Roche

Application Note

Crude sample bacterial whole-genome sequencing

Authors

Adriana Geldart
Sample Prep Support Manager

Rachel W. Kasinskas
Director, Support & Applications

Roche Sequencing Solutions
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
Davis, CA, USA

Matthew Settles
Manager, Bioinformatics Core

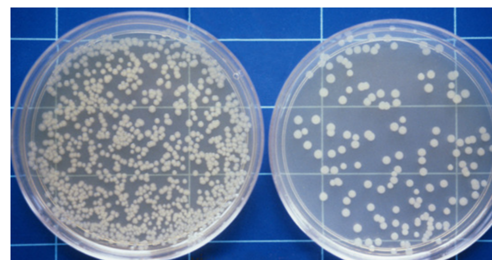
Bioinformatics Core,
University of California
Davis, CA, USA

Date of first publication:
June 2017

The KAPA HyperPlus Kit enables a streamlined bacterial whole-genome sequencing workflow, utilizing crude lysates from single colonies

Whole-genome sequencing of bacterial isolates typically involves many time-consuming and laborious steps, including bacterial cell culture and the purification of genomic DNA (gDNA). A crude DNA extraction protocol, followed directly by library construction would significantly streamline the sequencing workflow. To enable this process, the library construction method must be robust enough to tolerate inhibitors present in crude lysates. Here, we demonstrate the utility of the KAPA HyperPlus Kit for the preparation of gDNA libraries directly from single-colony crude cell lysates of both gram-negative and gram-positive bacteria, bypassing the need for liquid culture and DNA purification. Sequencing and data analysis confirmed that the resulting genome assemblies are comparable to those generated from purified bacterial gDNA.

Introduction



Established pipelines for whole-genome sequencing (WGS) of bacterial isolates are cumbersome, and require several steps prior to the construction of sequencing libraries. These include multiple iterations of plate streaking, sub-culturing in liquid media, purification of gDNA, and gDNA quality assessment. Preparation of the input DNA for library construction could potentially be streamlined by omitting the liquid culture and DNA purification steps, and instead obtaining DNA directly from a crude lysate of a single colony cultivated on a solid growth medium. However, such lysates may contain inhibitors that interfere with the enzymatic steps in library preparation.

The objective of this study was to establish a streamlined gDNA sequencing workflow employing mechanical lysis of a single bacterial colony. Using the KAPA HyperPlus Kit—a novel, streamlined workflow for DNA library construction with integrated enzymatic fragmentation—crude cell lysates and purified gDNA extracts from the same set of bacterial isolates were compared as inputs. Results demonstrate that high-quality libraries can be prepared with the KAPA HyperPlus Kit from single-colony crude lysates, and that minimal differences exist between the bacterial genome assemblies generated using the traditional and streamlined workflows. Unlike related studies, in which a similar approach was attempted with Nextera® library preparation kits (Illumina®) no additional purification of extracted DNA was needed.[1,2]

The streamlined workflow eliminates 2 – 3 days in overall turnaround time, as well as the costs associated with bacterial sub-culturing and DNA purification. This offers new possibilities in diagnostic microbiology, particularly in scenarios such as pathogen outbreaks and the management of hospital-acquired infections by antibiotic resistant bacteria, where the rapid identification of bacterial strains are critical.

KAPABIOSYSTEMS

For Research Use Only. Not for use in diagnostic procedures.

Materials and methods

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 previously classified by Sanger sequencing of a part of the 16S ribosomal RNA gene, employing 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 the cryobank storage vial, and plates were incubated at 28°C for 24 – 48 hours to promote colony formation. For each strain, three colonies were used to generate crude cell lysates by mechanical cell lysis, and three colonies were sub-cultured in liquid medium and used for the purification of gDNA.

To generate crude cell lysates, mechanical lysis was performed using a modified version of the method described by Köser *et al.*, 2013. Briefly, each 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 purification, sub-culturing in liquid medium is necessary to obtain sufficient cell biomass. Therefore, 5 mL aliquots of LB liquid media were inoculated with independent colonies, followed by incubation at 28°C for 24 – 48 hours. The growth medium was removed by centrifugation, and gDNA was extracted and purified 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 with the Qubit™ dsDNA Broad Range Quantification Kit, using a Qubit 2.0 Fluorometer (ThermoFisher).

