Application Note

Single-colony whole-genome sequencing

Authors

Adriana Geldart
Sample Prep Support Manager

Rachel W. Kasinskas
Director, Support & Applications

Kapa Biosystems
200 Ballardvale Street, Suite 250
Wilmington, MA, USA

Shawn Higdon
Graduate Student Researcher

Alan Bennett
Distinguished Professor of Plant Sciences

Richard Jeannotte
Assistant Professional Researcher

Department of Plant Sciences,
University of California
One Shields Avenue
Davis, CA, USA

Matthew Settles
Manager, Bioinformatics Core

Bioinformatics Core,
University of California
One Shields Avenue
Davis, CA, USA

Single-Colony Whole-Genome Sequencing of Crude Bacterial Isolates Utilizing the KAPA HyperPlus Kit

Whole-genome sequencing of microbial isolates often involves many laborious steps, as both the propagation and extraction of bacterial isolates are very time-consuming processes. Therefore, implementation of a crude extraction protocol followed directly by library preparation would streamline the sequencing workflow. To enable this process, the library preparation protocol utilized must be robust enough to tolerate inhibitory molecules that may be present in crude lysates. The results described here demonstrate the utility of the KAPA HyperPlus Kit in the preparation of genomic DNA (gDNA) libraries directly from crude cell lysates of both gram-negative and gram-positive bacteria, bypassing the need for liquid culture and DNA extraction. Subsequent sequencing and data analysis confirm that the resulting genome assemblies are comparable to those generated from purified bacterial gDNA.

Introduction

Established pipelines for whole-genome sequencing (WGS) of microbial isolates are cumbersome, involving several steps prior to the preparation of sequencing libraries. These include multiple iterations of plate streaking, sub-culturing in liquid media, extraction of gDNA, and gDNA quality assessment. Preparation of the input DNA sample could potentially be streamlined by omitting the liquid culture and DNA extraction steps, and instead performing a direct mechanical lysis on a single colony cultivated on solid growth medium. The resulting crude cell lysate could then be used as input into library construction. However, such lysates often contain inhibitory molecules that interfere with enzymatic steps in library preparation. To support this, several groups have recently tested whether current library construction workflows are robust enough to overcome the challenges of low-quality gDNA inputs, including crude cell lysates.[1,2]

The objective of this research was to establish a streamlined gDNA sequencing workflow involving the rapid lysis of a single bacterial colony. Using the KAPA HyperPlus Kit—a novel, streamlined workflow for DNA library construction including enzymatic fragmentation—crude cell lysates and purified gDNA extracts from the same set of bacterial isolates were compared as inputs. Our experimental results demonstrate that high-quality libraries can be prepared using the KAPA HyperPlus Kit from individual bacterial colonies immediately following mechanical cell lysis. Further, sequencing results reveal that minimal differences exist between the bacterial genome assemblies that were generated from each workflow. These findings provide a solid methodology that could be implemented in a new workflow for evaluating thousands of bacterial isolates in a streamlined fashion that accommodates high-throughput cell culture coupled with DNA library construction.

Experimental design and methods

The experimental design described here provides a rapid, streamlined method that yields high-quality libraries from crude, single-colony bacterial lysates.

Preparation of crude cell lysates and purified gDNA

Twelve bacterial strains representing a mixture of gram-positive and gram-negative bacteria were selected from a collection of cryopreserved isolates from Mexican corn landrace plants (Table 1). Each strain was identified by Sanger sequencing of the 16S ribosomal RNA gene, using amplicons generated with the universal primers **8F** (5'- AGAGTTTGATCCTGGCTCAG-3') and **1492R** (5'- CGGTTACCTTGTTACGACTT-3').

Each isolate was streaked onto LB-Agar plates directly from cryobank vial storage, and plates were incubated at 28°C for 24 – 48 hours to promote colony formation. For each strain, 3 colonies were used to create crude cell lysates by mechanical cell lysis, and 3 colonies were sub-cultured in liquid medium and used for the isolation of purified gDNA.

To generate crude cell lysates, mechanical lysis was performed using a modified version of the method described in Köser et al., 20131. Briefly, a single colony was resuspended in 40 µL of 10 mM Tris-Cl (pH 8.0) in the presence of acid-washed borosilicate beads (425 – 650 µm in diameter), vortexed at maximum intensity for 10 minutes, and centrifuged to remove cellular debris and beads; the supernatant (the crude cell lysate) was then transferred to a fresh tube.

