Quality Control of High-Throughput Library Construction Pipeline for KAPA HTP Library Using an Agilent 2200 TapeStation

Application Note

Authors

Nguyet Kong, Whitney Ng,
Azarene Foutouhi, B. Carol Huang, and
Bart C. Weimer
100K Pathogen Genome Project
Population Health and Reproduction,
School of Veterinary Medicine
University of California-Davis, Davis,
CA, USA

Lenore Kelly
Agilent Technologies, Inc.
Santa Clara, CA, USA

Abstract

Next Generation Sequencing requires the input of high molecular weight genomic DNA to construct quality libraries for whole genome bacterial sequencing. Large scale sequencing projects, such as the 100K Pathogen Genome Project, require methods to rapidly assess the quantity and quality of the input DNA using high-throughput methods that are fast and cost effective. In this study, the Agilent 2200 TapeStation and Agilent 2100 Bioanalyzer Systems were used to assess a few critical quality control steps for library construction. With minimal manual intervention, the Agilent 2200 TapeStation System determined the quality of genomic DNA, fragmented DNA, and final libraries constructed from multiple types of foodborne pathogens. The Agilent 2200 TapeStation System provided a single platform that effectively evaluated the necessary quality control steps, which provided a distinct advantage to decrease the time needed for library construction and a common instrument methodology for quality control.

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, resulting in the production of thousands of genomes. This contribution is especially important to food security because it represents a consortium of government, academic, and industrial partners in a global effort to make these sequences public. The 100K Pathogen Genome Project (http://100kgenome.vetmed.ucdavis.edu/) has committed to sequencing 100,000 bacterial zoonotic and foodborne pathogens from around the globe. As a large scale next generation sequencing project, it requires high-throughput procedures from DNA extraction to library construction and sequencing. This capacity demands an automated library construction workflow to enable the reliable and robust sequencing pipeline, which in turn creates a need for robust and high-throughput quality control steps to be established.

NGS library construction comprises repetitive processes, making it adaptable to automation. Automation offers many advantages, including sample throughput, reduced hands-on time, greater reproducibility, and improved process control [1]. However, the development and validation of automated protocols is not trivial. In the preparation for sequencing such a large number of microbial genomes, DNA quantification (and size qualification) is extremely important in many steps of the pipeline, and is a challenge with current automated instruments (Figure 1). The additional quality checks needed during library construction creates a need for a streamlined process to rapidly assess genomic DNA (gDNA) as well as fragmented DNA. A method to check the quality of gDNA extracts, DNA shearing protocols, and final library constructions directly in a 96-well plate format would significantly increase throughput.

Genomic DNA extracts are often analyzed on agarose gels, but this approach is not suitable for a high-throughput workflow. Size estimation against a ladder coupled with densitometry to determine concentration offers low resolution and cannot be automated. Currently, the Agilent 2100 Bioanalyzer System can size fragments up to 12,000 base pairs (bp), which is much smaller than gDNA. Consequently, the 2100 Bioanalyzer System is not applicable to DNA extraction QC steps. The 2100 Bioanalyzer System can run 12 samples at a time, which are hand-pipetted and cannot be loaded onto the instrument automatically from a 96-well plate. Therefore, evaluation of gDNA quality (that is, lack of degradation) is crucial because the next step in library preparation for automated sequencing is DNA shearing, which depends on consistent high molecular weight gDNA [2].

The gDNA is sheared to an optimal size range depending on the sequencing technology. Sheared gDNA was evaluated to confirm a normal size distribution centered around ~300 bp before the samples were placed on the Agilent NGS Workstation for library construction [3].

Agilent Technologies recently introduced the Agilent 2200 TapeStation System for DNA quantification and sizing. The 96-well plate sample format offers multiplexing, which leads to faster analysis times, and offers streamlined workflows for microbial library preparation for a project of this magnitude. DNA analyses on the 2200 TapeStation system are performed using the ready-to-use Genomic DNA ScreenTape and Genomic DNA Reagents [4,5]. The 2200 TapeStation instrument automatically loads the prepared samples from the 96-well plate onto the Genomic DNA ScreenTape. Electrophoresis and imaging of an electropherogram or a gel image are all automated within the instrument. This application note presents a comparative study between the quantification of microbial gDNA, sheared DNA, and the final library obtained with the 2200 TapeStation System using Genomic DNA ScreenTape and High Sensitivity D1000 ScreenTape System to Agilent 2100 Bioanalyzer System, the accepted standard for DNA/RNA fragment QC assessment [5,6,7,8].

