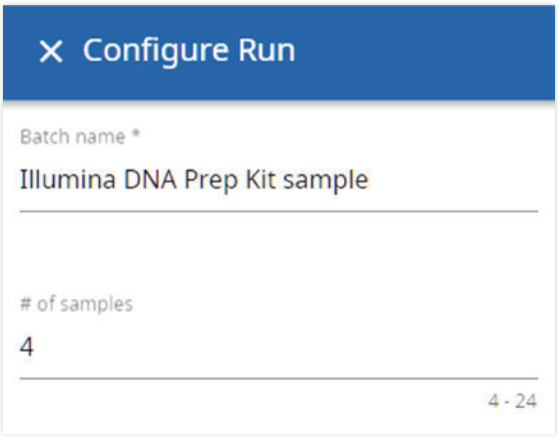

Figure 3. Batch name and number of samples for batch run.

1.1.2. Batch info

App Settings contains variables specific to the library kit that may be changed between runs or may be locked by the lab administrator **(Figure 4). Table 3** lists the app settings and descriptions of each setting.

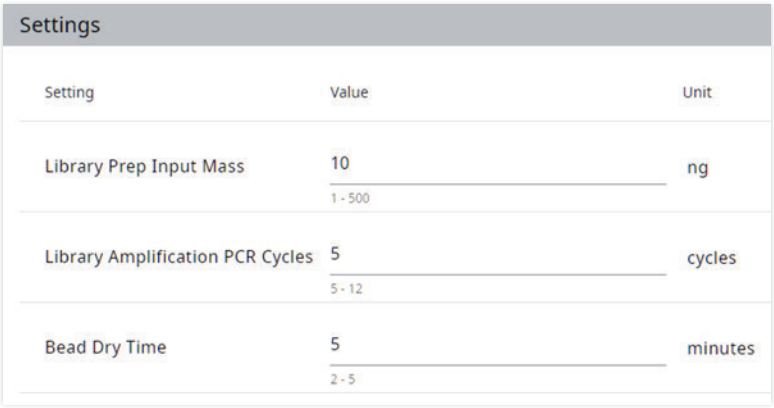

Figure 4. App settings for batch run.

Setting	Description
Library Prep Input Mass	The mass of genomic DNA (ng) used at the beginning of the library preparation workflow.
Library Amplification PCR Cycles	The number of cycles of PCR amplification to perform. The Illumina DNA Prep instructions provide recommendations for the number of PCR cycles, depending on the mass of input DNA.
Bead Dry Time	The amount of time to let beads containing prepped libraries dry after post-PCR cleanup EtOH wash. Values are between 2-5 minutes, never exceeding 5 minutes.

Table 3. App Settings and descriptions for each setting.

1.1.3. Sections

The next section of data to be filled out is **Sections (Figure 5)**. Illumina DNA Prep has three potential sections. Users can select where to start in the process just below the Sections marker in **Figure 5**. Some users may prefer to do the first section, **Normalize Samples**, by hand. If so, they can elect to utilize the **Start at section** drop-down menu to select section #2. A drop-down menu allows the user to select any starting point available. Starting points are determined by safe stops defined in the instructions for use of the library prep kit, which are also suitable for a safe stop on the Biomek NGeniuS system. The blue slider to the left of the sections allows the user to select a safe stop to end processing of samples. The instrument is designed to run unattended, but users can elect to stop processing and store samples safely before resuming the run at the next shift.

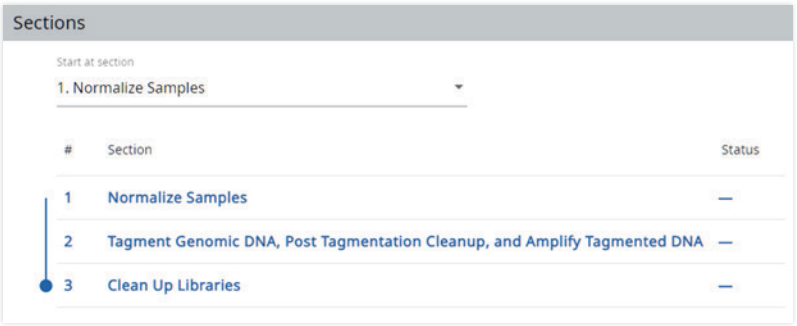

Figure 5. Sections for Illumina DNA Prep Application.