For conventional gDNA extraction, sub-culturing in liquid medium is necessary to obtain sufficient cell biomass. Therefore, independent colonies were inoculated in 5 mL of LB liquid media, followed by incubation at 28°C for 24 – 48 hours. Growth media was removed by centrifugation, and gDNA was extracted from the cell pellets using the Ultraclean® Microbial DNA Extraction Kit (Mo Bio Laboratories).

The DNA content of both the crude cell lysates and the purified gDNA samples was assessed via the Qubit™ dsDNA Broad Range Quantification Kit on the Qubit 2.0 Fluorometer (ThermoFisher).

Generation of sequencing libraries using the KAPA HyperPlus Kit

For each isolate, biological triplicate libraries were generated from both crude cell lysates and purified gDNA. In each case, 30 ng of material was used as input into the KAPA HyperPlus Kit, and libraries were prepared according to the KAPA HyperPlus Kit Technical Data Sheet. Enzymatic fragmentation was performed for 9 minutes at 37°C to obtain an average insert size of 400 – 500 bp. Custom dual-indexed adapters were obtained from Integrated DNA Technologies and ligated to the fragmented DNA at an adapter:insert ratio of 200:1. Post-ligation double-sided bead-based size selection (0.5X/0.7X) was performed with Agencourt® AMPure® XP (Beckman Coulter).

Table 1. Bacterial isolates used in this study

Isolate	Genus	Species/Accession*	Cell Wall Type
A	*Pantoea*	*dispersa*	Gram Negative
B	*Citrobacter*	*freundii*	Gram Negative
C	*Serratia*	*liquefaciens*	Gram Negative
D	*Pseudomonas*	*putida*	Gram Negative
E	*Stenotrophomonas*	*maltophilia*	Gram Negative
F	*Rahnella*	*aquatilis*	Gram Negative
G	*Klebsiella*	*oxytoca*	Gram Negative
H	*Sphingomonas*	*pseudosanguinis*	Gram Negative
I	*Micrococcus*	*CZBRD3*	Gram Positive
J	*Rhodococcus*	*jialingiae*	Gram Positive
K	*Curtobacterium*	*oceanosedimentum*	Gram Positive
L	*Staphylococcus*	*warneri*	Gram Positive

*As identified by 16S sequencing.

Prior to library amplification, library concentrations were assessed to determine the appropriate number of cycles to obtain 250 – 500 ng of final library (to ensure sufficient material for downstream processing). At this stage and following amplification and cleanup, library concentrations were determined with the KAPA Library Quantification Kit using the StepOne Plus Real-Time PCR System (ThermoFisher). Final library size distribution was assessed with a 2100 Bioanalyzer instrument and using the High Sensitivity DNA Kit (Agilent Technologies). The overall workflow is depicted in Figure 1.

Sequencing and data analysis

Paired-end 150 bp sequencing was performed on an Illumina® HiSeq 4000 by the UC Davis DNA Technologies Core Facility. Raw sequence fastq files were preprocessed by first identifying and removing PhiX spike-in sequences using bowtie2 (v2.2.8), followed by removal of PCR duplicate reads with Super Deduper (github.com/dstreett/Super-Deduper); trimming of poor-quality 5' and 3' ends with sickle (github.com/najoshi/sickle); and overlapping and adapter removal using FLASH2 (github.com/dstreett/FLASH2). Reads shorter than 50 bp were then discarded. Resulting paired- and single (merged) reads were assembled using the spades assembler (v3.9.0, skipping error correction), and then mapped back to the assembly using bwa mem (v0.7.13) in order to generate additional assembly statistics.

