Automated Library Construction Using KAPA Library Preparation Kits on the Agilent NGS Workstation Yields High-Quality Libraries for Whole-Genome Sequencing on the Illumina Platform

Application Note

Authors

Nguyet Kong, Kao Thao, Carol Huang, and Bart C. Weimer
Population Health and Reproduction Department, School of Veterinary Medicine, University of California-Davis, Davis, CA, USA

Maryke Appel
Kapa Biosystems, Inc.
Wilmington, MA USA

Stephen Lappin, Lisa Knapp, and Lenore Kelly
Agilent Technologies, Inc.
Santa Clara, CA, USA

Abstract

A new method was developed to automate the KAPA HTP Library Preparation kit for microbial whole genome sequencing. This method uses the Agilent NGS Workstation, consisting of the NGS Bravo liquid handling platform with its accessories for heating, cooling, shaking, and magnetic bead manipulations in a 96-well format. User intervention in multistep protocols is minimized through the use of other components of the workstation such as the BenchCel 4R Microplate Handler and Labware MiniHub for labware storage and movement. This method has been validated for sequencing on the Illumina platform and consists of three protocols: the first is for end repair to post-ligation cleanup; the second is used for library amplification setup; and the third is for the post-amplification cleanup. The modular design provides the end-user with the flexibility to complete library construction over two days, and is suitable for the construction of high-quality libraries from bacteria of various GC content. This combined solution produced a workflow that is suitable for production-scale sequencing projects such as the 100K Pathogen Genome Project.

Agilent Technologies

Introduction

Reduced costs and higher throughput have rendered microbial whole genome sequencing (WGS) accessible to many applications in infectious disease, food safety, and public health to produce genomes on an unprecedented scale. The 100K Pathogen Genome Project hosted at UC Davis and founded by Agilent Technologies, the FDA, and UC Davis, is a novel public/private/government consortium to sequence 100,000 bacterial zoonotic and food-borne pathogens that are of significant importance to the public, using next generation sequencing (NGS) technologies. The project has sufficient sequencing capacity to produce as many as 25,000 bacterial genomes per year using automated workflows. This capacity in sequencing created a challenge to establish an automated library construction workflow to fully enable the reliable and robust sequencing pipeline.

NGS library construction comprises repetitive processes, making it amenable to automation. Automation offers many advantages, including sample throughput, reduced hands-on time, greater reproducibility, and improved process control. However, the development and validation of automated protocols is not trivial. Optimal workflows require robust chemistry, robotics, and meticulous optimization of many parameters to achieve acceptable libraries that yield sequence data of quality similar to those produced by skilled, experienced, and attentive technicians in a low-throughput setting.

The NGS Workstation platform offers reliable and reproducible automation of library preparation processes. With minimal user-intervention to produce robust libraries, the graphical user interface of the Agilent VWorks Automation Control Software makes the system easy to operate in a production setting, while the open architecture of the software facilitates new method development and customization to optimize specific steps of the protocol.

Library preparation kits from Kapa Biosystems offer robust library construction methods for a range of DNA inputs (100 pg–5 µg) for a variety of sequencing applications, including WGS, targeted sequencing, ChIP-Seq, RNA-Seq, and Methyl-Seq. Reagents are formulated for optimal activity and stability, and exhibit excellent conversion rates of input DNA to adapter-ligated libraries through the use of a highly optimized, automation-friendly protocol using beads [1]. Kits contain the engineered KAPA HiFi DNA Polymerase, which has become widely accepted for high-efficiency, high-fidelity, low-bias NGS library amplification [2,3,4,5]. These reagents were selected by the 100K Pathogen Genome Project for automated library construction due to the excellent sequence results in the initial phase of the project using manual approaches. However, increased capacity required automation of this protocol to meet the high-throughput demands. This application note details the automated production of high quality bacterial libraries that are reproducible and provide excellent sequencing results using the KAPA HTP Library Preparation Kit on the Agilent NGS Workstation for use in the 100K Pathogen Genome Project.

