ATCGGCA
TTTGGCC
CGCATCG
GACTACG
NGS

A novel, single-tube enzymatic fragmentation and library construction method enables fast turnaround times and improved data quality for microbial whole-genome sequencing

Next-generation whole genome sequencing of microbes demands rapid, robust, and scalable library construction workflows, capable of generating high-quality sequence data across a wide range of genome sizes, complexities and genomic GC content. In this Application Note, we describe a streamlined library preparation method that results in minimal bias, high uniform coverage, and facilitates de novo *assembly of microbial genomes.*

Authors

Bronwen Miller
Victoria van Kets
R & D Scientists

Beverley van Rooyen
Senior Support & Applications Scientist

Heather Whitehorn
Support & Applications Scientist

Piet Jones
Bioinformatics Scientist

Martin Ranik
R & D Team Leader

Adriana Geldart
Support & Applications Manager

Eric van der Walt
R & D Manager

Maryke Appel
Technical Director

Kapa Biosystems
Cape Town, South Africa and
Wilmington, MA, USA

Nguyet Kong
Research Associate

Carol Huang
Research Associate

Dylan Storey
Post-doctoral Fellow

Bart C. Weimer
Professor and Director

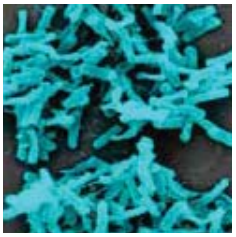
UC Davis School of Veterinary Medicine & 100K Pathogen Genome Project
Davis, CA, USA

INTRODUCTION

Whole-genome sequencing (WGS) of microbes employing next-generation sequencing (NGS) technologies enables pathogen identification, differentiation, and surveillance on an unprecedented scale and level of resolution—thereby profoundly impacting diagnostic microbiology and public health. To fully capitalize on the benefits of greater sequencing capacity, faster sequencing technologies and lower per-genome costs, rapid and robust NGS library construction workflows are needed to support both *de novo* and re-sequencing applications.

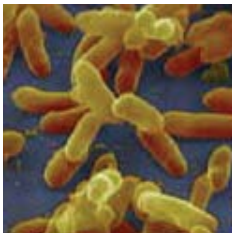
A major focus area in NGS library construction for microbial WGS has been the elimination of mechanical shearing—which requires expensive, specialized equipment and consumables, and is both laborious and difficult to scale. Alternative, enzymatic fragmentation solutions based on transposases ("tagmentation") or mixtures of DNA endonucleases and nicking enzymes, offer significant benefits in terms of throughput and turnaround times. However, these often come at a cost to reproducibility, control over fragment length, and sequence data quality (coverage depth and uniformity)—particularly for organisms with extreme (highly GC- or AT-rich) genomic content.

The KAPA HyperPlus Kit is a robust and versatile kit for the construction of DNA libraries for Illumina sequencing from a range of sample types and inputs (1 ng – 1 µg). The streamlined, one-tube workflow—which includes enzymatic fragmentation with a novel enzyme cocktail—offers the speed and convenience of tagmentation-based methods, but the control and performance of ligation-based library construction from Covaris-sheared DNA.



Clostridium difficile (29% GC)

Escherichia coli (51% GC)



Bordetella pertussis (68% GC)

Figure 1. Bacterial species used for library preparation and sequencing.

EXPERIMENTAL DESIGN

Current Illumina® library construction methods employing non-mechanical solutions for DNA fragmentation have three key limitations, namely: (i) poor control over fragment length, which is related to sensitivity with respect to DNA input; (ii) low library construction efficiency; and (iii) sequence biases, introduced during fragmentation and/or compulsory library amplification. Combined, these factors limit the amount and alter the representation of input DNA that is converted to usable reads—ultimately affecting coverage depth and uniformity, and the quality and completeness of *de novo* genomes.

To address these concerns and illustrate the benefits of the KAPA HyperPlus workflow for the production of high-quality libraries for microbial WGS, we sequenced the genomic DNA of three bacteria from whole genome shotgun libraries prepared using four different fast library construction strategies. The four methods were compared with respect to key library construction, sequencing, and *de novo* assembly metrics.

The bacterial species (Figure 1), *Clostridium difficile* (29% GC), *Escherichia coli* (51% GC), and *Bordetella pertussis* (68% GC), are all relevant for human health and were selected to represent a wide range of genomic GC content.

Library preparation methods are summarized in Table 1. The KAPA Hyper Prep Kit with Covaris-sheared DNA represents the industry standard for high-efficiency DNA library preparation. The KAPA HyperPlus Kit contains the novel KAPA Frag reagent for enzymatic fragmentation, developed by Kapa Biosystems (Wilmington, MA) to overcome the drawbacks of current non-mechanical fragmentation solutions, and work synergistically with the KAPA Hyper Prep chemistry to improve library construction efficiency. Fragmentase from New England Biolabs (Ipswich, MA) employs a combination of a dsDNA nicking enzyme and an endonuclease. Both the KAPA Hyper Prep and NEBNext Ultra kits offer streamlined, single-tube, ligation-based library preparation protocols. The Nextera XT DNA Library Preparation Kit from Illumina (San Diego, CA) is based on tagmentation technology.