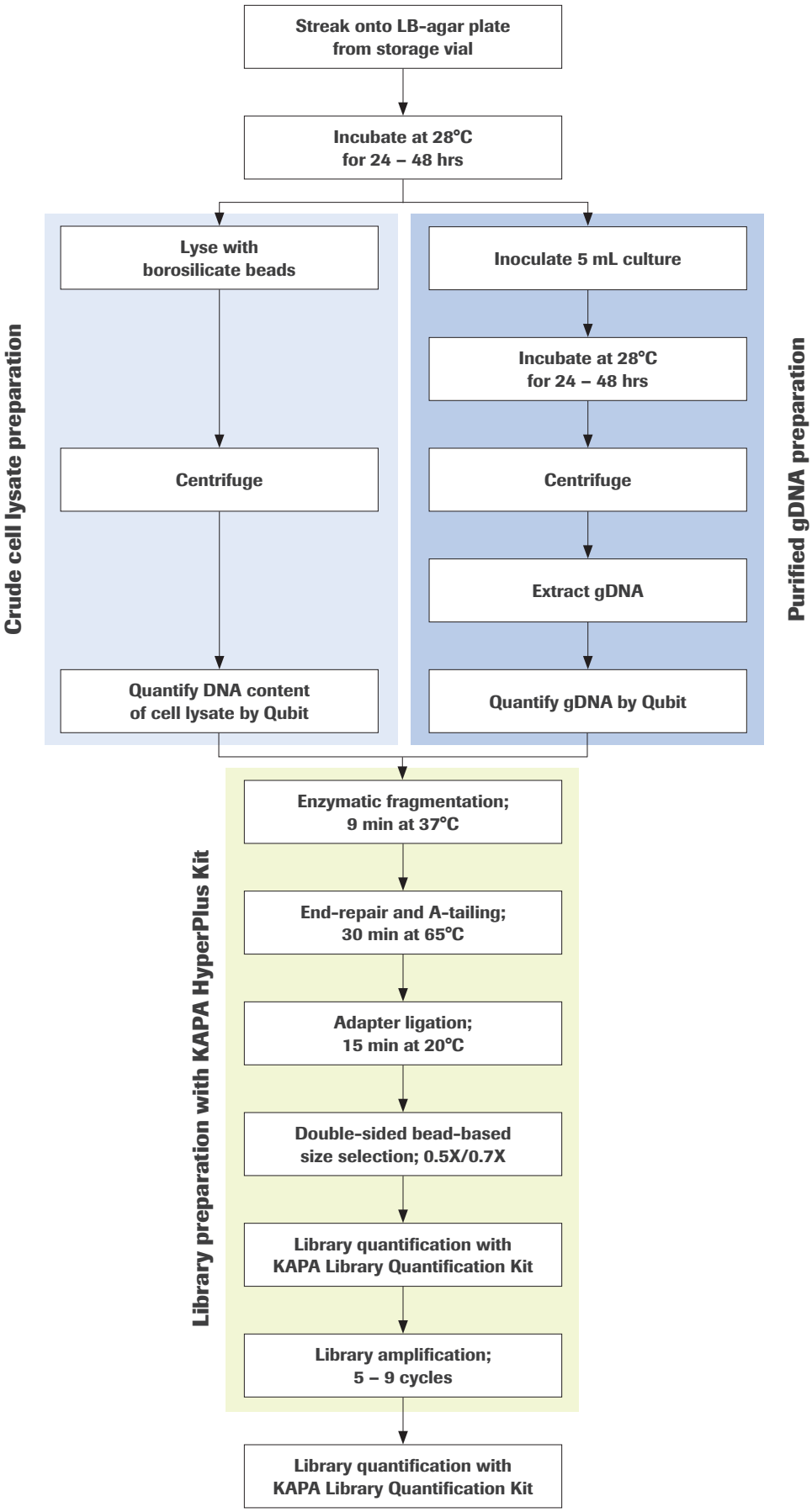

Figure 1: Workflow schematic

Results and discussion

To evaluate the efficiency of library construction, post-size-selection concentrations were quantified via qPCR using the KAPA Library Quantification Kit. With the exception of one bacterial isolate, the crude cell lysates produced library concentrations roughly equal to or greater than concentrations produced by the purified gDNA inputs (Figure 2). The concentrations of libraries at this step were then used to determine the number of amplification cycles (between 5 – 9 cycles) needed to generate sufficient material for sequencing and archiving (250 – 500 ng) while minimizing over-amplification. After amplification, final library yields were quantified with the KAPA Library Quantification Kit (Table 2).