1.1.4. Sample Data

The final step in setting up a batch to run is to input the sample data **(Figure 6**). In the sample data section, users can click the **DOWNLOAD SAMPLE DATA TEMPLATE** and fill in the appropriate information. This is a .csv file that is filled out and uploaded into the sample data by clicking the **Upload** button. Users can utilize tool tips to determine what information goes into each column by hovering over the header of each column. Illumina DNA Prep has four different data pieces that are required for tube-based index processing. The first column is the sample **Well** location in alphanumeric notation. The second column is the **Sample_ID** of each sample. The third item, **Index_i5**, is the H5xx index to be used. The fourth column, **Index_i7**, is the H7xx index to be used. The final column, **Initial Concentration**, is the concentration of DNA that will be placed into each well for dilution and processing for library preparation. Once the data is entered in the template and saved, the user clicks the **UPLOAD** button. If there are unexpected values in the sample data file, a red box will appear indicating the source of the problem. Users can fix the data file and upload again if needed. The final step is to click the **READY TO RUN** button in the top right of the screen. The batch can be initiated at any Biomek NGeniuS system within the same lab.

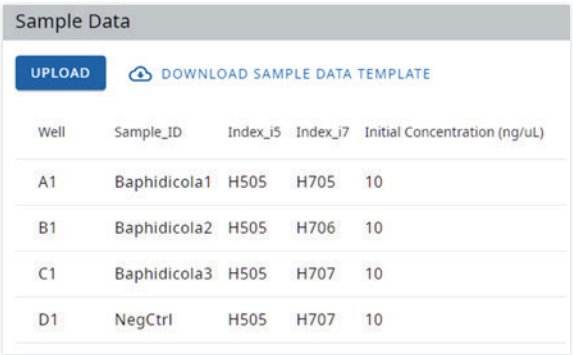

Figure 6. Simulated Sample Data information. The four columns starting with **Sample_ID** must be user defined in the sample-data-template.csv file and uploaded to the Biomek NGeniuS Portal.

1.2. Consumables Used

Reagents	Manufacturer	Part Number
Illumina® DNA Prep, (M) Tagmentation (24 Samples)	Illumina	20018704
Nextera™ DNA CD Indexes (24 Indexes, 24 Samples)	Illumina	20018707
Qubit dsDNA BR Assay Kit	Thermo Fisher Scientific	Q32850
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
High Sensitivity D1000 ScreenTape	Agilent	5067-5584
High Sensitivity D1000 Reagents	Agilent	5067-5587
NextSeq 500/550 High Output Kit v2.5 (300 Cycles)	Illumina	20024908
AMPure XP Beads	Beckman Coulter Life Sciences	A63882
PCR grade Water	Invitrogen-Life Technology	10977-015
Ethanol	American Bio	AB00515-00500

Table 4. Reagents used in preparation of libraries with Illumina DNA Prep kit and sequencing on Illumina sequencer.

Equipment	Manufacturer
Biomek NGeniuS Sample Prep System	Beckman Coulter Life Sciences
NextSeq 550 Sequencer	Illumina
Allegra X-14 Centrifuge	Beckman Coulter Life Sciences
Qubit	Thermo Fisher Scientific
4200 TapeStation	Agilent

Table 5. Equipment used in sample preparation and processing of Illumina DNA Prep.

Consumable	Manufacturer/ Part Number
Qubit Tubes	Thermo Fisher Scientific # Q32851
Foil Plate Seals	Beckman Coulter 538619
Biomek NGeniuS Instrument Reaction Vessel, 24 well	Beckman Coulter C62705
Biomek NGeniuS Instrument Lid, 24 well	Beckman Coulter C62706
NGeniuS Bulk reservoirs	Beckman Coulter C62707
NGeniuS seal pads	Beckman Coulter C70665
NGeniuS reagent plugs	Beckman Coulter C62706
1025 µL Conductive Filtered Tips, Case	Beckman Coulter C59585
70 µL Conductive Filtered Tips, Case	Beckman Coulter C62712
Empty Tip box 1025 µL, Case	Beckman Coulter C70672
Empty Tip box 70 µL, Case	Beckman Coulter C70673

Table 6. Consumables required for sample processing.

1.3. Sequencing the NGeniuS Produced Libraries

Samples of *Bacillus cereus* and *Homo sapiens* gDNA were processed on the Biomek NGeniuS system using reagents, equipment, and consumables detailed in **Tables 4, 5, and 6**. System requested reagents from the Illumina DNA Prep kit and bulk reagents (**Table 4**) along with Biomek NGeniuS consumables (**Table 6**) were loaded onto the system for processing.