To demonstrate the practical utility and benefits of the KAPA HyperPlus workflow for large-scale microbial genome projects, sequencing metrics for selected draft genomes, released by the 100K Human Pathogen Genome Project (UC Davis, Davis, CA) are included at the end of this Note.

Table 1. Library construction methodologies used in this study.

Abbreviation	Fragmentation method/kit	Library preparation kit	Prep time
Hyper Prep	Covaris shearing	KAPA Hyper Prep Kit	4 h
HyperPlus	KAPA Frag reagent for Enzymatic Fragmentation	KAPA Hyper Prep Kit	3 h
NEBNext	NEBNext dsDNA Fragmentase	NEBNext Ultra DNA Library Prep Kit for Illumina	4 h
Nextera	Tagmentation, Nextera XT DNA Library Preparation Kit		2.5 h

MATERIALS AND METHODS

Comparative Library Construction

Libraries were prepared in duplicate from 1 ng of bacterial genomic DNA (the optimal input for the Nextera XT chemistry), obtained from the American Type Culture Collection (ATCC; Manassas, VA). Strains and accession numbers were as follows: *C. difficile* (Hall and O'Toole) Prevot, strain 630 (BAA-1382); *E. coli* (Migula) Castellani and Chalmers, strain MG1655 (700926) and *B. pertussis* (Bergey, et al.) Moreno-Lopez, strain Tohama 1 (BAA-589).

Unless indicated otherwise, library construction was performed with reagents supplied in the respective library preparation kits, following standard protocols.

Hyper Prep workflow: Input DNA was sheared in 130 µL microtubes with a Covaris E220 Focused Ultrasonicator (Covaris; Woburn, MA), using parameters optimized for a median fragment length of 500 bp. Fragmented DNA was used directly for library construction using the KAPA Hyper Prep Kit (Kapa Biosystems; Wilmington, MA).

HyperPlus workflow: Libraries were prepared with the KAPA HyperPlus Library Preparation Kit (Kapa Biosystems), with enzymatic fragmentation for 5 min at 37°C.

Dual-indexed adapter oligos used for both the Hyper Prep and HyperPlus methods were obtained from Integrated DNA Technologies (IDT; Coralville, IA). For both workflows, post-ligation size selection (0.5 – 0.7X) was performed with Agencourt AMPure XP reagent (Beckman Coulter; Beverly, MA). Libraries were amplified for 14 cycles.

NEBNext workflow: Input DNA was digested with NEBNext dsDNA Fragmentase (New England Biolabs; Ipswich, MA) for 32.5 min at 37°C, followed by library preparation with the NEBNext Ultra DNA Library Prep Kit for Illumina. Post-ligation size selection was performed with parameters recommended for an insert size range of 500 – 700 bp. Libraries were amplified for 15 cycles.

Nextera workflow: Libraries were prepared according to the standard protocol, which includes no size selection and 12 cycles of library amplification.

All libraries were quantified after the post-amplification cleanup with the qPCR-based KAPA Library Quantification Kit for Illumina platforms (Kapa Biosystems). Library size distributions were confirmed with a 2100 Bioanalyzer instrument and Agilent High Sensitivity DNA Kit (Agilent Technologies; Santa Clara, CA).

Libraries were normalized and combined into four separate pools for 2 x 300 bp paired-end sequencing on an MiSeq Desktop Sequencer, using a MiSeq Reagent Kit v3 (Illumina; San Diego, CA).

Adapter and quality trimming was performed using Trimmomatic v. 0.30. GC bias was calculated using Picard v. 1.128, and coverage with Bedtools genomecov v. 2.22. For reference genome assembly, reads were trimmed and aligned with BWA MEM v. 0.7.12 and down-sampled to the lowest common number of reads (~900,000). *De novo* assembly was performed using Spades v. 3.5, and metrics collected using Quast v. 2.3.

100K Pathogen Genome Project Workflow

High-molecular weight genomic DNA was extracted from cultured bacterial isolates; and DNA concentration and quality assessed using previously described methods (Kong, et al., Agilent Technologies Application Note 5991-3722EN; Jeannotte, et al., Agilent Technologies Application Note 5991-4003EN). Input into library construction ranged between 200 – 400 ng.

Libraries were prepared with the KAPA HyperPlus Library Preparation Kit according to the manufacturer's instructions. DNA was fragmented enzymatically for 10 min at 37°C. NEXflex-96 DNA Barcodes (Bioo Scientific; Austin, TX) were used for adapter ligation. Dual-SPRI size selection (0.6 – 0.8X) was performed after the post-ligation cleanup. Libraries were amplified with KAPA HiFi HotStart ReadyMix and KAPA Library Amplification Primer Mix for Illumina (Kapa Biosystems), using 8 cycles of amplification.

Library size distributions were confirmed with a 2100 Bioanalyzer instrument and Agilent High Sensitivity DNA Kit. The KAPA Library Quantification Kit for Illumina platforms was used for qPCR-based library quantification, prior to normalization and pooling (96 libraries/lane), for 2 x 100 bp paired-end sequencing on a HiSeq 2500 Sequencer, using v4 HiSeq SBS and Cluster Kits (Illumina; San Diego, CA).