Materials and Methods

The 100K Pathogen Genome Project sample preparation workflow for multiplexed, paired-end (2 × 100 bp) Illumina sequencing libraries uses accepted methods for bacterial sequencing projects (Figure 1). Collection of high molecular weight genomic DNA is critical for successful library construction. Optimization of the cell lysis and DNA isolation is described in previous documentation [6]. High-throughput QC of the high molecular weight genomic DNA using the Agilent 2200 TapeStation System is also documented elsewhere [7].

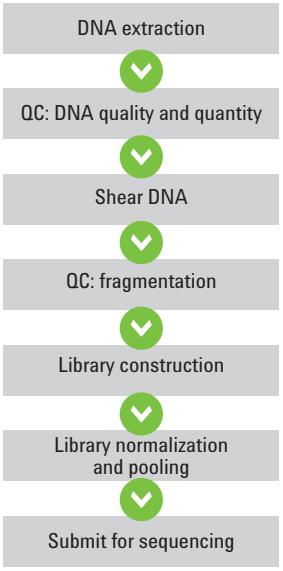


Figure 1. 100K Pathogen Genome Project sample preparation workflow for multiplexed, short-read Illumina sequencing.

To validate the automated method, DNA for NGS library construction was extracted from selected bacterial isolates with different GC content. These microbes are listed in Table 1. Organisms were lysed using the KAPA Express Extract Kit (KK7102) [8], after which the DNA was purified with a Qiagen QIAamp DNA Mini Kit (51306) using the manufacturer's instructions [9]. Before shearing, the extracted DNA was analyzed using an Agilent 2200 TapeStation system with the Genomic DNA ScreenTape assay for integrity of high molecular weight DNA [7,10,11]. Fragmented DNA was quantified with the method described by Jeannotte *et al*. [7] using an Agilent 2200 TapeStation system as the 96-well plate high-throughput workflow for quantitation and sizing of gDNA samples for library construction [9].

Table 1. Library Construction Metrics for Libraries Prepared from Different Bacteria

Bacterium	Gram reaction	Approximate genome size (MB)	GC content (%)	Average library size (bp)	Final library yield (ng)
Lactococcus	Positive	2	35	430	679
Listeria	Positive	2	38	307	949
Vibrio	Negative	5 (two chromosomes)	41	304	198
Escherichia	Negative	5	51	347	361
Salmonella	Negative	5	52	299	287

DNA was sheared in batches of 96 samples using microtubes with the Covaris E220 Focused Ultrasonicator [12]. The fragmented DNA size was determined with the Agilent 2100 Bioanalyzer system and High Sensitivity DNA Kit, to confirm a normal size distribution around a ~300 bp peak.

The input into library construction with the KAPA HTP Library Preparation Kit (KK8234, Figure 2) was normalized to 1–5 µg for all samples [13]. The standard KAPA protocol with dual-SPRI size selection after adapter-ligation was used to achieve libraries with a fragment distribution in the range of 250–450 bp. Library amplification was done for eight cycles using the KAPA HiFi HotStart ReadyMix, followed by a final 1X SPRI bead cleanup step. The size distributions of amplified libraries were confirmed to be in the range of 200–500 bp, using the Agilent 2100 Bioanalyzer system with the High Sensitivity DNA Kit [14,15]. Libraries were quantified with the qPCR-based KAPA Library Quantification Kit (KK4824) prior to normalization and pooling for sequencing [16] at BGI@UCD (Sacramento, CA) on the Illumina HiSeq 2000.