Materials and Methods

The 100K Pathogen Genome Project sample preparation workflow begins with high molecular weight gDNA followed by production of sheared DNA for library construction. Isolates were chosen to validate the entire library construction pipeline quality control using 2200 TapeStation System from specific bacterial isolates across a range of different GC content and genomes sizes (Table 1). DNA was extracted and followed by a cleanup with Qiagen QIAamp DNA Mini Kit (51306) using the manufacturer's instructions [9] to produce gDNA for all isolates tested.

The extracted gDNA was analyzed using the 2200 TapeStation for high molecular weight gDNA before shearing and library construction (Table 1) [5,6]. gDNA was sheared in batches of 96 samples using microtubes with the Covaris E220 Focused Ultrasonicator [10,11]. The sheared DNA size range was determined with the High Sensitivity DNA Kit on the Agilent 2100 Bioanalyzer system and 2200 TapeStation System with the High Sensitivity D1000 ScreenTape assay to measure the fragment distribution with desired sizes between ~200 to 300-bp (Table 1) [5,6,7,4]. Fragmented DNA was quantified using the method detailed by Jeannotte *et al*. (5991-4003EN) that describes use of the Agilent 2200 TapeStation system [12].

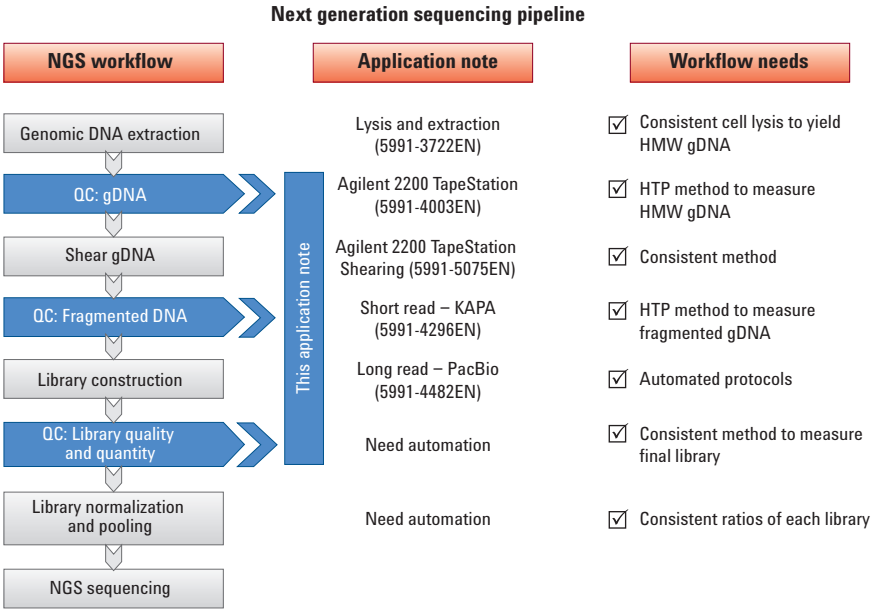


Figure 1. 100K Pathogen Genome Project sample preparation workflow for multiplexed, short read Illumina sequencing using the Agilent 2200 TapeStation System and Agilent 2100 Bioanalyzer for quality control.