The variables that were selected for processing are seen in **Table 7**. After all reagents and consumables had been allocated to proper indicated storage locations, the user was instructed to remove excess reagents and notified of an estimated time of completion for the library prep based off selections the user input at the start of the protocol.

Following normalization of the input samples, Illumina DNA Prep libraries were constructed on the system. After completion of the runs, the resulting libraries were analyzed using the 4200 TapeStation with D1000 High Sensitivity ScreenTape (Agilent) to determine library size and the Qubit dsDNA HS assay (Thermo Fisher Scientific) to determine library concentration. Libraries were then sent to Illumina for sequencing. *Bacillus cereus* libraries were sequenced on a single run of an Illumina NextSeq 550 system using a NextSeq 500/550 High Output Kit v2.5 (300 Cycles) with a 2x151 bp sequencing run with a loading concentration of 1.2 pM. *Homo sapiens* libraries were first sequenced on NextSeq 500/550 High Output Kit v2.5 (300 Cycles) with a 2x151 bp sequencing run.

Additionally, 12 of the *Homo sapiens* libraries (three 500 ng input libraries and nine 50 ng input libraries) were re-sequenced later using a 2x151 bp sequencing run on an Illumina NovaSeq sequencer with an S4 flowcell to achieve higher depth of coverage.

Data was analyzed using the DRAGEN Germline App (version 3.10.4) on Illumina BaseSpace. For *Homo sapiens* libraries, the Human HG38 Alt-Masked v2 reference genome provided by BaseSpace was used. *Bacillus cereus* libraries were mapped against the reference genome supplied by ATCC (genomes.atcc.org). When analyzed using the DRAGEN Germline App, the *B. cereus* libraries were downsampled to 2,000,000 reads for each library to accommodate the smaller size of the B. cereus genome (5.4 Mb) compared to *H. sapiens* (3.1 Gb).

Library Prep Input Mass (ng)	Sample type	Indices	PCR cycles	Bead Dry Time (min)
1 ng	*B. cereus*	Nextera™ DNA CD Indexes (24 Indexes, 24 Samples)	12	5
50 ng	*H. sapiens*	Nextera™ DNA CD Indexes (24 Indexes, 24 Samples)	5	5
500 ng	*H. sapiens*	Nextera™ DNA CD Indexes (24 Indexes, 24 Samples)	5	5

Table 7. Method variables and selections for Illumina DNA Prep library kit. Parameters outlined in this table correspond to selections made in **Settings** when initiating a run (**Figure 4**).

2. Results & Discussion

After completion of the runs by the Biomek NGeniuS system, the resulting libraries were analyzed using the 4200 TapeStation with D1000 High Sensitivity ScreenTape to determine library size and concentration.