Reads were de-multiplexed and basic quality control performed. *De novo* assembly and annotation were carried out using Abyss and Prokka, respectively.

RESULTS AND DISCUSSION

Comparative Library Construction Metrics

Average yields of purified, amplified libraries for the Hyper Prep and HyperPlus workflows ranged between 190 – 290 ng for all three bacterial species, whereas yields for the NEBNext and Nextera workflows were significantly lower (20 – 150 ng) and more variable (Table 2). When taking the number of amplification cycles into account, the NEBNext workflow performed worst. Yields obtained with the Kapa workflows were much higher than needed for library QC and multiplexed sequencing (theoretically, ~30 ng of each library would have sufficed), indicating that the number of amplification cycles could have been reduced by 2 – 3 cycles for these workflows. Higher consistency across species suggests that the Hyper Prep and HyperPlus workflows are more robust, and better suited for high-throughput pipelines than the NEBNext and Nextera methods.

Table 2. Final library yields

Species and GC content	Average yield (ng) and number of amplification cycles			
	Hyper Prep (14 cycles)	**HyperPlus** (14 cycles)	**NEBNext** (15 cycles)	**Nextera** (12 cycles)
C. difficile (29%)	267	223	85	22
E. coli (51%)	279	290	148	97
B. pertussis (68%)	264	189	25	54
Average (all species)	**270 ± 8**	**237 ± 51**	**43 ± 37**	**58 ± 38**

Electropherograms of final libraries generated with each of the four workflows are given in Figure 2 on the next page. Mode fragment lengths from the electrophoretic analysis vs. mode insert sizes calculated from trimmed, aligned reads are summarized in Table 3.

Fragment lengths determined with the Bioanalyzer for size-selected libraries prepared with ligation-based methods (Hyper Prep, HyperPlus and NEBNext) were within the expected range of 600 – 800 bp, and very similar for all three of the bacterial species. In contrast, Nextera libraries had a mode library fragment length >1 kb, and displayed a wide variation across bacteria. Since long library molecules are not expected to cluster and sequence efficiently, the effective yield of sequenceable library achieved with the Nextera workflow is lower than reflected in Table 2.

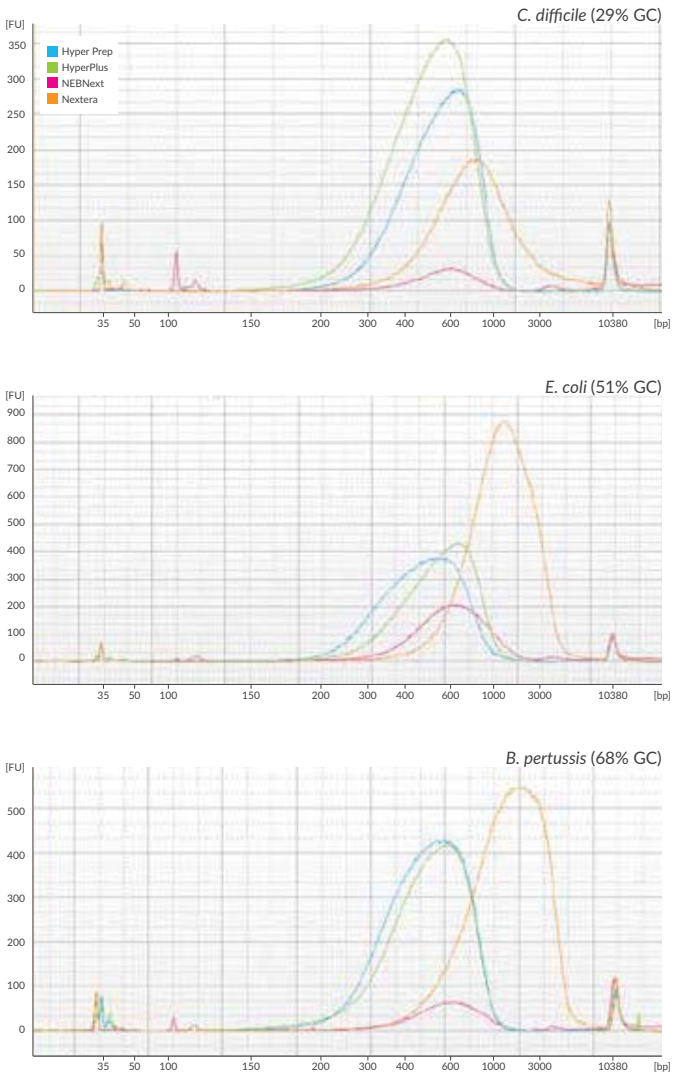


Figure 2. Size distribution of final libraries

Libraries prepared from *C. difficile*, *E. coli*, and *B. pertussis* gDNA using the Hyper Prep, HyperPlus, NEBNext and Nextera workflows were analyzed using a 2100 Bioanalyzer instrument and High Sensitivity DNA Kit (Agilent Technologies). Peak sizes do not correspond with mass-based library yields given in Table 2, as libraries were recovered in different final volumes for different workflows. Nextera libraries were analyzed without dilution; NEBNext libraries were diluted 1/5, whereas Hyper Prep and HyperPlus libraries were diluted to 5 ng/µL for analysis.