Table 1. Comparison of library construction metrics for libraries prepared from different bacteria using the Agilent 2100 Bioanalyzer and the Agilent 2200 TapeStation Systems

				Average Shearing Size QC (bp)			Average Library Size QC (bp)		
Bacterium	**Gram reaction**	**Approximate genome size (Mb)**	**GC content (%)**	**Agilent 2100 Bioanalyzer**	**Agilent 2200 TapeStation**	**Difference (bp)**	**Agilent 2100 Bioanalyzer**	**Agilent 2200 TapeStation**	**Difference (bp)**
Campylobacter	Negative	1.7	30	205	248	43	304	310	6
Staphylococcus	Positive	2.8	32	180	205	25	303	302	1
Lactococcus	Positive	2	35	179	235	56	290	301	11
Listeria	Positive	2	38	204	234	30	300	289	11
Vibrio	Negative	5	43	208	205	3	301	287	14
Escherichia	Negative	5	51	260	255	5	320	312	8
Salmonella	Negative	5	52	263	270	7	279	307	28
Pseudomonas	Negative	5.2	66	250	250	0	294	295	1
Micrococcus	Positive	2.5	70	245	240	5	310	320	10

The sheared DNA input used for library construction (KAPA HTP Library Preparation Kit KK8234; Kapa Biosystems, Boston, MA) was normalized to 1–5 µg for all samples (Figure 2) [13]. The standard KAPA protocol with dual-SPRI size selection after adapter-ligation was used to construct libraries that resulted in fragment distribution of 250–499 bp. After library amplification, final libraries were confirmed with the High Sensitivity DNA Kit on the 2100 Bioanalyzer System and 2200 TapeStation System with DNA High Sensitivity Kit to be between 200–500 bp. Libraries were quantified with the qPCR-based KAPA Library Quantification Kit (KK4824) prior to normalization and pooling before sequencing [14] at BGI@UCD Davis (Sacramento, CA) using the Illumina HiSeq 2000 (San Diego, CA).

Results and Discussion

Prior to shearing, the extracted DNA was analyzed using the 2200 TapeStation System with Genomic DNA Kit to determine the high molecular weight gDNA that was subsequently used for library construction. The 2100 Bioanalyzer System can size fragments up to 12,000 base pairs (bp), which is much smaller than gDNA. The sheared gDNA sizes were within the expected size range using each technology. Differences between the 2100 Bioanalyzer System and the 2200 TapeStation System observed for Gram Positive organisms were < 56 bp, while the differences for Gram Negative microbes were < 43 bp. These sizes were within the performance specifications for the instruments, and were considered within acceptable ranges.

The library QC metrics for libraries prepared from these samples were within acceptable ranges with differences between the final library size < 20 bp. The final libraries showed no bias from gram reaction and GC content (%) ranging from 30 to 70 based on insert size, and passed the established QC limits for sequencing. The 2200 TapeStation System was considered equivalent to the 2100 Bioanalyzer System for determining the final library size.

