# Automation of PacBio SMRTbell 10 kb Template Preparation on an Agilent NGS Workstation

Application Note


## Authors

Nguyet Kong, Kao Thao, Whitney Ng, and Bart C. Weimer
100K Pathogen Genome Project
Department of Population Health and Reproduction
School of Veterinary Medicine
University of California, Davis
Davis, CA, USA

Kristi Spittle Kim, Jonas Korlach, and Luke Hickey
Pacific Biosciences
Menlo Park, CA 94025

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



## Abstract

Automation of the PacBio 10 kb Template Preparation using the SMRTbell Template Prep kit for the sequencing of bacterial genomes on the Agilent Automated Liquid Handling Platform is described. Four bacterial genomes of differing GC content are used to prepare libraries without bias reading for next generation sequencing. The gDNA quality assessment was done using spectrophotometric and automated electrophoresis for high molecular weight DNA with an Agilent 2200 TapeStation system with the Agilent Genomic DNA ScreenTape assay. The sheared DNA was measured with the Agilent Bioanalyzer system and the DNA 12000 Kit.


Agilent Technologies

## Introduction

Genome sequencing of microbial pathogens continue to reveal impactful insights into the complexities of infectious disease agents, public health, human disease, and clinical diagnostics. Recent advancements in DNA sequencing technology allow whole genome sequencing to be accomplished with more cost-effective and high-throughput methods at an unprecedented rate and sequence quality. In particular, the use of next generation sequencing (NGS) methods has increased productivity and revealed new information about the structure and function of bacterial genomes, leading to increasingly large data sets for analysis. Largely, NGS methods relying on short reads to produce draft genomes continue to experience sequencing limitations due to GC content bias and repeat regions that provide sufficient information to allow closed genome assemblies. The PacBio RS II approach provides an alternative that quickly produces closed genomes by using a single molecule, real-time (SMRT) DNA sequencing, a technology that provides extra-long read lengths. This method reduces the number of contigs and increases accuracy of *de novo* assembly of bacterial whole genomes. The real-time technology is the only sequencing technology that allows determination of the genomic DNA (gDNA) sequence, plasmid DNA sequence, epigenetic modifications, and closed genome assembly simultaneously. Due to these advantages, the 100K Pathogen Genome Project (http://100kgenome.vetmed.ucdavis.edu) selected SMRT technology to produce 1,000 closed genomes. Production of genomes on this scale required automation of the sequencing library construction.

This application note describes the automation of the PacBio 10 kb Template Preparation using the SMRTbell Template Prep Kit to provide rapid sequencing of bacterial genomes for the 100K Pathogen Genome project, and obtain additional genomic information linked to specific traits important to food-associated bacterial pathogens. This solution consists of the industry's most compact and versatile Agilent Liquid Handling Platform (Figure 1), predeveloped protocols, and solutions for labware management.

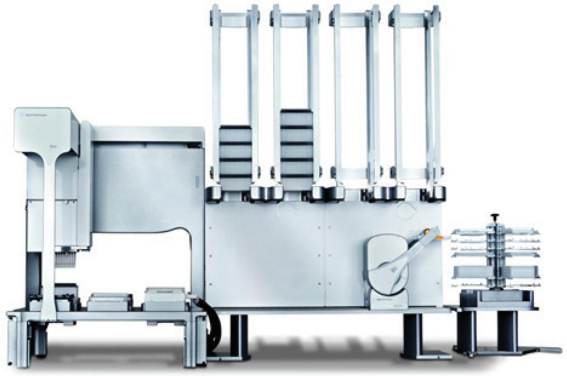

*Figure 1. The Agilent NGS Workstation. This workstation comprises the highly reliable Agilent Bravo Automated Liquid Handling Platform, Agilent Bravo accessories for heating, cooling, shaking, and magnetic bead manipulations. User intervention in multistep protocols is minimized through the use of the BenchCel 4R Microplate Handler and Labware MiniHub for labware storage and movement.*

## Materials and Methods

The SMRTbell Template Preparation kit from Pacific Biosciences provides a simple, fast, and scalable solution for producing a variety of insert-sized libraries that can be used in a broad range of NGS applications. The input gDNA must be pure and of high molecular weight prior to fragmentation to achieve acceptable library size for optimal sequencing quality.