Table 2: Average concentration of final libraries

Isolate	Average library Purified gDNA (nM)*	Average library Crude Cell Lysate (nM)*
A	65.1 (SD 11.1)	42.1 (SD 20.1)
B	49.2 (SD 25.9)	37.9 (SD 5.3)
C	43 (SD 4.2)	37.2 (SD 6.2)
D	46 (SD 10.5)	26.4 (SD 20)
E	68.8 (SD 7.3)	25.3 (SD 5.8)
F	45.4 (SD 3.1)	26.2 (SD 6.6)
G	61.5 (SD 27.1)	23.8 (SD 6.4)
H	50.3 (SD 45.4)	26.2 (SD 14.8)
I	42.2 (SD 11.2)	21.5 (SD 12.5)
J	43 (SD 11.0)	47 (SD 3.9)
K	45 (SD 11.4)	17.4 (SD 4.6)
L	58.3 (SD 7.8)	22.3 (SD 9.9)

*Concentrations were determined using the KAPA Library Quantification Kit.

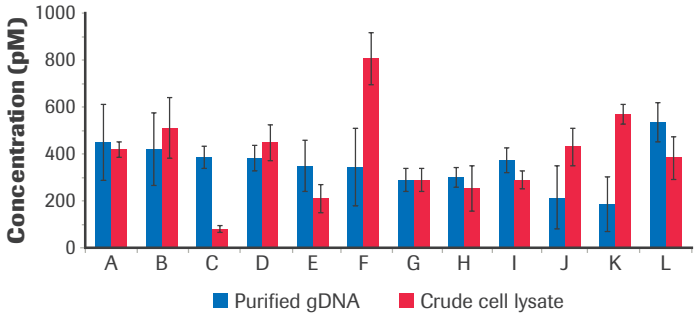

Figure 2: Quantification of adapter-ligated libraries prior to amplification. Concentrations are shown for libraries prepared from purified gDNA (blue) and crude cell lysates (red) for each strain tested. Concentrations were determined using the KAPA Library Quantification Kit.

To determine whether the two DNA preparation methods lead to any bias in the size of the final libraries due to the potential impact of inhibitors in the crude cell lysate, library size distribution was assessed. Electropherograms are shown for libraries created from both purified gDNA and crude cell lysate for two representative bacterial isolates, Isolate C (*S. liquefaciens*) and Isolate D (*P. putida*) (Figure 3). Mean peak sizes are between 485 – 620 bp, consistent with the fragmentation conditions used and the double-sided bead-based size selection protocol that was performed. Size distributions are similar between libraries prepared from purified gDNA and crude cell lysates, with the exception of isolate D where the mean insert size was slightly larger for the crude cell lysate preparation.