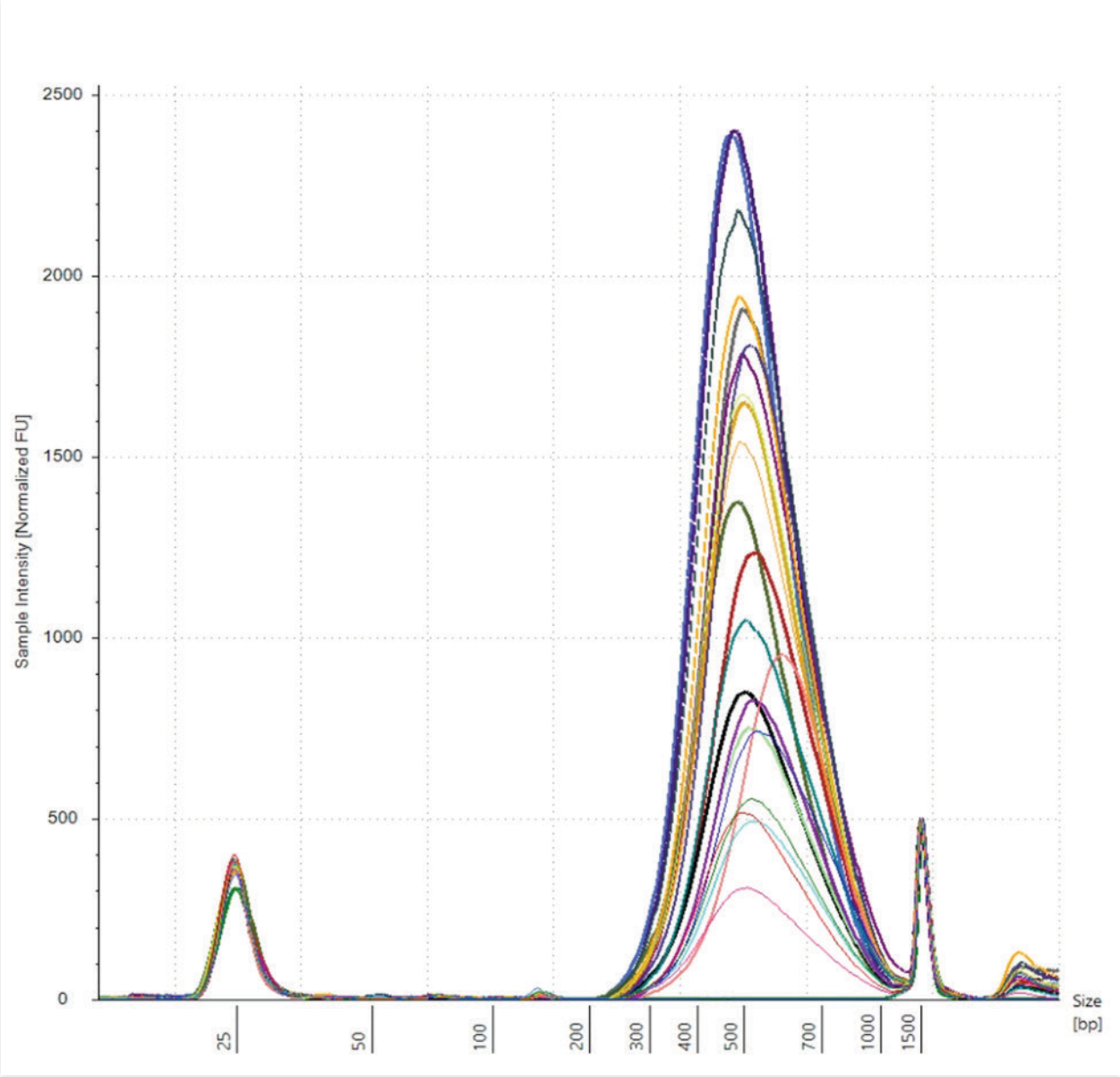

Figure 7. Agilent TapeStation trace results from libraries created on the Biomek NGeniuS system from 22 *Bacillus cereus* gDNA samples (1ng input mass) and two negative controls using the Illumina DNA Prep library prep kit. Libraries have an average fragment size of 560 bp, averaged across all libraries created from these samples.