For this work, gDNA for PacBio SMRTbell 10 kb Library construction was extracted from selected bacterial isolates with increasing GC content (Table 1). DNA was obtained with the QIAamp DNA Mini Kit (51306) using manufacturer's instructions [1,2,3,4]. The gDNA quality assessment was done using spectrophotometric and automated electrophoresis for high molecular weight DNA Agilent 2200 TapeStation system with the Agilent Genomic DNA ScreenTape assay, following manufacturer instructions [5,6].

*Table 1. Bacteria Used in This Study*

| ID no. | Lab ID | Color | Bacterium | GC content (%) |
|---|---|---|---|---|
| 1 | BCW_5811 | | *Campylobacter jejuni* | 30 |
| 2 | BCW_3862 | | *Listeria monocytogenes* | 38 |
| 3 | BCW_5789 | | *Vibrio fluvialis* | 41 |
| 4 | BCW_4887 | | *Salmonella enterica serovar enteritidis* | 52 |

After evaluation for gDNA size and quantity, fragmentation was done using the Covaris g-TUBE device (520079) following manufacturer instructions [7]. A sample loss between 20–50% was expected after fragmentation. The fragmented gDNA was sized on an Agilent 2100 Bioanalyzer system with the Agilent DNA 12000 Kit (part number 5067-1508). The broad linear range of this separation gives both sizing and quantitation with minimal sample consumption [8,9]. The fragmented DNA was quantified using a Life Technologies Qubit 2.0 Fluorometer using a dsDNA HS assay, following manufacturer protocol [10]. The sheared gDNA input into library construction with the PacBio SMRTbell 10 kb Library Preparation Kit was normalized for all samples between 1–5 µg. If preparing libraries with DNA input amounts greater than 5 µg, all the reaction volumes can be scaled proportionally [11].

A streamlined workflow of the main steps to construct final DNA libraries for sequencing was used:

1. Fragment gDNA using a Covaris g-TUBE device.
2. Repair DNA damage and repair ends of fragmented DNA.
3. Purify the DNA.
4. Blunt-end ligate using blunt adapters.
5. Purify template for submission to a sequencer (Figure 2).

The workflow will take as little as 7 hours for post-shearing cleanup and to construct DNA libraries for 96 samples ready for submission using the NGS Workstation [12]. After library construction, the sequencing facility anneals sequencing primer and polymerase to the template before loading the library onto the PacBio RS II.