Construction of sequencing libraries with 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 DNA (as determined using the Qubit assay) was used as input, and libraries were prepared according to the instructions in the KAPA HyperPlus Kit Technical Data Sheet. Enzymatic fragmentation was performed for 9 minutes at 37°C to obtain average insert sizes 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 reagent (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 classified by Sanger sequencing of a part of the 16S rRNA gene.

Prior to library amplification, libraries were quantified by qPCR to determine the number of cycles needed to obtain 250 – 500 ng of final library (the amount of material required for downstream processing). At this stage, and after the post-amplification cleanup, library concentrations were determined with the KAPA Library Quantification Kit, using a StepOne Plus Real-Time PCR System (ThermoFisher). Final library size distributions were assessed with a 2100 Bioanalyzer instrument and a High Sensitivity DNA Kit (Agilent Technologies). Both the conventional and streamlined workflows are depicted in Figure 1.

Sequencing and data analysis

Paired-end sequencing (2 x 150 bp) was performed by the by the UC Davis DNA Technologies Core Facility, using an Illumina® HiSeq 4000 instrument. 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 removal of overlapping and adapter sequences 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.

Data on file.
For Research Use Only. Not for use in diagnostic procedures.

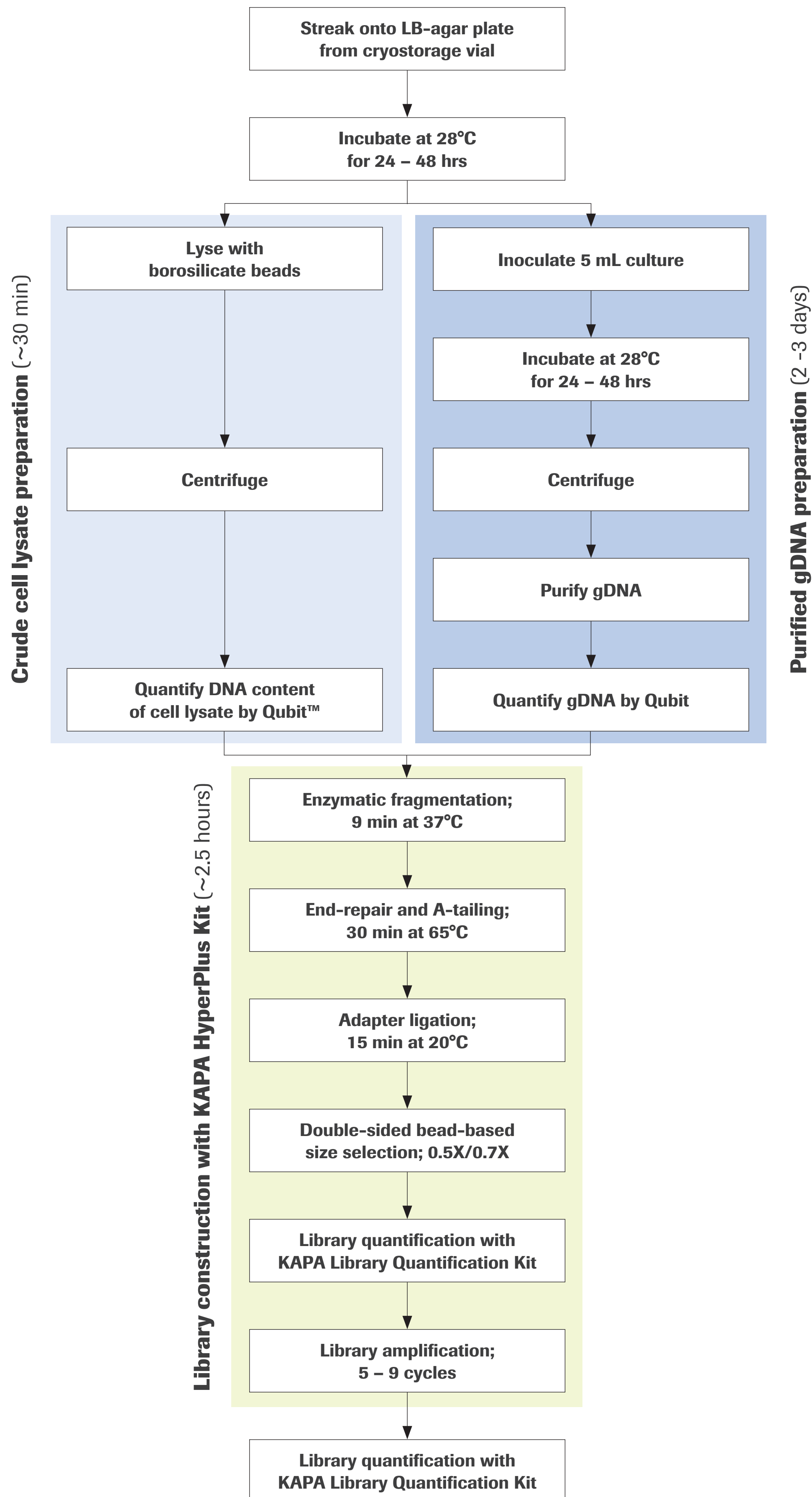


Figure 1: Overview of the streamlined and conventional workflows used in this study. The library construction time does not include qPCR-based quantification with the KAPA Library Quantification Kit. Each quantification assay takes approximately 3 hours to complete. Post-ligation quantification may be omitted once the workflow has been optimized.