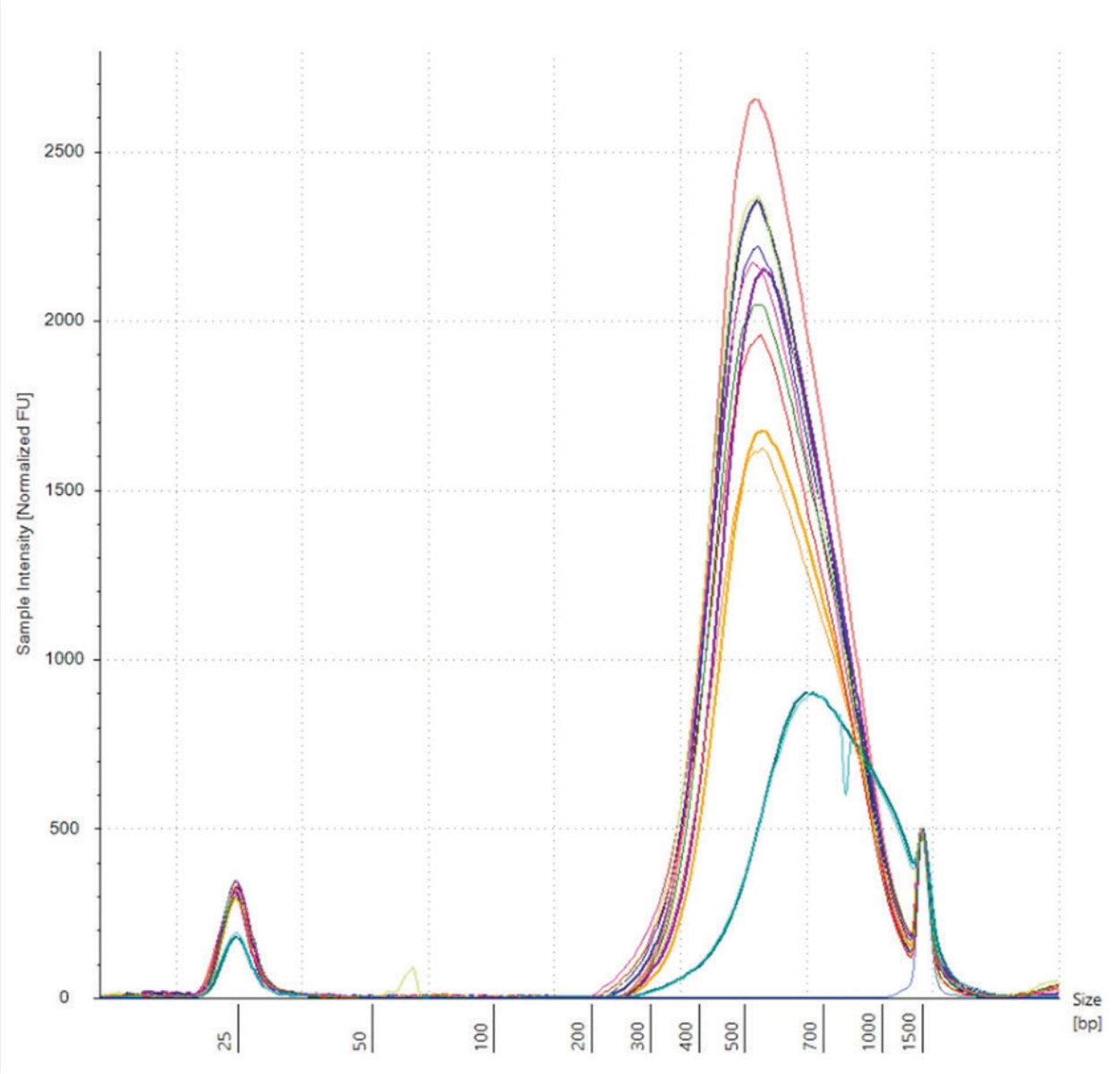

Figure 8. Agilent TapeStation trace results from libraries created on the Biomek NGeniuS system from 12 Coriell Institute NA12878 samples (50 ng input mass) and one negative control using the Illumina DNA Prep library prep kit. Libraries have an average fragment size of 628 bp, averaged across all libraries created from these samples.