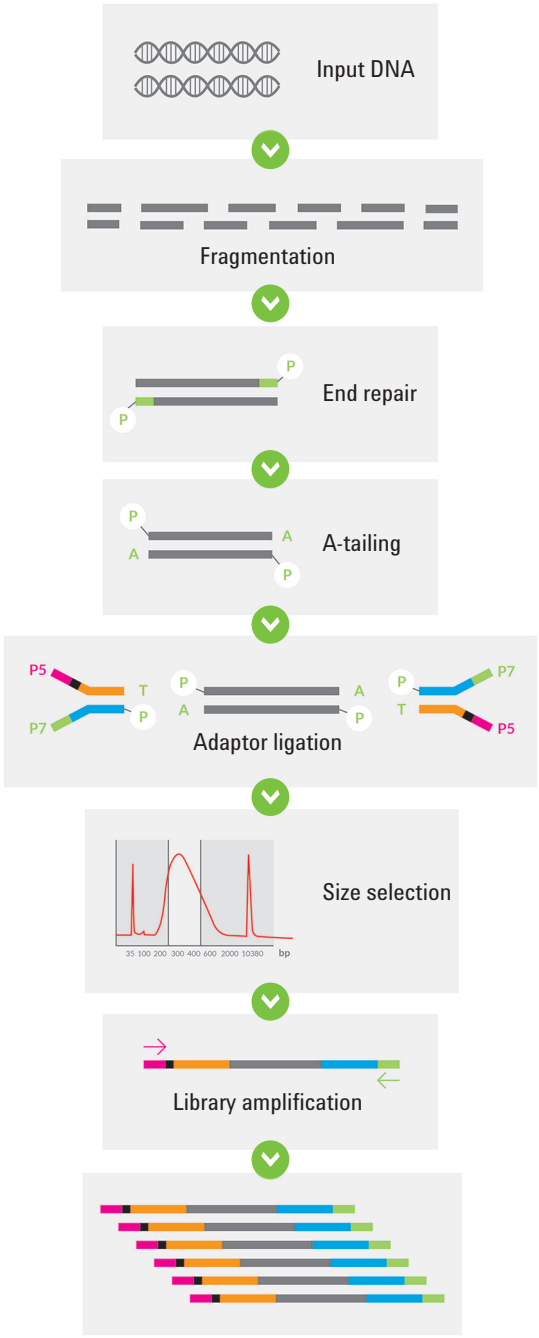


Figure 2. Detailed KAPA HTP Library Preparation protocol. The input into library construction is fragmented DNA or cDNA. Each enzymatic reaction is followed by a SPRI-bead cleanup (1.7X after End Repair; 1.8X after A-Tailing, and two consecutive 1X cleanups after adapter ligation). The "with-bead" protocol uses a single aliquot of SPRI beads for all cleanups prior to library amplification, and significantly reduces the loss of library fragments associated with the physical transfer of material between enzymatic reactions. This results in higher yields of adapter-ligated libraries, and reduces the number of amplification cycles required to generate sufficient material for Library QC and sequencing.

The entire library construction workflow, from fragmentation to the post-amplification cleanup, can be completed in as little as 9 hours. The validated, automated KAPA HTP Library Preparation method for the NGS Workstation (p/n G5522A; Figure 3) consists of three protocols, each with a graphical user interface or Form (Figure 4A–4C). Three variations of the first protocol (end repair to post-ligation cleanup) are available to provide the end user with the option of preparing libraries without size selection, employing a SPRI-bead based size selection, or performing off-deck size selection. The second protocol is used for library amplification setup, and the third for the post-amplification cleanup. This modular design provides the end-user with the flexibility to complete library construction over two days, or implement physical separation of pre- and post-PCR procedures. The VWorks Forms are supplemented by a Microsoft Excel workbook (Figure 4D) to facilitate reagent preparation and instrument setup, and can be used to keep a complete record of each experiment.

Consumables used on the NGS Workstation were selected for optimal performance. Labware movements and pipetting parameters for the KAPA HTP Library Preparation protocols were carefully optimized to minimize reagent and consumable waste and ensure equal or better library yields and quality compared to manual preparation methods.