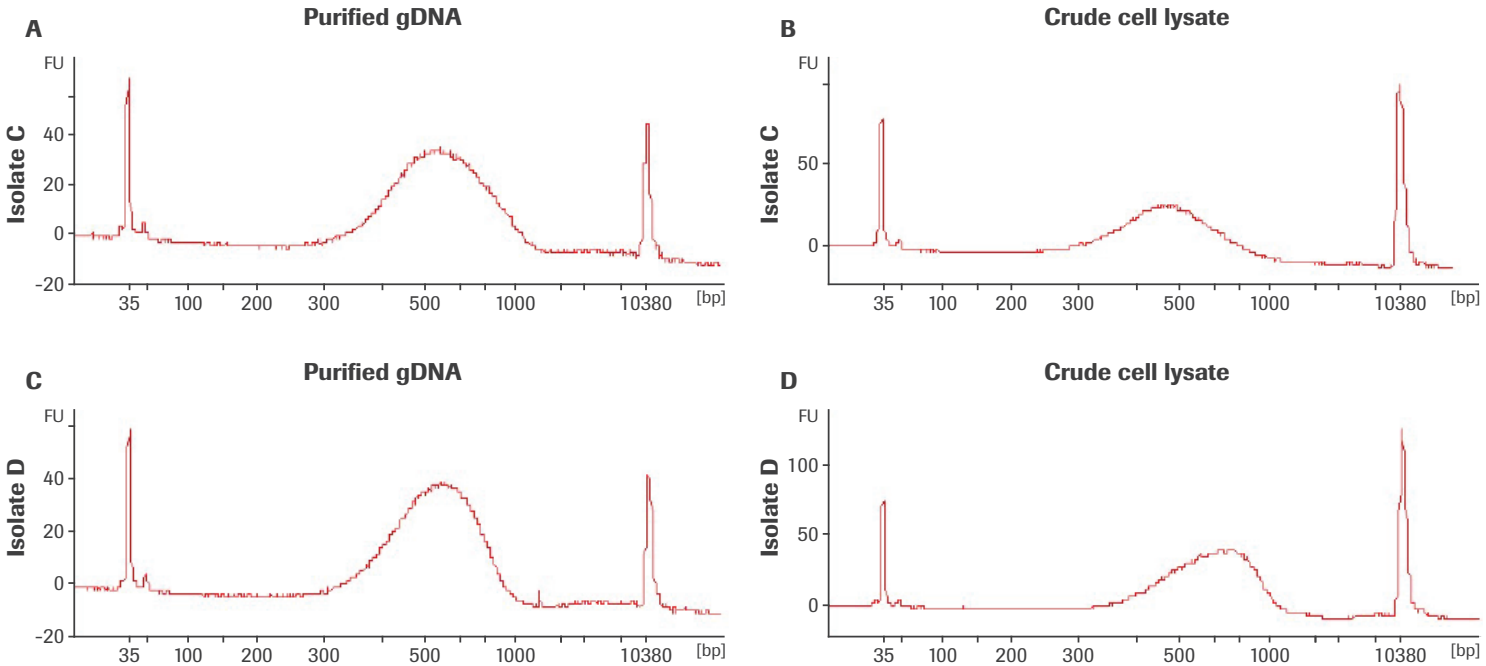

Figure 3: Size distribution of final libraries. Libraries prepared from two representative bacterial isolates for both purified gDNA and crude cell lysates are shown: Isolate C, **(A)** purified gDNA and **(B)** crude cell lysate; Isolate D, **(C)** purified gDNA and **(D)** crude cell lysate. Analysis was performed using a 2100 Bioanalyzer Instrument and a High Sensitivity DNA Kit (Agilent Technologies).

Table 3: *De novo* assembly metrics for both purified gDNA and crude cell lysate library preparations

Isolate	Sample type	Number of reads assembled	Average coverage	Calculated % GC	Assembly size (kbp)	Assembled contigs	N50 length (kbp)
A	Purified	5,770,877	31	58	4,727.7	33	500.4
	Crude	4,530,187	35		4,729.5	34	500.6
B	Purified	5,339,707	25	52	4,990.1	28	608.7
	Crude	4,876,363	34		4,991.7	30	608.7
C	Purified	6,071,137	35	55	5,279.3	39	573.1
	Crude	4,035,264	30		5,279.6	41	573.1
G	Purified	6,698,683	30	56	5,511.5	42	554.8
	Crude	5,378,585	30		5,530.8	76	425.9
I	Purified	5,386,721	46	73	2,592.4	66	152.2
	Crude	5,616,599	67		2,495.7	64	152.2
L	Purified	5,208,844	57	33	2,535.7	18	530.9
	Crude	5,416,857	62		2,542.1	30	507.9

Upon analysis of the sequencing results, several of the isolates were found to contain multiple organisms, or to have been incorrectly classified based on the original 16S sequence results (i.e., only 15% of reads map to the reference). These isolates were thus excluded from further analysis. The remaining 6 isolates represent both gram-positive and gram-negative bacteria, and a wide range of GC content (33% – 73%).

Libraries prepared from crude cell lysates and purified gDNA inputs were compared with respect to key *de novo* assembly metrics: number of contigs; length of longest contig; and N50 length (a weighted median contig length where 50% of the entire assembly is contained in contigs equal to or larger than the value represented) (Figure 4). Fewer, longer contigs and a greater N50 length are the most desirable as they readily facilitate downstream analysis. With the exception of Isolates G and L, the number of contigs is similar for libraries created with crude cell lysates and with purified gDNA, and the longest contig and N50 lengths are indistinguishable between the two input types.

Additional *de novo* assembly data for selected isolates are shown in Table 3, including the number of reads assembled, average coverage, GC content, and assembly size. Values for all metrics are similar for both purified and crude cell lysate library preparations. It should be noted that the most GC-rich isolate (Isolate I) resulted in the poorest assembly as compared to all samples investigated regardless of the input type, with a greater numer of contigs, shorter average contig size, and a significantly lower N50 length than the remaining isolates.