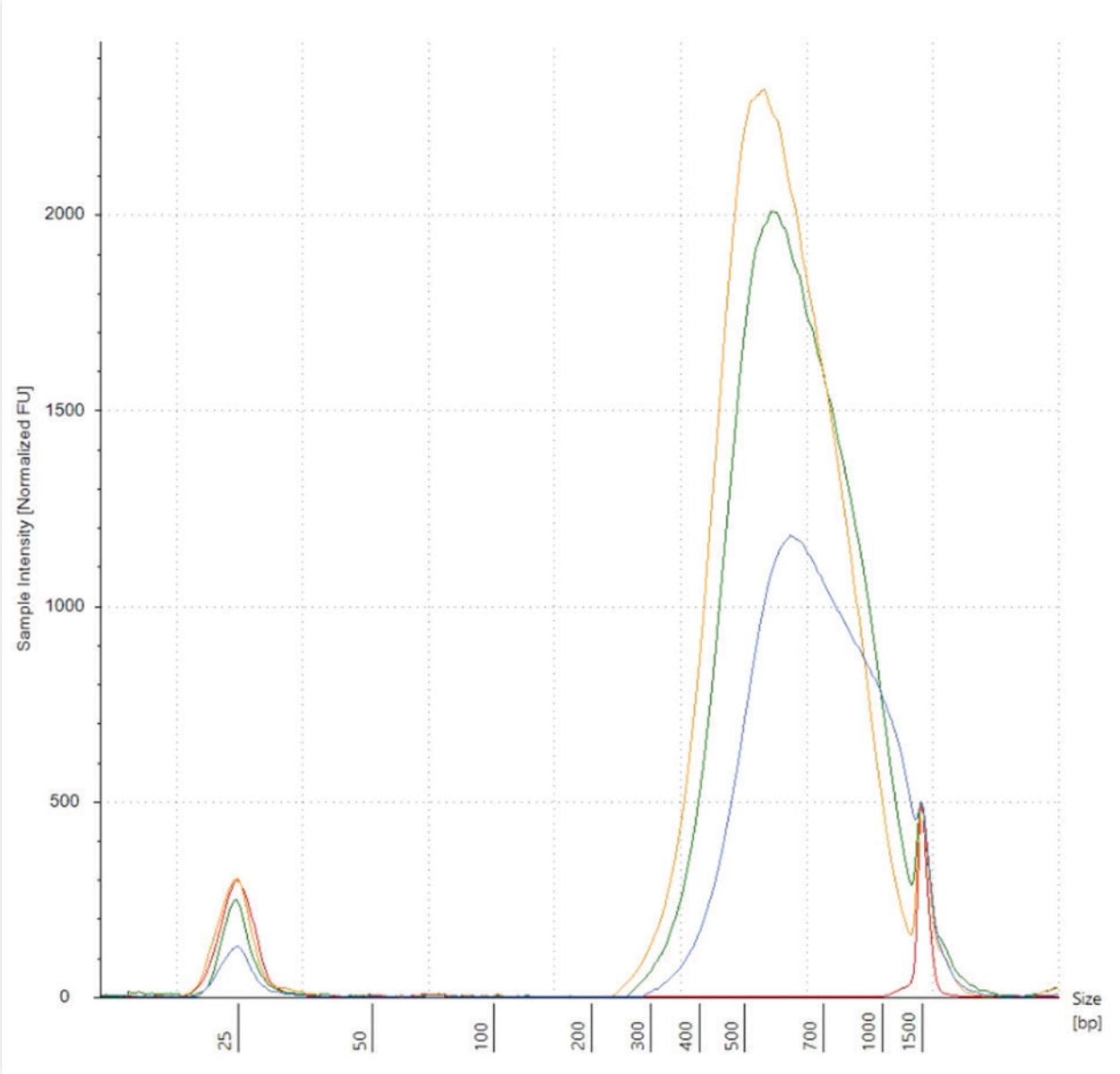

Figure 9. Agilent TapeStation trace results from libraries created on the Biomek NGeniuS system from three Coriell Institute NA12878 samples (500ng input mass) and one negative control using the Illumina DNA Prep library prep kit. Libraries have an average fragment size of 664 bp, averaged across all libraries created from these samples.

The *B. cereus* str. 971 gDNA (ATCC) 1 ng input DNA run produced 22 libraries (plus two negative controls) with an average size of 560 bp (**Figure 7**). The two runs utilizing *H. sapiens* NA12878 gDNA (Coriell Institute) produced 12 libraries (plus one negative control) from 50 ng of DNA input per library and three libraries (along with one negative control) from 500 ng of DNA input, with average library sizes of 628 bp (**Figure 8**) and 664 bp (**Figure 9**) respectively. Library sizes between *B. cereus* and *H. sapiens* showed a difference in distribution (**Figure 10**) while concentrations of the final libraries did not vary appreciably by input mass or input DNA type (**Figure 11**).

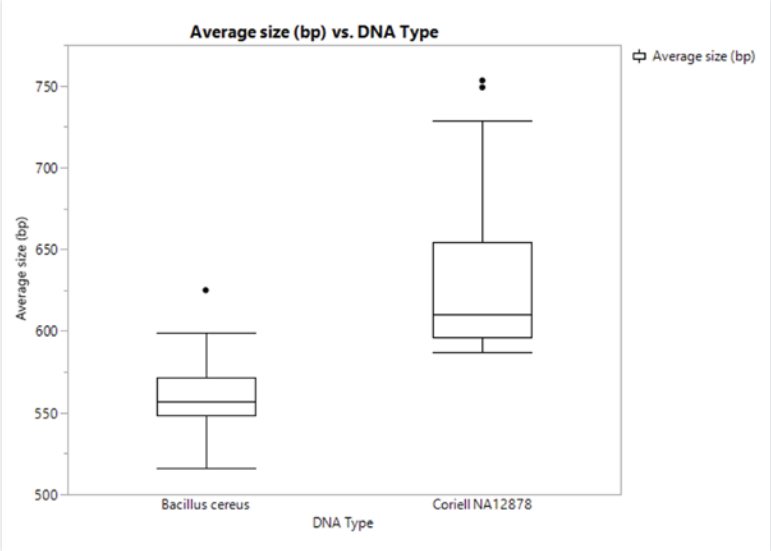

Figure 10. Comparison of library average sizes by input DNA type.