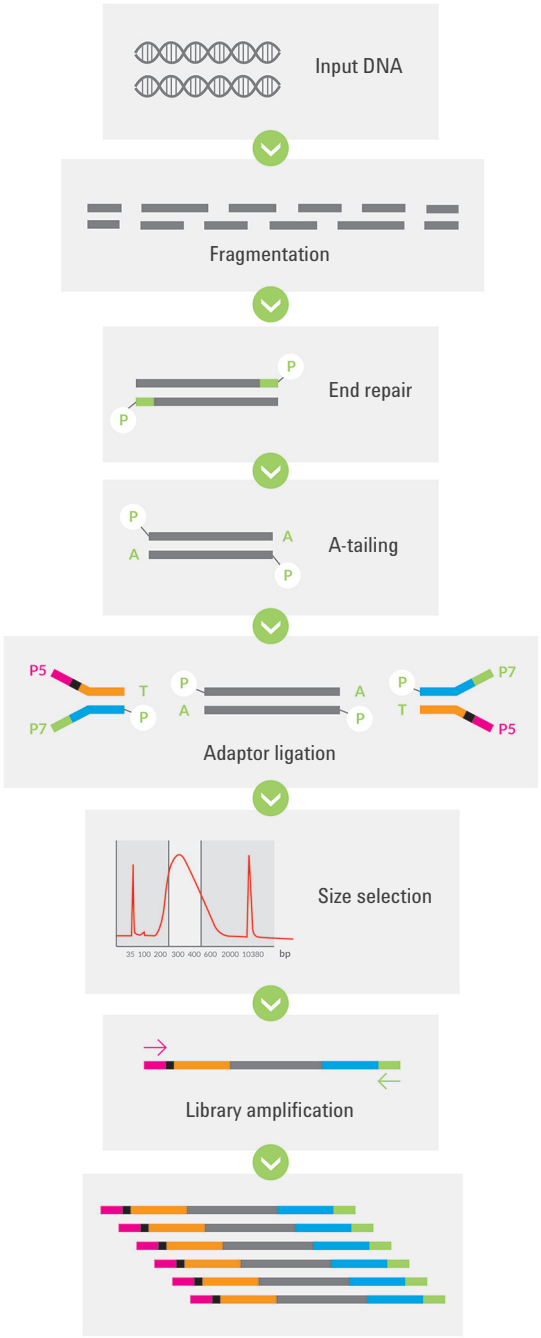


Figure 2. Detailed KAPA HTP Preparation Protocol. The input into library construction is fragmented DNA or cDNA. Each enzymatic reaction is followed by a SPRI-bead cleanup (1.7X after End Repair; 1.8X after A-Tailing, and two consecutive 1X cleanups after adapter ligation). The "with-bead" protocol uses 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.

Electropherograms for three different gDNA, sheared DNA, and final libraries were represented using the 2200 TapeStation or 2100 Bioanalyzer Systems (Figures 3A, 3B, and 3C). Each of the isolates had similar library sizes that subsequently produced adequate sequence results. The final yield suggested over-amplification using the 2100 Bioanalyzer System, but not with the 2200 TapeStation System. Thus, the enteric pathogens produced similar sized libraries that produced excellent sequence results.

Conclusion

Genomic DNA from nine bacteria with three different isolates was extracted and analyzed for quality and molecular weight using the Agilent 2200 TapeStation System with Genomic DNA ScreenTape assay, which determined the gDNA was high molecular weight and not degraded. Using the same instrumentation, the fragmented DNA and final libraries were determined. In both QC steps, the two instruments produced equivalent results.

In the future, the Agilent NGS Workstation can be used to prepare samples to complete the quality control for the library preparation workflow. Thus, the Agilent 2200 TapeStation System met the needs of the high-throughput workloads because 96 samples in a well plate at a time can be measured. Using a single platform, we can evaluate gDNA, fragmented DNA, and final libraries, streamlining the process and the workflow for high volume DNA library preparation.