Sequencing Metrics

The four library construction methods used in this study were compared with respect to three key sequencing metrics, namely start site bias, GC bias, and coverage uniformity. Biases associated with fragmentation—which has traditionally been a concern with non-mechanical methods—and bias introduced during library amplification are two major factors that impact the depth and uniformity with which genomes are covered. Typically, genomic regions with a balanced GC content are

Table 3. Mode library fragment lengths, determined by electrophoretic analysis (BioA) or from sequence data (Seq).

Species and GC content	Average length (bp)							
	Hyper Prep		HyperPlus		NEBNext		Nextera	
	BioA	Seq	BioA	Seq	BioA	Seq	BioA	Seq
C. difficile (29%)	659	385	712	358	694	478	872	566
E. coli (51%)	650	361	749	383	774	438	1563	444
B. pertussis (68%)	629	361	683	374	788	452	1905	532
Average (all species)	**646**	**369**	**715**	**371**	**752**	**456**	**1447**	**514**
Std dev (bp)	13	11	27	10	41	16	430	51

Mode fragment lengths determined by electrophoretic analysis of final, amplified libraries are inclusive of adapters, whereas mode lengths calculated from sequencing metrics are not.

"easy" to sequence, resulting in surplus coverage for these regions—at the expense of AT- and GC-rich regions, which are underrepresented or often absent from draft genomes.

Start site complexity plots (Figure 3) show the nucleotide content of all aligned reads in a 40-bp window (-10 to +30 bp) relative to the alignment start. As expected, the Hyper Prep workflow—which employed mechanical shearing—displayed the least amount of start site bias for all three bacteria, whereas enzymatic fragmentation methods all displayed varying degrees of start site bias. The KAPA Frag reagent for Enzymatic Fragmentation used in the HyperPlus workflow performed significantly better than Fragmentase (NEBNext workflow) and the tagmentation-based Nextera workflow.

Start site bias potentially impacts library diversity (number of unique reads representing each genome position). Other library construction parameters that impact library diversity are the amount and quality of input DNA (identical for all four methods in this study), and the efficiency with which sequenceable adapter-flanked molecules are generated. Library amplification only creates duplicates (but is necessary to complete adapter sequences and/or generate a sufficient amount of material for QC and sequencing if the input into library construction is low), and can profoundly skew the ratio in which unique adapter-flanked fragments are represented in the final library due to intrinsic biases of DNA polymerases. Besides factors intrinsic to the sequencing platform, these are the primary determinants of coverage depth and uniformity, and ultimately the amount of sequencing that has to be done.

All of the enzymatic fragmentation methods displayed more start site bias than Covaris shearing (the current industry standard). Nevertheless, coverage uniformity plots (Figure 4) and GC bias plots (Figure 5) indicated the following:

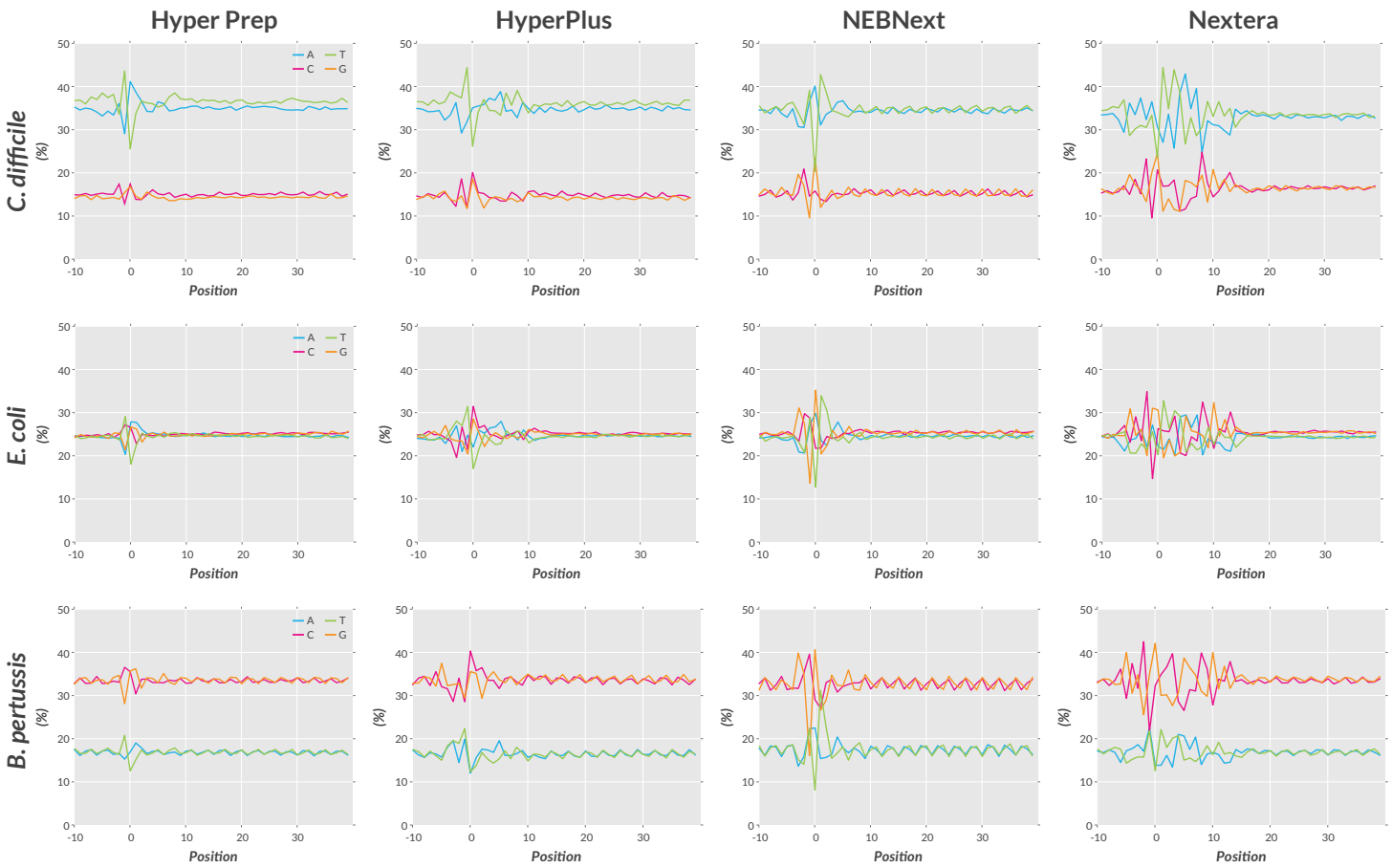


Figure 3. Start site complexity plots

Nucleotide content over a 40 bp window (-10 to +30 bp relative to read alignment start) for *C. difficile* (29% GC), *E. coli* (51% GC), and *B. pertussis* (68% GC), for libraries prepared with the Hyper Prep, HyperPlus, NEBNext and Nextera workflows.

If all three library construction processes (fragmentation, adapter addition, and library amplification) were completely sequence non-specific, each base (A, C, G, and T) would be represented by a perfectly flat, horizontal line with a y-axis value corresponding to the average prevalence of that nucleotide in the genome. For example, the A and T plots for *C. difficile* (29% genomic GC content) would be superimposed, and have a value of ~35% for each position, whereas the C and G plots would both have a value of ~15% at each position.

- Bias associated with enzymatic fragmentation in the HyperPlus workflow had no impact on overall coverage depth and uniformity, or GC bias—which were virtually identical for Hyper Prep and HyperPlus. In the HyperPlus method, minor start site bias is offset by the integrated workflow (which eliminates the physical transfer of material between fragmentation and library construction, and the associated loss of input DNA), and synergy between the fragmentation and library construction chemistries.
- The NEBNext workflow battles with both AT- and GC-rich genomes, whereas Nextera performs poorly with AT-rich sequence—presumably as the result of more biased fragmentation and library amplification. This results in lumpy coverage and coverage hotspots, i.e., over-representation of "easy" (more GC-balanced) regions and under-representation of "difficult" (AT- and GC-rich) regions. With these methods, more sequencing has to be performed to achieve the requisite coverage for these regions, which increases cost and turnaround times.

Integrative Genomics Viewer (IGV) plots (Figure 6 on p. 6) illustrate the impact of different library construction methodologies on a more local level, for specific regions of the genome. Coverage tracks for 7 – 8 kb portions of the *C. difficile* toxin genes (*tcdA* and *tcdB*); and an 11 kb genomic region of *B. pertussis* spanning genes encoding the pertussis toxin, are given in these examples. With the Hyper Prep and HyperPlus workflows, similar and highly uniform coverage was achieved across each region. The NEBNext and Nextera methods yielded a much more uneven distribution of aligned reads, especially for the AT-rich *C. difficile* toxin genes. The NEBNext data has two gaps in the GC-rich *B. pertussis* toxin encoding region, for which virtually no reads are available.