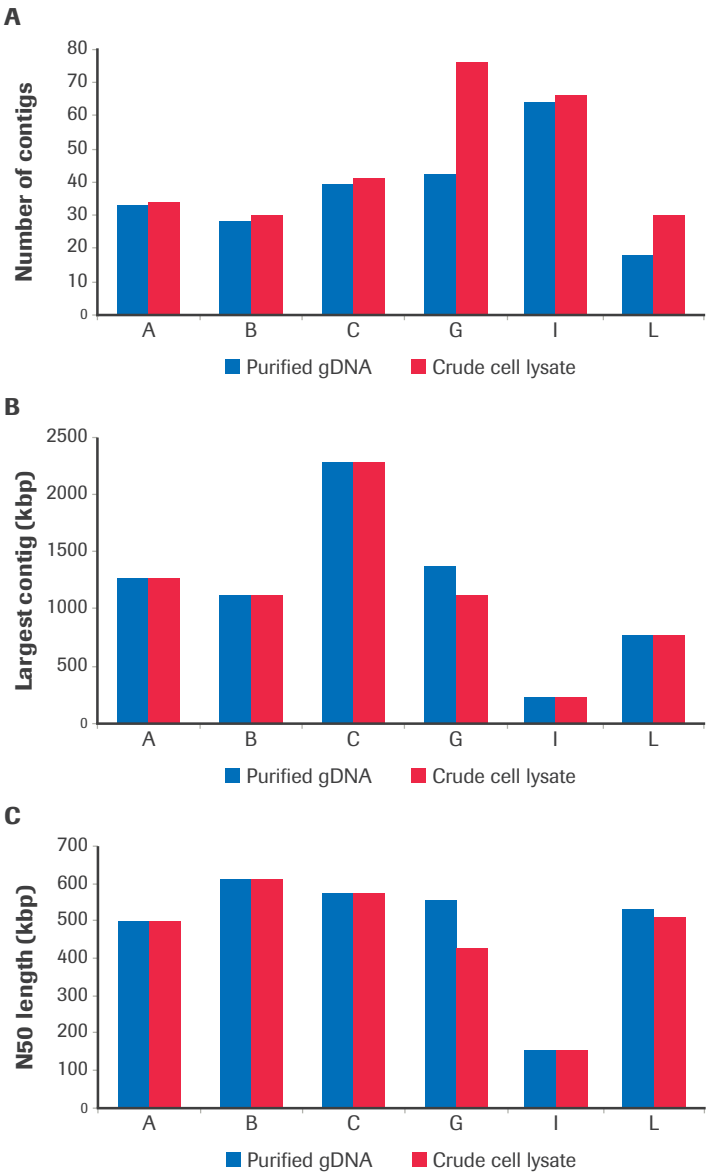

Figure 4. Key *de novo* assembly metrics for positively identified strains. **(A)** Number of contigs; **(B)** Longest contig; and **(C)** N50 length.

Conclusions

The KAPA HyperPlus Kit is a robust solution for the rapid generation of high-quality sequencing libraries directly from mechanically lysed bacterial colonies, and is effective across a wide range of GC content and bacterial types. The libraries and genome assemblies produced from crude cell lysates are highly comparable to those generated using purified gDNA as input. These results establish the KAPA HyperPlus Kit as a rapid bacterial genome sequencing workflow that eliminates the need for conventional, time-consuming bacterial culturing techniques.

References

1. Köser, Claudio U., et al., "Rapid single-colony whole genome sequencing of bacterial pathogens." *Journal of Antimicrobial Chemotherapy* **69.5** (2014): 1275 – 1281.
2. Wright, Meredith S., et al., "SISPA-Seq for rapid whole genome surveys of bacterial isolates." *Infection, Genetics and Evolution* **32** (2015): 191 – 198.

Published by:

Roche Sequencing & Life Science
9115 Hague Road
Indianapolis, IN 46256

sequencing.roche.com

Data on file.
For Research Use Only. Not for use in diagnostic procedures.
PP-US-11859
KAPA is a trademark of Roche. All other product names and trademarks are the property of their respective owners.

AP109005 A250 06/17