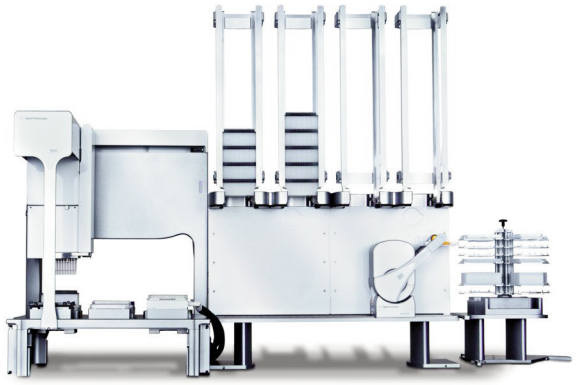
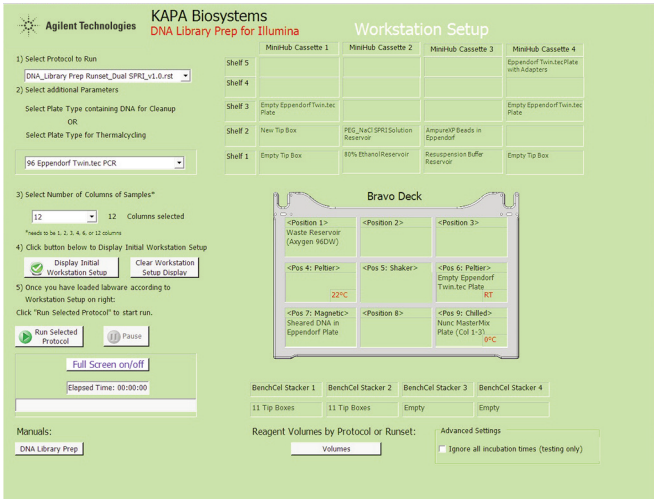
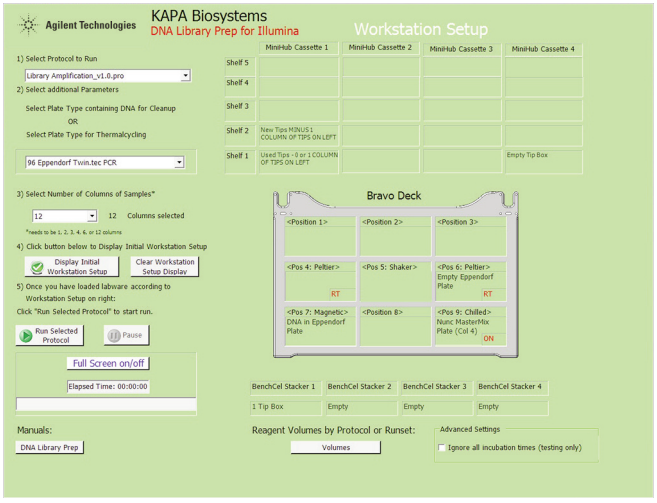


Figure 3. The Agilent NGS Workstation. This instrument comprises the highly reliable Agilent Bravo Automated Liquid Handling Platform, Agilent Bravo accessories for heating, cooling, shaking, and magnetic bead manipulations. User intervention in multistep protocols is minimized through the use of the BenchCel 4R Microplate Handler and Labware MiniHub for labware storage and movement.