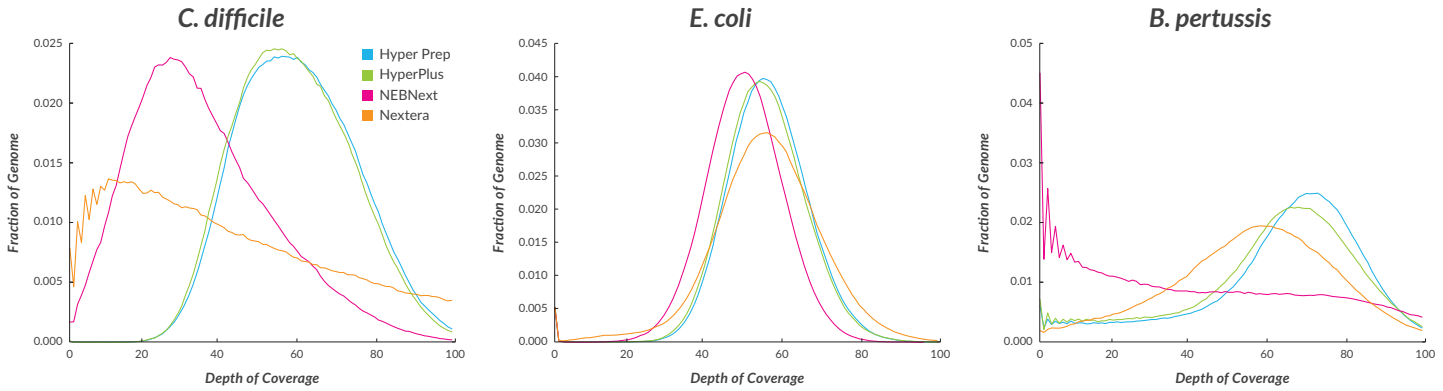


Figure 4. Coverage uniformity plots

Data for all libraries were down-sampled to ~900,000 reads and coverage calculated using Bedtools. The Hyper Prep and HyperPlus workflows yielded highly similar coverage profiles, with a sharp peak and negligible tails for all three bacteria, indicating uniform coverage. In contrast, the NEBNext and Nextera workflows yielded a broader distribution for the genomes with unbalanced GC content, and/or lower mode coverage depth.

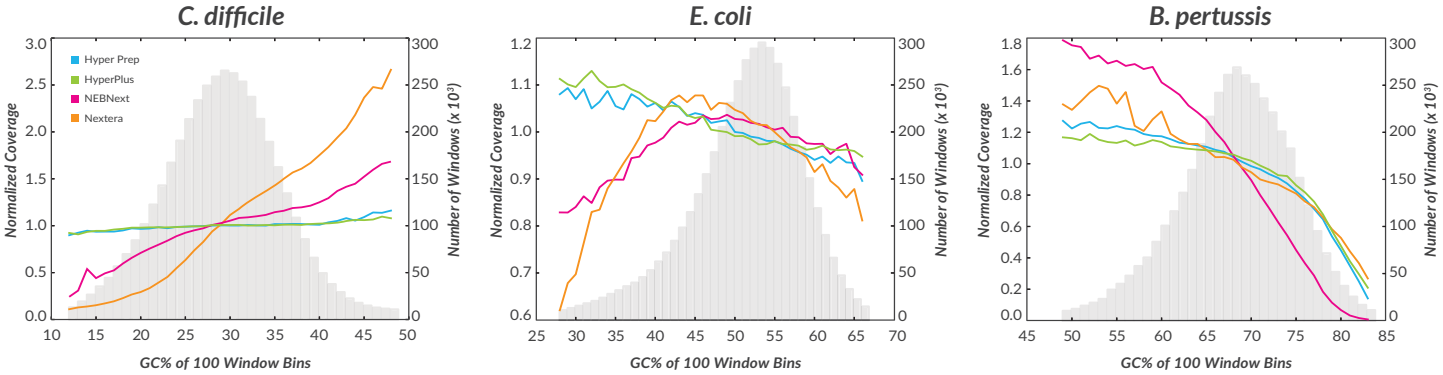
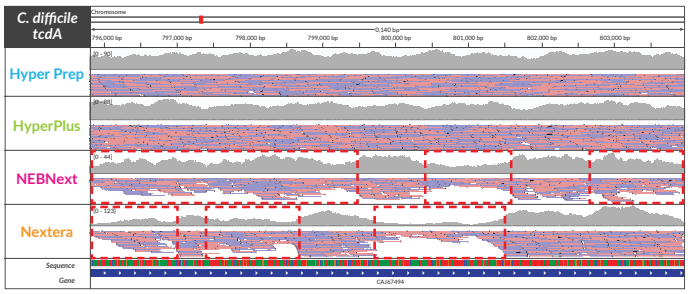
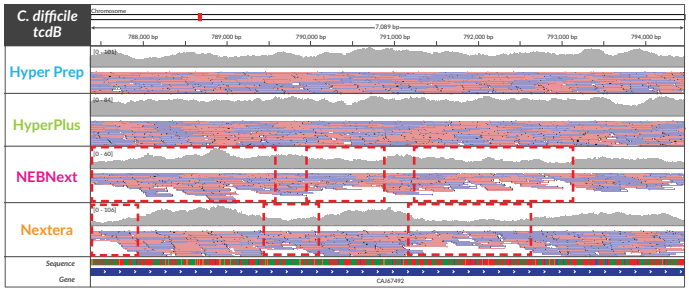


Figure 5. GC-bias plots

Plots were generated with Picard CollectGCBiasMetrics. Gray histograms represent the distribution of genomic GC content for each bacterium, calculated for the reference sequence in 100 bp bins. GC bias was assessed by plotting normalized coverage for each bin—for the Hyper Prep, HyperPlus, NEBNext, and Nextera workflows. If all sample-to-data processes (fragmentation, adapter addition, library amplification, cluster amplification, sequencing, and data analysis) were completely unbiased, all bins would be equally represented, i.e., the plot for each workflow would be a horizontal distribution centered on a normalized coverage of 1.

For *C. difficile*, near-perfect data was obtained for both the Hyper Prep and HyperPlus workflows. In contrast, bins with a more balanced GC content (30 – 50% GC) were over-represented in the NEBNext and Nextera *C. difficile* data, at the expense of bins with extremely low GC content (<30% GC). The NEBNext and Nextera workflows generally performed better in balanced and slightly GC-rich regions (40 – 65% GC), as compared to AT-rich regions. All workflows performed poorly with respect to the extremely GC-rich (>70% GC) bins of *B. pertussis*, where limitations inherent to the sequencing technology start to dominate.