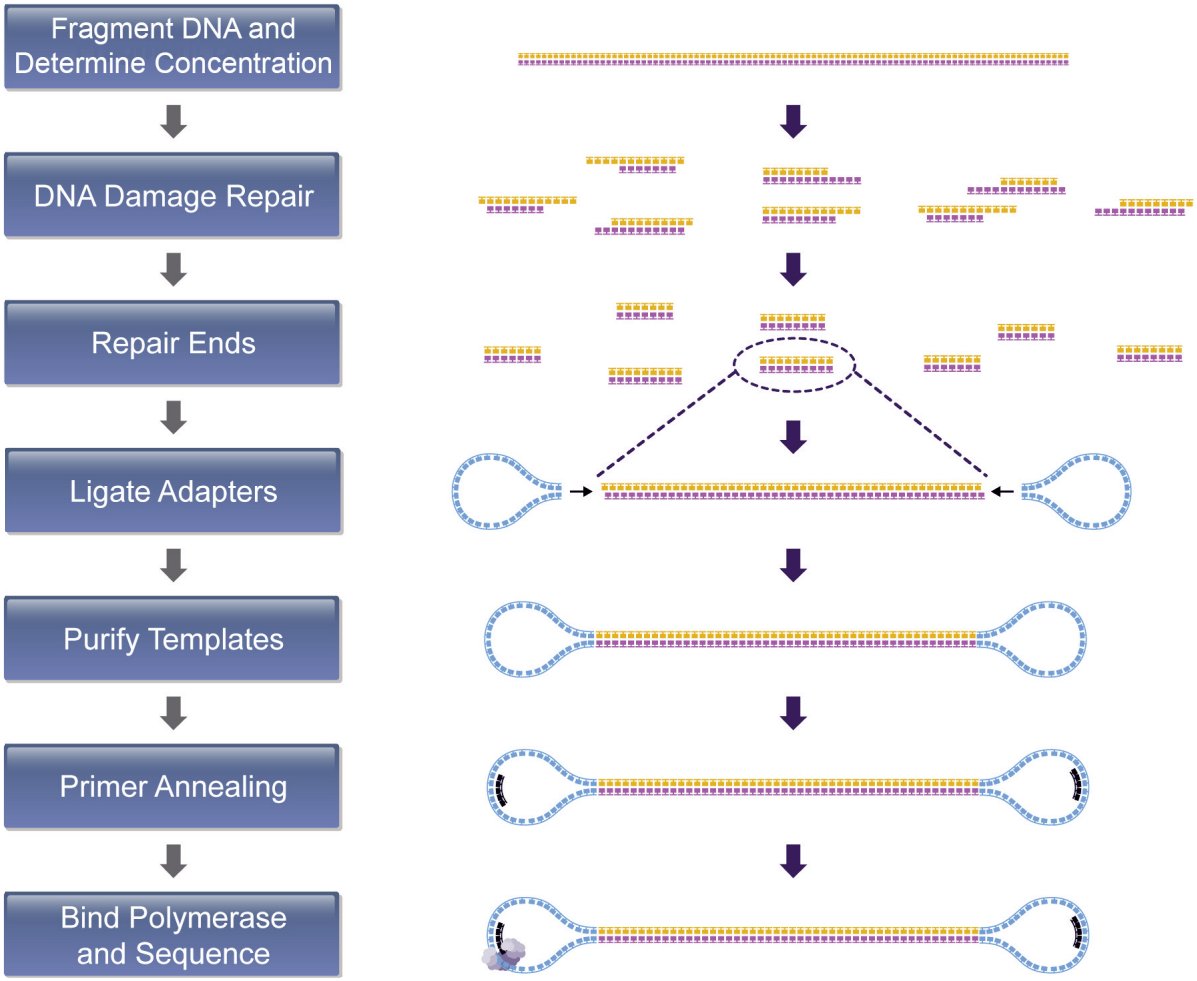


*Figure 2. SMRTbell Template Preparation Workflow for PacBio RS II system. This workflow is used to prepare libraries from fragmented and concentrated DNA using Covaris g-TUBE and concentrated using the AMPure magnetic beads before following PacBio SMRTbell 10 kb Library Preparation procedures. If preparing libraries with DNA input that are greater than 5 µg, scale the reaction volume proportionally.*

In cooperation with Agilent Technologies, the 100K Pathogen Genome Project and Pacific Biosciences automated and validated the PacBio SMRTbell 10kb Library Preparation process on the Agilent NGS Workstation (G5522A) [12]. Two protocols, each with an easy-to-use graphic user interface, were produced (Figure 3) to enable quick scale-up of the protocol. The VWorks Form is supplemented by an Excel Workbook, which facilitates reagent preparation and instruction setup, and can be used to keep a complete record of each experiment.