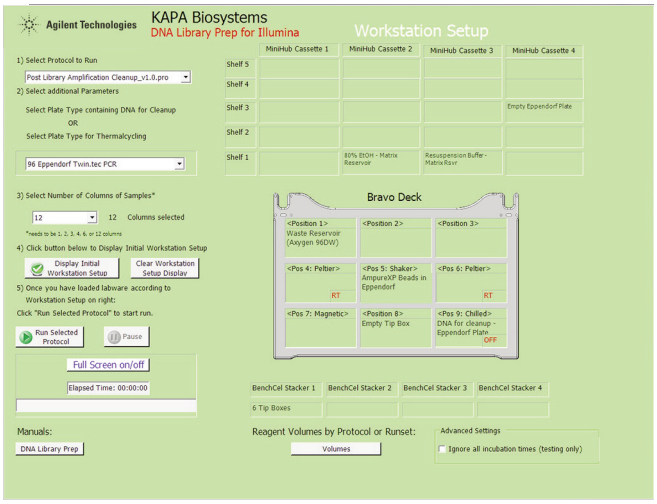
A



B



C



D

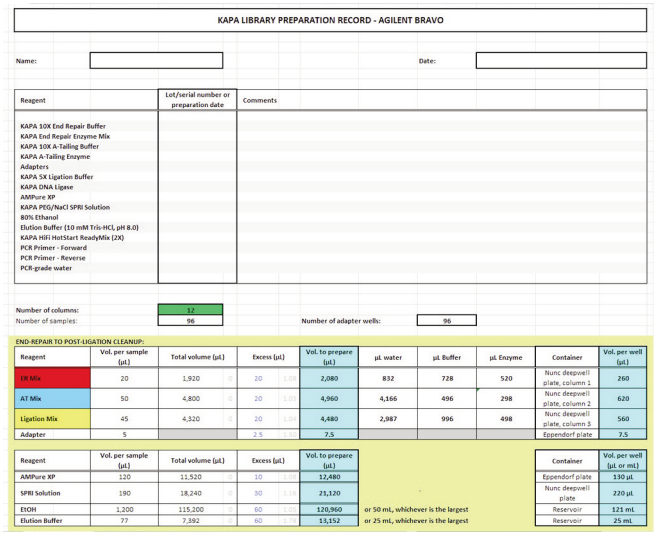


Figure 4. VWorks Forms and a Microsoft Excel workbook for the KAPA HTP Library Preparation method. A) End Repair to Post-ligation Cleanup Form, B) Library Amplification Setup Form, C) Post-amplification Cleanup Form, D) Microsoft Excel workbook. Each run is configured by selecting the number of plate sample columns (1, 2, 3, 4, 6, or 12 columns, and 8, 16, 24, 32, 48, or 96 samples), after which, the interface displays the deck setup and workstation setup for the protocol. The Microsoft Excel workbook is designed to guide reagent preparation and instrument setup, and can be used to create a record of each experiment. The VWorks software has many additional, convenient features, including various ways of tracking the progress during a run, and intuitive manual control of over instrument components to facilitate error recovery.

Results and Discussion

Prior to shearing, the extracted DNA was analyzed using the Agilent 2200 TapeStation system with the Genomic DNA ScreenTape assay (Figure 5). A typical electropherogram for fragmented bacterial DNA used for library construction is given in Figure 6. Electropherograms and virtual gel images for representative libraries prepared from *Salmonella, Escherichia, Vibrio, Listeria* and *Lactococcus* isolates using the KAPA HTP Library Preparation method on an Agilent NGS Workstation are given in Figure 7. Interestingly, libraries generated from different *lactococcal* isolates displayed different apparent library fragment sizes, but all of these libraries produced adequate sequence results. The higher molecular weight peak in the electropherograms for *Listeria* libraries are typical of over-amplification (primer depletion during library amplification). Since the average yield of adapter-ligated library was higher for *Listeria* than for the other bacteria, the number of amplification cycles could have been reduced. The enteric pathogens also produced similar sized libraries that produced excellent sequence results.

Library construction metrics (average library size and final library yields) for libraries prepared from different bacteria are summarized in Table 1.

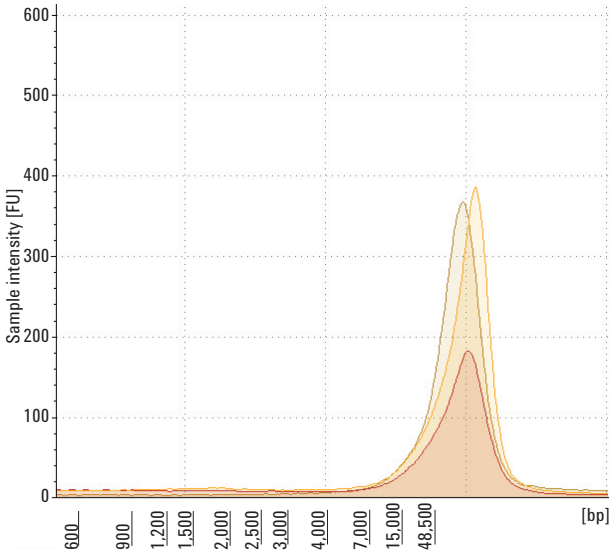


Figure 5. Typical electropherogram generated with the Agilent 2200 TapeStation system and Genomic DNA ScreenTape assay to assess integrity of high molecular weight genomic DNA used for library construction. Each line represents an individual Listeria isolate.