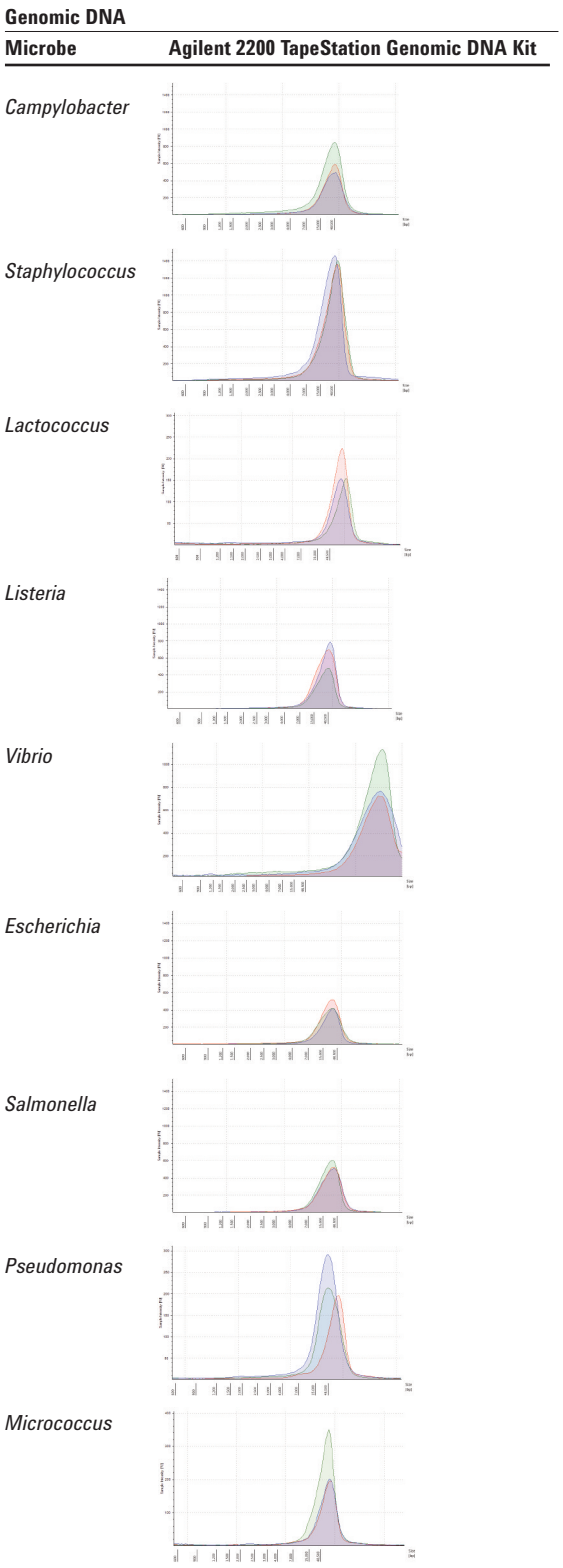


Figure 3A. The electropherograms generated with the Agilent 2200 TapeStation Genomic DNA Kit to determine high molecular weight genomic DNA used for library construction. Each trace represents an individual isolate.

Fragmented DNA

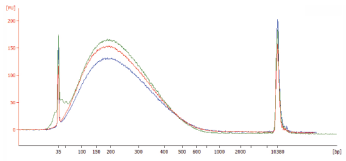
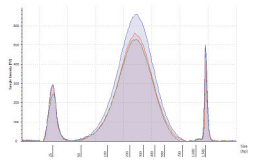
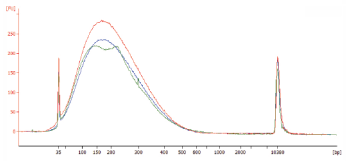
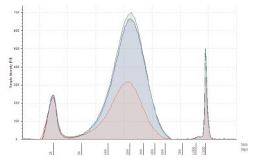
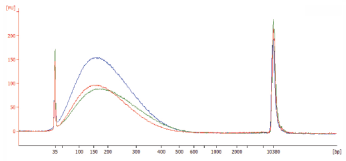
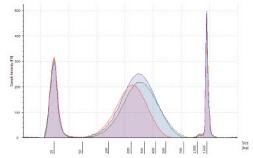
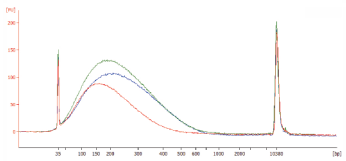
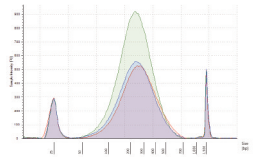
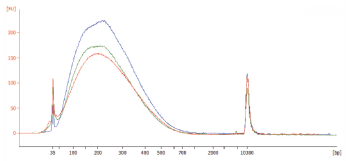
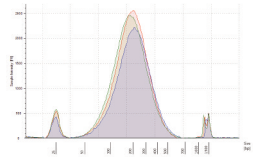
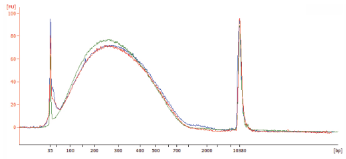
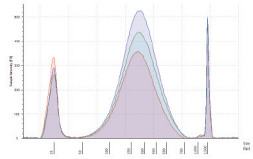
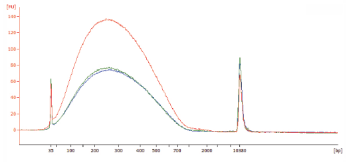
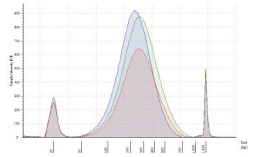
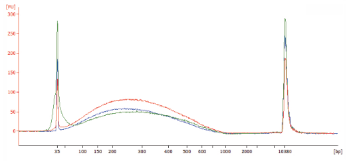
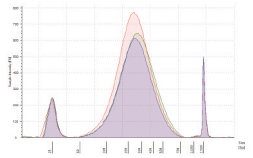
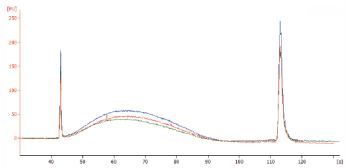
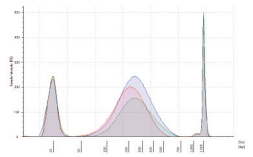
Microbe	Agilent 2100 Bioanalyzer System	Agilent 2200 TapeStation System
Campylobacter		
Staphylococcus		
Lactococcus		
Listeria		
Vibrio		
Escherichia		
Salmonella		
Pseudomonas		
Micrococcus		