Each protocol has a unique deck layout displayed. The labware used in the protocol was selected for optimal performance to minimize dead volumes, conserve critical reagents, and reduce cost. The Post Shearing Cleanup Form and 10 kb Library Prep Runset Dual SPRI Form, have many convenient features, including various ways of tracking the progress during a run, and intuitive manual control over instrument components to facilitate error recovery. The PacBio Library Excel Workbook configures the number of plate sample columns (1, 2, 3, 4, 6, or 12, that is, 8, 16, 24, 32, 48, or 96 samples), after which the interface displays the deck setup and workstation setup for the protocol. The Excel workbook is designed to guide reagent preparation and instrument setup, and can be used to create a record of each experiment. Reagent volumes are calculated on the workbook, where lot numbers and notable information can be recorded.

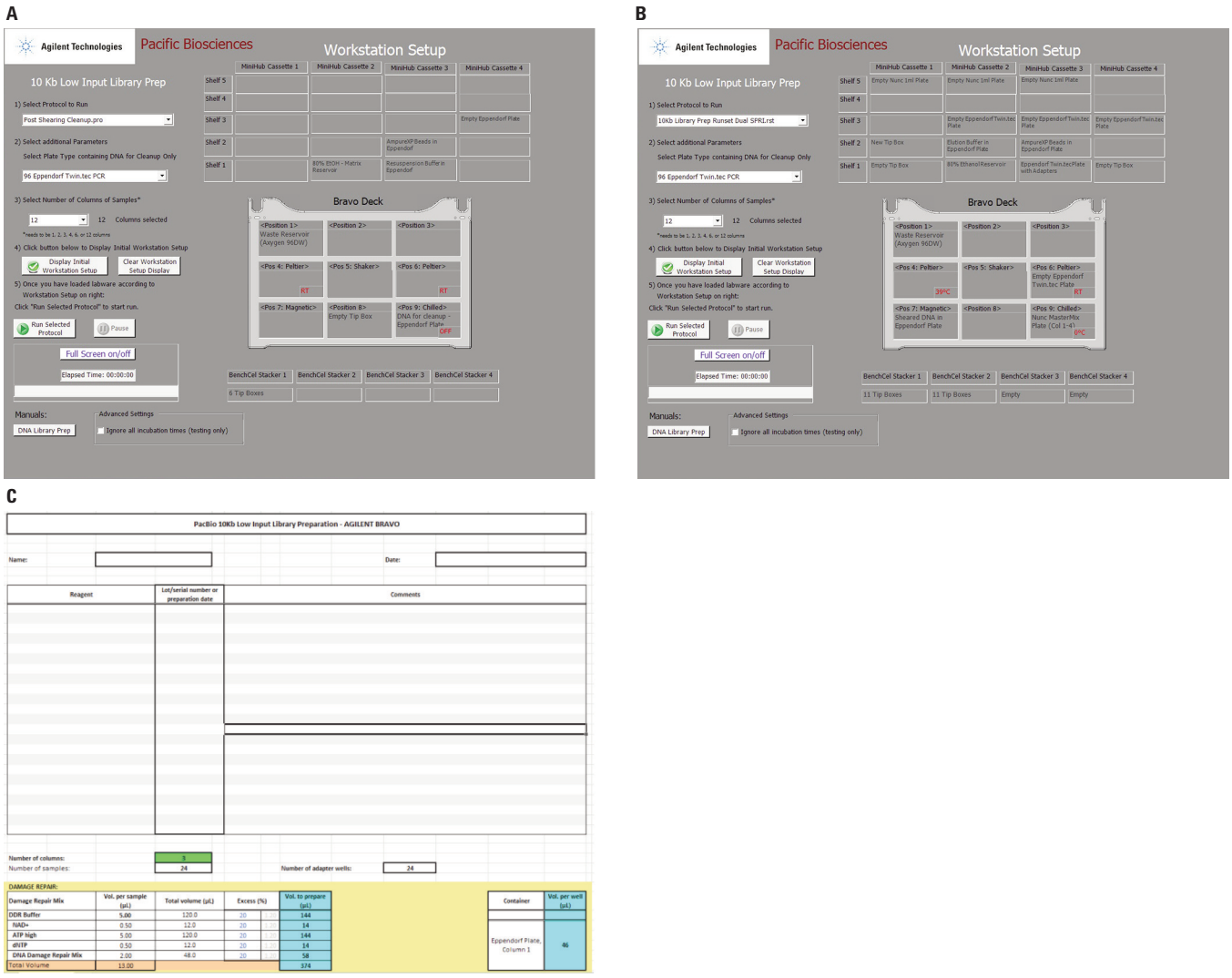


*Figure 3. VWorks protocol forms and Excel workbook for PacBio Library Preparation method provide an interactive, visual layout for the end user. A) Post Shearing Cleanup Form. B) 10 kb Library Prep Runset Dual SPRI Form. C) PacBio Library Excel Workbook.*