Data on file.
For Research Use Only. Not for use in diagnostic procedures.

Results and discussion

To evaluate the efficiency of library construction, post-size-selection library concentrations were determined by qPCR, using the KAPA Library Quantification Kit. With the exception of one bacterial isolate, crude cell lysates produced library concentrations roughly equal to or greater than those obtained from purified gDNA inputs (Figure 2). The concentrations of libraries at this stage of the process were 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 also assessed with the KAPA Library Quantification Kit (Table 2).

Library quantification suggested that potential inhibitors in crude cell lysates did not appear to have any significant impact on the success of library construction with the KAPA HyperPlus Kit. To confirm this, library size distribution was also assessed. Electropherograms are shown for libraries constructed from both purified gDNA and crude cell lysates, for two representative bacterial isolates, namely Isolate C (*S. liquefaciens*) and Isolate D (*P. putida*) (Figure 3). Mean peak sizes were between 485 – 620 bp, consistent with the fragmentation conditions and the double-sided bead-based size selection parameters that were used. Size distributions were similar for libraries prepared from purified gDNA and crude cell lysates, with the exception of Isolate D, for which the mean insert size was slightly larger for the crude cell lysate.

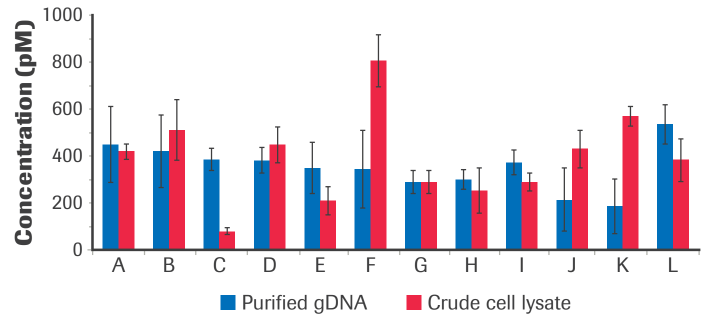


Figure 2: Quantification of adapter-ligated libraries prior to amplification. Concentrations are shown for libraries prepared from purified gDNA (blue) and crude cell lysates (red), for each of the 12 strains used in this study. Concentrations were determined with the KAPA Library Quantification Kit.

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 with the KAPA Library Quantification Kit.