Figure 3B. Electropherograms generated from Covaris sheared DNA used for library construction (on the Agilent 2100 Bioanalyzer system with the High Sensitivity DNA Kit and the Agilent 2200 Tapestation System with the High Sensitivity D1000 Kit). The far right and left peaks are internal standards, the broad center peak represents the sheared DNA. 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.

Final library

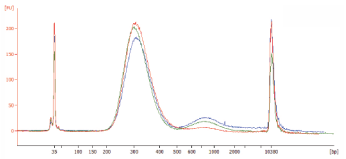
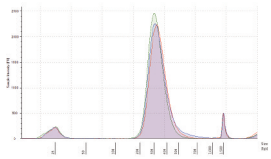
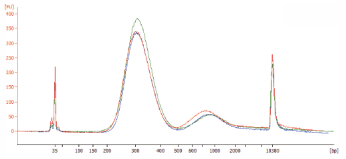
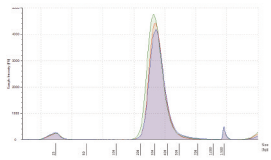
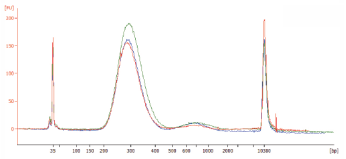
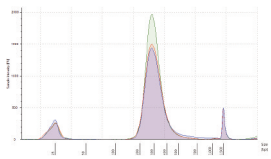
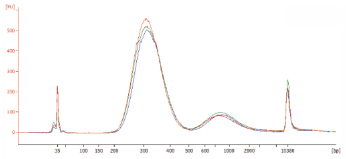
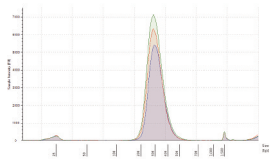
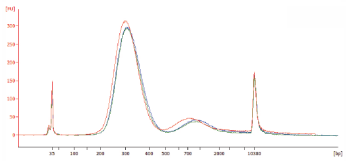
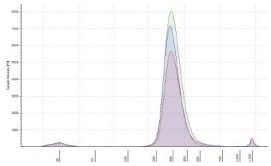
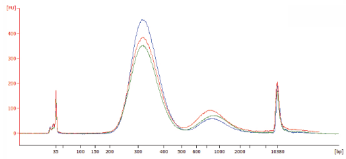
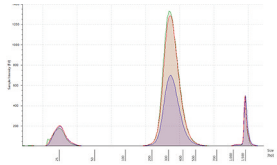
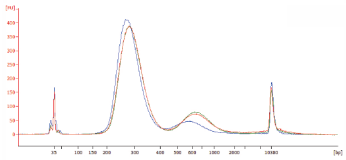
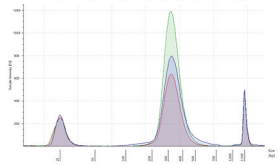
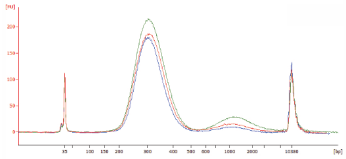
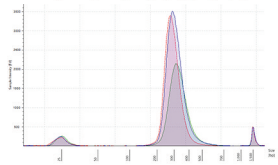
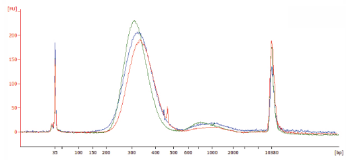
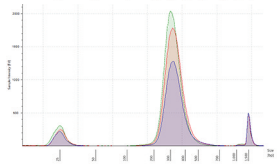
Microbe	Agilent 2100 Bioanalyzer System	Agilent 2200 TapeStation System
Campylobacter		
Staphylococcus		
Lactococcus		
Listeria		
Vibrio		
Escherichia		
Salmonella		
Pseudomonas		
Micrococcus		