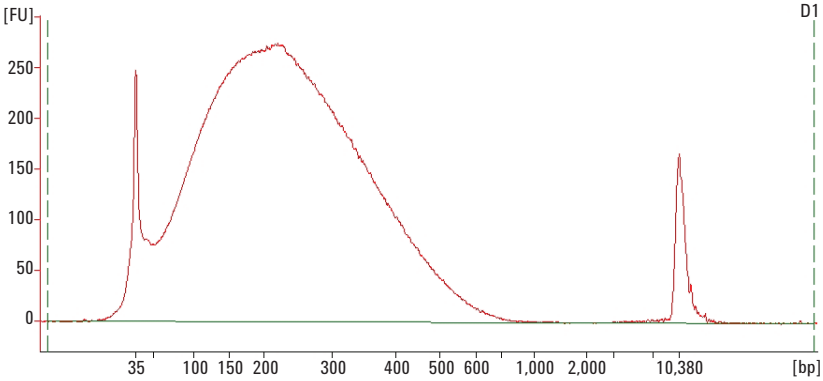


Figure 6. A typical electropherogram (generated on the Agilent 2100 Bioanalyzer system and High Sensitivity DNA Kit) of Covaris-sheared DNA used for library construction. Fragmentation parameters were selected to produce a fragment size distribution with a peak in the range of ~300 bp, which is ideal for 2 x 100 bp paired-end sequencing on the Illumina HiSeq 2000.