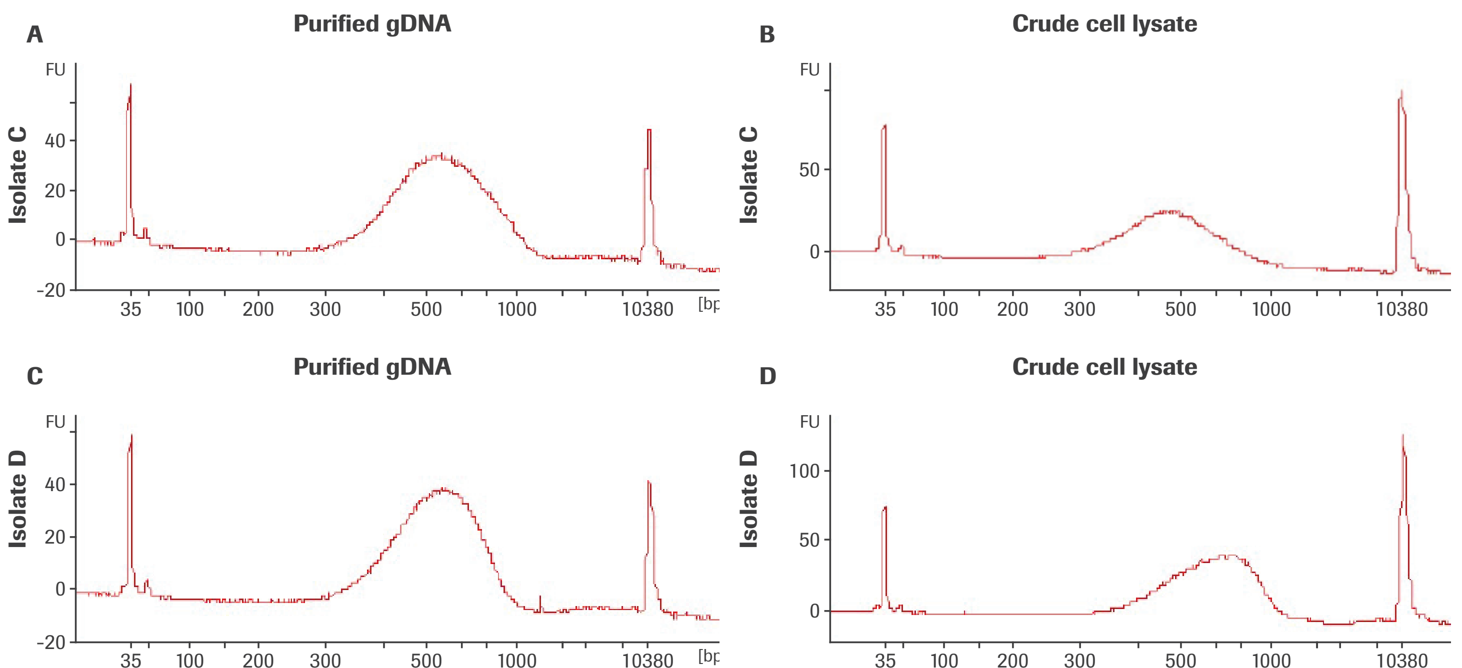


Figure 3: Size distributions of final libraries for representative isolates. Libraries prepared with both workflows are shown for Isolate C, **(A)** purified gDNA and **(B)** crude cell lysate; and Isolate D, **(C)** purified gDNA and **(D)** crude cell lysate. Analysis was performed using a 2100 Bioanalyzer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

Data on file.
For Research Use Only. Not for use in diagnostic procedures.

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 (only ~15% of reads mapped to the reference). These isolates were thus excluded from further analysis. The remaining six isolates represented both gram-positive and gram-negative bacteria, and a wide range of genomic GC content (33% – 73%).

Libraries prepared from crude cell lysates and purified gDNA inputs were compared with respect to key *de novo* assembly metrics, namely the 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 facilitate downstream analysis. With the exception of Isolates G and L, the number of contigs were similar for libraries created with crude cell lysates and with purified gDNA, and the longest contig and N50 lengths were 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 were similar for libraries prepared from purified gDNA *vs.* crude cell lysates. It should be noted that the most GC-rich isolate (Isolate I, 73% GC) resulted in the poorest assembly, regardless of the input type; with a greater number of contigs, shorter average contig size, and a significantly lower N50 length as compared to the other five isolates.