Figure 3C. Electropherograms generated on the Agilent 2100 Bioanalyzer system with the High Sensitivity DNA Kit and the Agilent 2200 Tapestation System with the High Sensitivity D1000 Kit. The bacterial libraries were prepared for whole genome sequencing with the KAPA HTP Library Preparation Kit on the Agilent NGS Workstation. The electropherogram average library size for each genus is indicated. The far right and left peaks are internal standards, the broad center peak represents the sheared DNA in the Bioanalyzer and TapeStation kit.

Acknowledgements

We gratefully acknowledge the technical assistance provided by Kao Thao, Poyin Chen, Narine Arabyan, Soraya Foutouhi, Allison Weis, Louis Sorieul, Alvin Leonardo, Lucy Cai, Alan Truong, Patrick Ancheta, Christina Kong, Vivian Lee, Regina Agulto, Kendra Liu, Derek Yip, San Mak, Preston Leung, Jordan Leung, and Madison Snider from the laboratory of Dr. Bart Weimer at University of California, Davis.

References

1. M. A. Quail, *et al*. "A large genome center's improvements to the Illumina sequencing system" *Nature Methods* **5**:1005-1010, 2008.
2. S. Wilkening, *et al*. "Genotyping 1000 yeast strains by next-generation sequencing" *BMC Genomics* **14**:90, 2013.
3. N. Kong, *et al*. "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" *Agilent Technologies*, publication number 5991-4296EN, **2014**.
4. "Agilent Genomic DNA ScreenTape System Quick Guide" *Agilent Technologies*, publication number G2964-90040 rev.B, **2013**.
5. "Agilent 2200 TapeStation User Manual" *Agilent Technologies*, publication number G2964-90002 Rev. B, **2013**.
6. "Agilent High Sensitivity D1000 ScreenTape System Quick Guide" *Agilent Technologies*, publication number G2964-90032 Rev B, **2013**.
7. "Agilent High Sensitivity DNA Kit Guide" *Agilent Technologies*, publication number G2938-90321 Rev. B, **2013**.
8. "Agilent 2100 Bioanalyzer User Manual" *Agilent Technologies*, publication number G2946-90003.
9. Qiagen QIAamp DNA Mini Kit: http://www.qiagen.com/us/search.aspx?q=qiaamp%20dna%20mini%20kit#&&p=1
10. DNA Shearing with microTUBEs, Covaris (p/n 400056 Rev G), 2011: http://covarisinc.com/products/afa-tubes-and-vials/microtube/
11. R. Jeannotte, *et al*. "Optimization of Covaris Settings for Shearing Bacterial Genomic DNA by Focused Ultrasonication and Analysis Using Agilent 2200 TapeStation" *Agilent Technologies*, publication number 5991-5075EN, **2014**.
12. R. Jeannotte, *et al*. "High-Throughput Analysis of Foodborne Bacterial Genomic DNA Using Agilent 2200 TapeStation and Genomic DNA ScreenTape System" *Agilent Technologies*, publication number 5991-4003EN, **2014**.
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. KAPA Library Quantification Kit - Illumina/Universal: http://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/library-quantification/

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.

See also:
www.agilent.com/genomics/tapestation
www.agilent.com/genomics/bioanalyzer

www.agilent.com/chem

Printed in the USA
September 18, 2014
5991-5141EN

Agilent Technologies