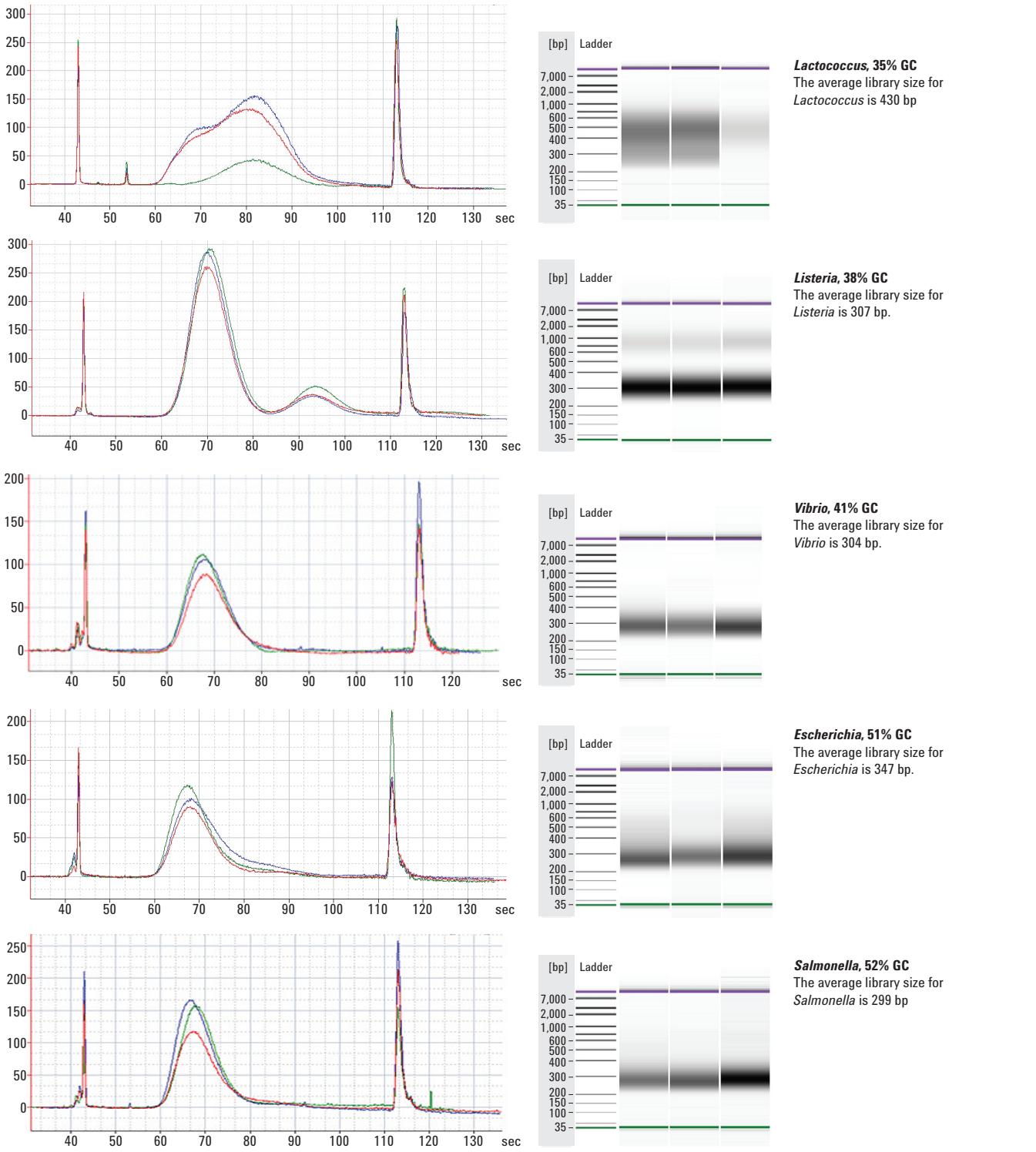


Figure 7. Representative electropherograms and virtual gel images (generated on the Agilent 2100 Bioanalyzer system with the High Sensitivity DNA Kit) of bacterial libraries prepared for whole genome sequencing with the KAPA HTP Library Preparation Kit on the Agilent NGS Workstation. The average library size for each genus was as indicated. Peaks at 35 and 10381 bp are internal standards used for alignment and quantitation determination with the Agilent 2100 Bioanalyzer system.

Conclusion

The KAPA HTP Library Preparation Kit and Agilent NGS Workstation provide a robust, high-throughput automation solution for the construction of high-quality libraries for microbial whole genome sequencing on the Illumina platform. This combined solution produced a workflow that is suitable for the construction of high-quality libraries for microbial whole genome sequencing in a production scale sequencing project like the 100K Pathogen Genome Project.

Acknowledgements

We gratefully acknowledge the technical assistance provided by Winnie Ng, Poyin Chen, Narine Arabyan, Soraya Foutouhi, Elizabeth Marta, Kerry Le, Sum Leung, Lucy Cai, Alan Truong, Christina Kong, Vivian Lee, Alvin Leonardo, Alex Weimer, Giovanni Tenorio, Patrick Ancheta, Azarene Foutouhi, and Kendra Liu from the laboratory of Dr. Bart Weimer at University of California, Davis.

References

1. S. Fisher, *et al*., *Genome Biology* **12**, R1 (2011).
2. M.A. Quail, *et al*., *Nature Methods* **9**, 10–11 (2012).
3. S.O. Oyola, *et al*., *BMC Genomics* **13**, 1 (2012).
4. M.A. Quail, *et al*., *BMC Genomics* **13**: 341 (2012).
5. M.G. Ross, *et al*., *Genome Biology* **14**: R51 (2013).
6. N. Kong, *et al*. Agilent Technologies Application Note (5991-3722EN): Production and Analysis of High Molecular Weight Genomic DNA for NGS Pipelines Using Agilent DNA Extraction Kit (p/n 200600).
7. R. Jeannotte, *et al*. Agilent Technologies Application Note (5991-4003EN): High-Throughput Analysis of Foodborne Bacterial Genomic DNA Using Agilent 2200 TapeStation and Genomic DNA ScreenTape System.
8. KAPA Express Extract Kits: http://www.kapabiosystems.com/products/name/kapa-express-extract-kits
9. Qiagen QIAamp DNA Mini Kit: http://www.qiagen.com/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/qiaamp-dna-mini-kit
10. Agilent 2200 TapeStation User Manual, Agilent Technologies (p/n G2964-90001).
11. Agilent Genomic DNA ScreenTape System Quick Guide, Agilent Technologies (p/n G2964-90040 rev.B).
12. Covaris E220 Focused-ultrasonicators: http://covaris-inc.com/products/afa-ultrasonication/e-series/
13. KAPA High Throughput Library Preparation Kit with SPRI solution and Standard PCR Library Amplification/Illumina series (96 libraries): http://www.kapabiosystems.com/products/name/kapa-library-preparation-kits
14. Agilent 2100 Bioanalyzer User Manual, Agilent Technologies (p/n G2946-90003).
15. Agilent High Sensitivity DNA Kit Guide (Rev. B), Agilent Technologies (p/n G2938-90321).
16. KAPA Library Quantification Kit - Illumina/Universal: http://www.kapabiosystems.com/products/name/kapa-library-quant-kits

For More Information

These data represent typical results. For more information on our products and services, visit our Web site at www.agilent.com/chem.

www.agilent.com/chem

Printed in the USA
May 19, 2014
5991-4296EN

Agilent Technologies