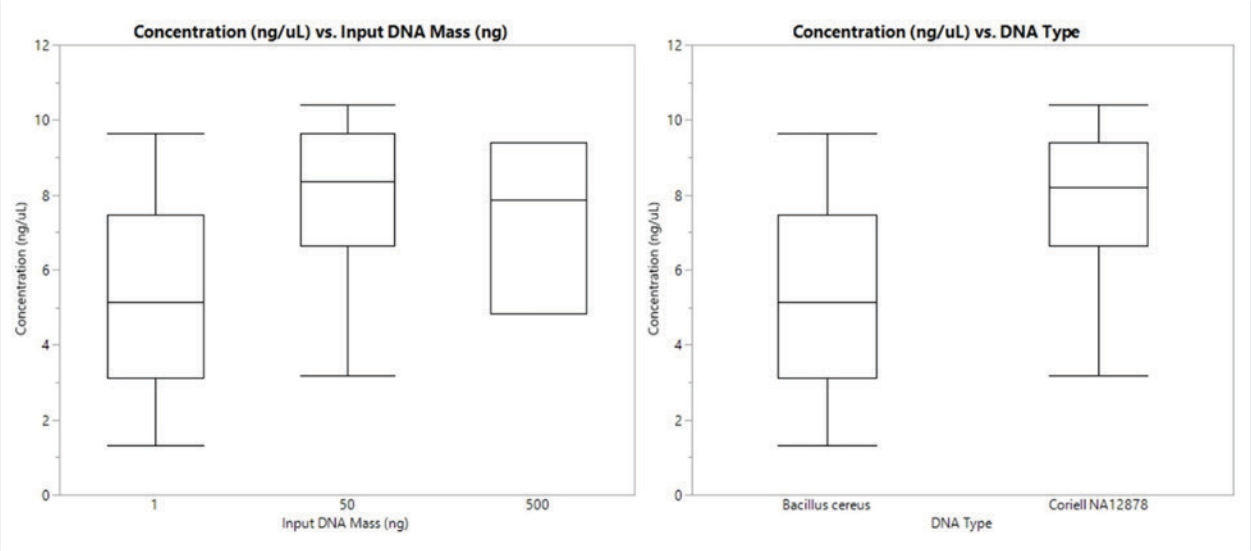

Figure 11. Comparison of final library concentration by Input DNA Mass (left) and by Input DNA type (right).

Sequencing results from the two NextSeq 550 runs returned a total of 379M pass filter reads, generating a total of 213.35 Gb of sequencing data with 88% of bases scored at Q30 or higher. The single NovaSeq run generated a total of 5.1 billion pass filter reads, resulting in 3.18 Tb of sequencing data with 85.8% of bases scored at Q30 or higher. Index CV for libraries run on the two NextSeq 550 runs were 0.1418 and 0.4988. Index CV for the libraries run on the single NovaSeq run was 0.3371.

After alignment to appropriate reference genome, the number of reads mapping back to the reference genomes was higher than 95% for all libraries. Percent properly paired reads were on average higher for *H. sapiens* libraries (average 98.5%) than the *B. cereus* libraries (average 93.4%). Reads marked as PCR duplicates were less than 3.2% for all libraries, with an average of 2.2%. Sequencing metrics for the NextSeq 550 runs are shown in Figure 11.