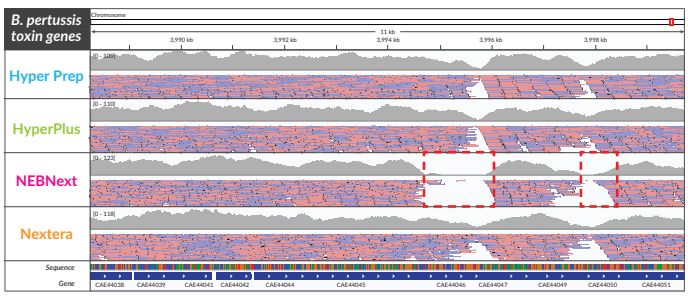


Figure 6. Selected IGV plots

Coverage tracks generated with the Hyper Prep, HyperPlus, NEBNext, and Nextera workflows, for 7 – 8 kb portions of the two *C. difficile* toxin genes, *tcdA* and *tcdB*, and an 11 kb region containing pertussis toxin-encoding genes. Each magenta (read 1) and purple (read 2) line represents a trimmed, aligned read. Gray plots represent coverage depth, whereas the colored track at the bottom of each plot represents the DNA sequence (A = green, C = blue, G = yellow and T = red). Areas of low or lumpy coverage for the AT-rich *C. difficile* toxin genes with the NEBNext and Nextera workflows are highlighted, as are two regions of the GC-rich pertussis toxin genomic region for which virtually no reads were obtained with NEBNext.

De novo Assembly

Microbial WGS often requires *de novo* sequence assembly, e.g., when a reference genome is not available or novel genes are being interrogated. Since longer inserts with a tight size distribution facilitate *de novo* assembly, library construction protocols that provide for tunable fragmentation and size selection are essential. All three of the ligation-based library construction methods used in this study (Hyper Prep, HyperPlus, and NEBNext) met these criteria, whereas the Nextera protocol offered significantly less flexibility.

The four library construction methods were compared with respect to key *de novo* assembly metrics, namely number of contigs, length of longest contig, and N50 length (Figure 7). The Hyper Prep and HyperPlus workflows outperformed NEBNext and Nextera with respect to all metrics.

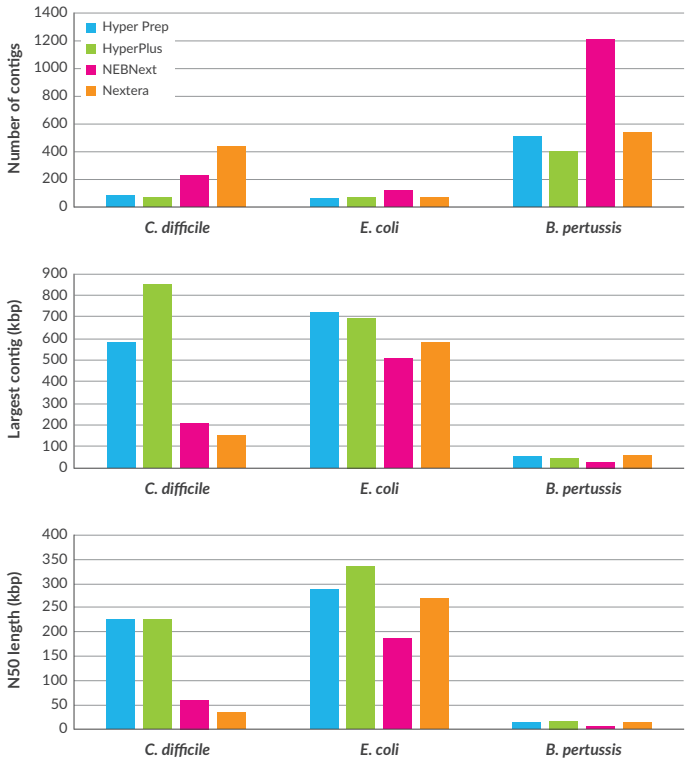


Figure 7. *De novo* assembly metrics

The Hyper Prep, HyperPlus, NEBNext, and Nextera workflows were compared with respect to three key *de novo* assembly metrics. *De novo* assembly is achieved by the appropriate arrangement of overlapping contigs (collections of overlapping reads without gaps). High coverage depth and uniformity, and low bias results in longer and fewer contigs, and longer N50 lengths, which facilitate assembly. The N50 length is a weighted median contig length (50% of the entire assembly is contained in contigs equal to or larger than this value).

The GC-rich *B. pertussis* genome proved to be the most difficult to assemble from data generated with all four methods, with more and shorter contigs and a significantly lower N50 length as compared to the *E. coli* and *C. difficile* genomes.

Despite the fact that the Hyper Prep and HyperPlus methods yielded the shortest mode fragment length (see Table 3), higher and more uniform coverage translated to fewer and longer contigs, and longer N50 lengths, particularly for the AT-rich *C. difficile* genome. The HyperPlus method performed similarly or better than the industry-leading Hyper Prep method across all three assemblies, confirming that minor start site bias associated with enzymatic fragmentation is efficiently offset by other benefits of the integrated workflow.