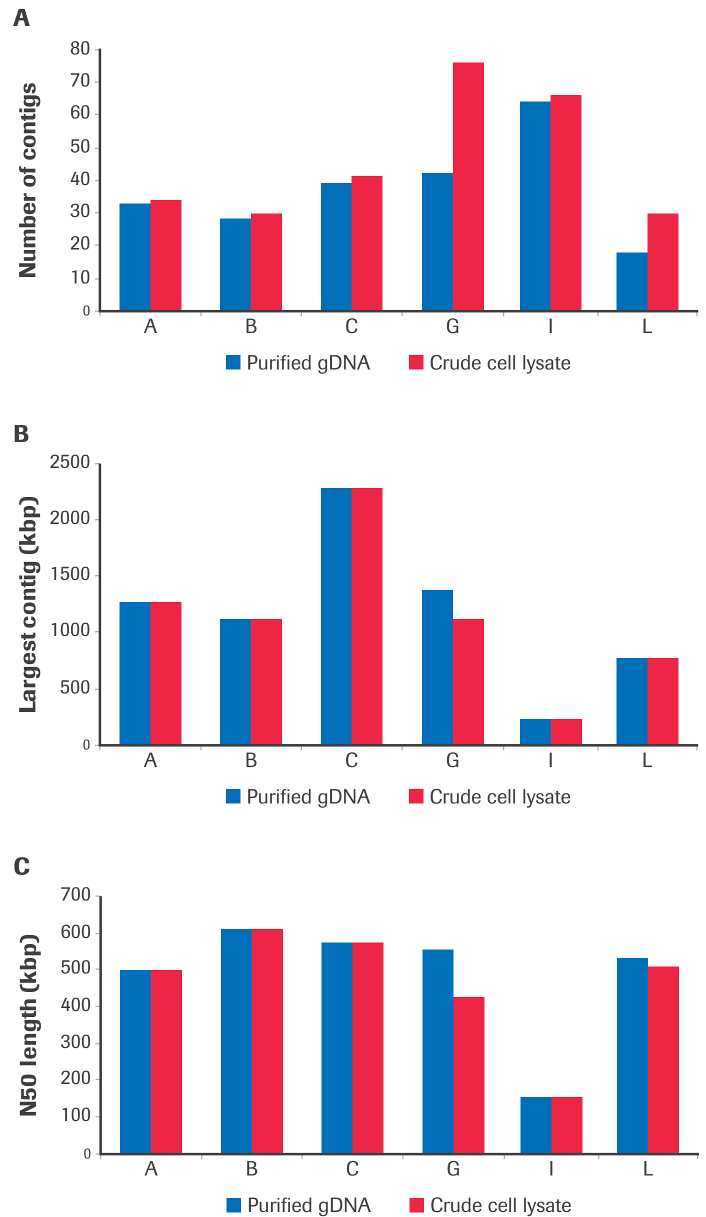


Figure 4. Key *de novo* assembly metrics for positively identified strains. **(A)** Number of contigs; **(B)** Longest contig; and **(C)** N50 length.

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

Data on file.
For Research Use Only. Not for use in diagnostic procedures.

Conclusions

The KAPA HyperPlus Kit enables the preparation of high-quality whole-genome sequencing libraries directly from mechanically lysed single bacterial colonies, and is effective across a wide range of bacterial types with divers genomic GC content. The libraries and genome assemblies produced from crude cell lysates are highly comparable to those generated using conventional workflows that require bacterial sub-culturing and the purification of gDNA. The streamlined and cost-effective crude sample KAPA HyperPlus workflow potentially expands the scenarios and settings in bacterial which whole-genome sequencing may be applied.

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 Solutions, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588

sequencing.roche.com

Data on file.
For Research Use Only. Not for use in diagnostic procedures.
KAPA is a trademark of Roche. All other product names and trademarks are the property of their respective owners.
SEQ100142 09/17
© 2017 Roche Sequencing Solutions, Inc. All rights reserved