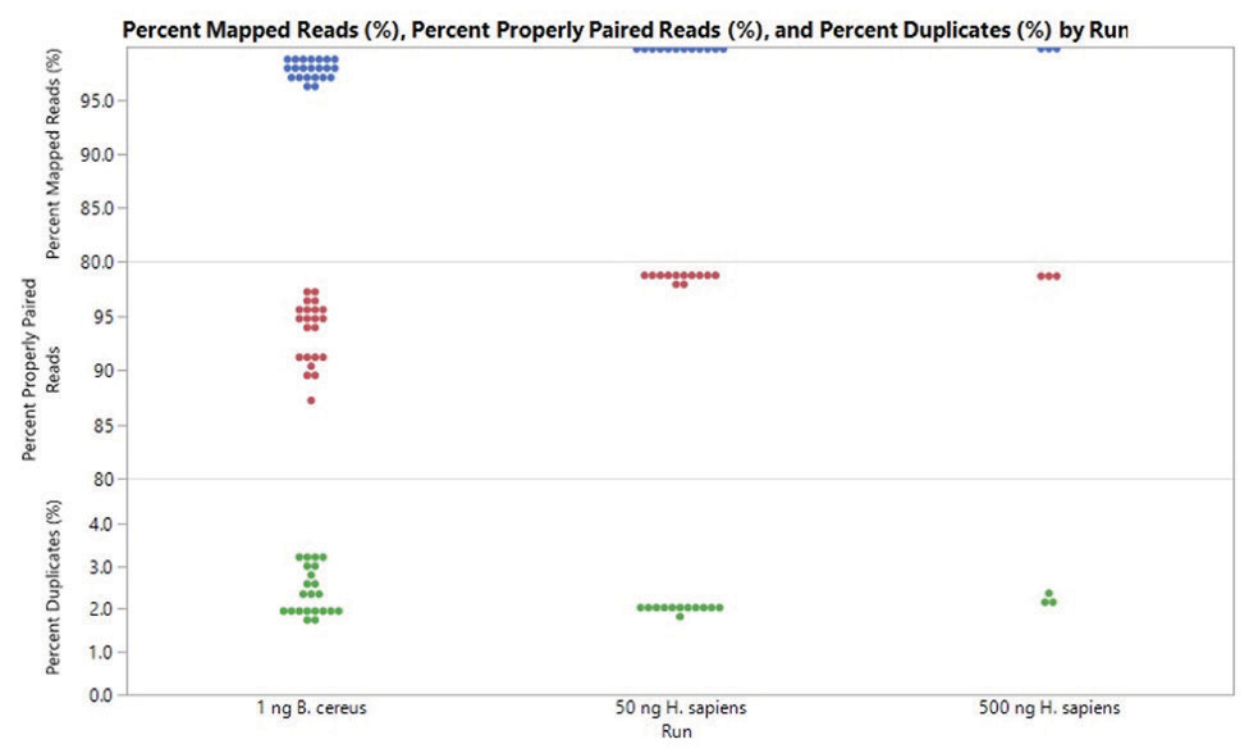

Figure 12. NextSeq 550 sequencing results. Results fall within passing parameters defined by Illumina.[2]

Following the two NextSeq 550 sequencing runs, 12 of the *Homo sapiens* libraries were re-sequenced on an Illumina NovaSeq sequencer using an S4 flowcell with a loading concentration of 400 pM along with a 1% phiX spike in. The averaged results are presented below (**Table 8**) and track closely with the NextSeq 550 data.

Library and Sequencing Run	Average Total Reads	Average % of reads mapped to reference genome	Average Depth of Coverage	Average % duplicates
50 ng *H. sapiens* on NextSeq 550	41,218,135	98.96%	1.89 X	1.92%
500 ng *H. sapiens* on NextSeq 550	52,927,405	99.82%	2.4 X	2.19%
50 ng *H. sapiens* on NovaSeq	1,401,113,834	99.56 %	61.2 X	5%
500 ng *H. sapiens* on NovaSeq	2,215,244,297	99.4%	94.23 X	6.2%

Table 8. Sequencing results from the NovaSeq sequencing run.

Summary

Library yields, sizes and sequencing data demonstrated that the automation of Illumina DNA Prep library kit using the Biomek NGeniuS system produces libraries with >97.7% alignment, low percentage of duplicate reads, and >93.48% of fragments properly paired to reference genomes. Generated libraries fall within the recommended library size range of the Illumina DNA Prep kit.

We demonstrated the Biomek NGeniuS system can successfully produce high-quality whole genome sequencing libraries suitable for sequencing on the Illumina platforms using the Illumina DNA Prep kit.

References

1. Bruinsma S, Burgess J, Schlingman D, Czyz A, Morrell N, Ballenger C, et al. Bead-linked transposomes enable a normalization-free workflow for NGS library preparation. *BMC Genomics*. 2018 Oct 1;19(1):722. doi: 10.1186/s12864-018-5096-9. PMID: 30285621; PMCID: PMC6167868.
2. Illumina DNA Prep Reference Guide Document 10000000025416 V09, June 2020, **https://sapac.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-09.pdf**
3. Illumina Nextera DNA Flex Library Preparation Kit Technical Data Sheet, **https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/nextera-dna-flex-data-sheet-770-2017-011.pdf**

For Beckman Coulter's worldwide office locations and phone numbers, please visit Contact Us at **beckman.com**

2022-GBL-EN-100603-v1