Meeting the Challenges of Large-scale Genome Projects

The 100K Pathogen Genome Project is an innovative collaboration between the US Food and Drug Administration (FDA); the University of California, Davis; Agilent Technologies, and the Centers for Disease Control and Prevention (CDC). The project aims to create the largest public database of foodborne pathogen draft genomes—to support public health and research activities related to pathogen surveillance and outbreak management; the diagnosis and epidemiology of emerging pathogens, microbial genome variation and evolution, and new gene discovery.

The 100K Project originally selected the KAPA HTP Library Preparation Kit for library construction, due to the high library construction efficiency and coverage uniformity achievable with Kapa's optimally formulated and evolved enzymes, and highly-optimized "with bead" protocol. Approximately 10,000 draft genomes have been assembled to date from libraries prepared with this chemistry. However, mechanical shearing proved to be a major bottleneck in establishing a fast and robust sample preparation pipeline, prompting the transition to the streamlined, fully automatable KAPA HyperPlus workflow, with integrated enzymatic fragmentation.

Initial evaluation of the KAPA HyperPlus workflow focused on the quality and utility of data generated using Kapa's novel enzymatic fragmentation solution instead of Covaris shearing. Routine quality control analysis (not shown) confirmed that the standard workflow with post-ligation size selection produced high-quality sequence data. Slight bias in the nucleotide distribution for the first 10 positions of trimmed reads was observed, but this did not appear to have any impact on the uniqueness or GC content distribution of reads.

Table 4. ***De novo*** **assembly metrics for representative genomes produced by the 100K Pathogen Genome Project**

Genus and species	Read count (PF reads)	Predicted coverage	Average coverage (calculated from assembly)	Predicted genome %GC	Calculated %GC	Estimated genome size (bp)	Assembly size (bp)	Assembled contigs	N50 length	Number of annotated genes
Staphylococcus areus	3,658,490	252	168	32	32	2,900,000	3,098,172	44	195,208	3,055
Staphylococcus areus	3,263,540	225	169	32	32	2,900,000	2,936,030	45	303,273	2,798
Micrococcus sp.	917,292	73	172	N/A	34	2,500,000	2,686,627	41	236,126	2,737
Micrococcus sp.	1,168,170	93	176	N/A	34	2,500,000	2,675,965	44	477,672	2,709
Listeria monocytogenes	2,404,730	161	164	38	36	2,990,000	2,926,186	16	477,671	2,999
Listeria monocytogenes	2,433,480	163	165	38	36	2,990,000	2,915,543	21	475,683	2,983
Vibrio parahaemolytica	6,058,830	233	166	43	44	5,200,000	5,398,332	77	225,233	4,996
Vibrio parahaemolytica	3,200,250	123	167	43	44	5,200,000	5,370,088	72	251,504	4,977
Salmonella bnamdala	4,360,100	194	173	51	50	4,500,000	4,879,757	55	266,584	4,605
Pseudomonas tremae	1,421,780	46	170	66	61	6,200,000	6,846,272	25	832,752	6,293
Pseudomonas tremae	5,258,040	170	174	66	62	6,200,000	6,846,700	26	768,971	6,288
Microbacterium sp.	1,676,190	91	175	65-75	69	3,700,000	3,504,750	13	786,597	3,396
Microbacterium sp.	1,583,760	86	171	65	70	3,700,000	3,502,614	26	251,299	3,396

De novo assembly data for fourteen representative genomes, spanning a genome GC content range from 30 – 70%, are given in Table 4. Average coverage was high (~170X) and very consistent across all the genomes. Calculated GC content correlated extremely well with predicted GC content, where available. The number of assembled contigs ranged between 10 and 80, with *Vibrio* (43% GC) proving the most difficult to assemble. The average N50 length for the set of assemblies (426,813) was significantly (3X) longer than the N50 length achieved in the comparative library construction experiment. With the KAPA HyperPlus workflow, automated liquid handling, and high-throughput QC assays, the turnaround time for 96 samples—from isolated DNA to sequencing-ready pool—has been reduced by approximately 50%, with a concomitant improvement in success rates. A higher degree of multiplexing (384 libraries/lane) is currently being implemented to further optimize coverage and overall sequencing cost.

CONCLUSIONS

The KAPA HyperPlus Library Preparation Kit is the ideal solution for high-throughput microbial whole genome sequencing. The rapid, one-tube protocol is fully automatable; robust across a wide range of genome GC contents; and offers flexibility with respect to the amount of input DNA, library fragment size, adapter design, barcoding strategy, and library amplification—also supporting PCR-free workflows. The combination of a novel, low-bias fragmentation reagent, highly efficient library construction chemistry, and a low-bias amplification enzyme yields high and uniform coverage, thereby facilitating *de novo* assembly and maximizing sequencing cost.

KAPABIOSYSTEMS

Headquarters, United States
Wilmington, Massachusetts
Tel: 781.497.2933
Fax: 781.497.2934
sales@kapabiosystems.com

International Office
Cape Town, South Africa
Tel: +27.21.448.8200
Fax: +27.21.448.6503
sales@kapabiosystems.com

United Kingdom Office
London, England
Tel: +44.845.512.0641
Fax: +44.203.745.5862
uksales@kapabiosystems.com