## Results and Discussion

High molecular weight gDNA used for library construction was determined using an Agilent 2200 TapeStation system (Figure 4). Overlayed electropherograms and a virtual gel image for the multiple sequencing libraries constructed using PacBio 10 kb Library Preparation on the Agilent NGS Workstation were used for QC. From 1 to 5 µg of sheared gDNA from four different bacterial isolates of varying GC content (Figure 5), were used to construct libraries and sequenced on the PacBio RS II sequencer. Figure 6 shows the electropherogram for the final DNA libraries for the organisms listed in Table 1. The distribution of reads from each sample is plotted to determine the depth of coverage across the genome.

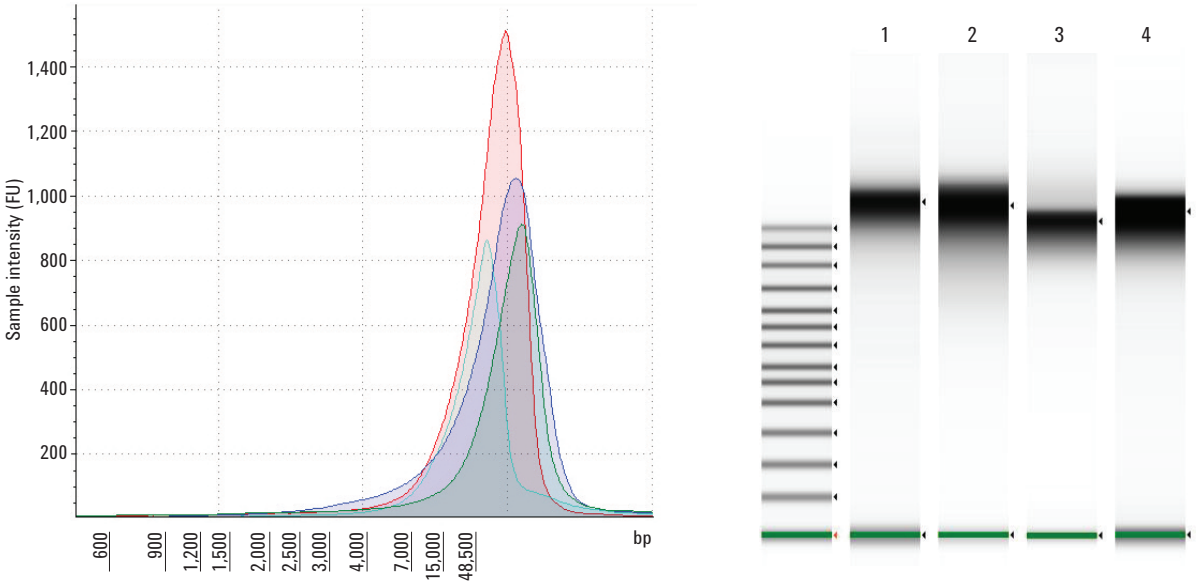


*Figure 4. Electropherogram and gel image of high molecular weight genomic DNA from an Agilent 2200 TapeStation system using the Genomic DNA ScreenTape Assay. The electropherogram has been edited so that the lower marker is to the left of the y-axis, to focus on gDNA sample.*

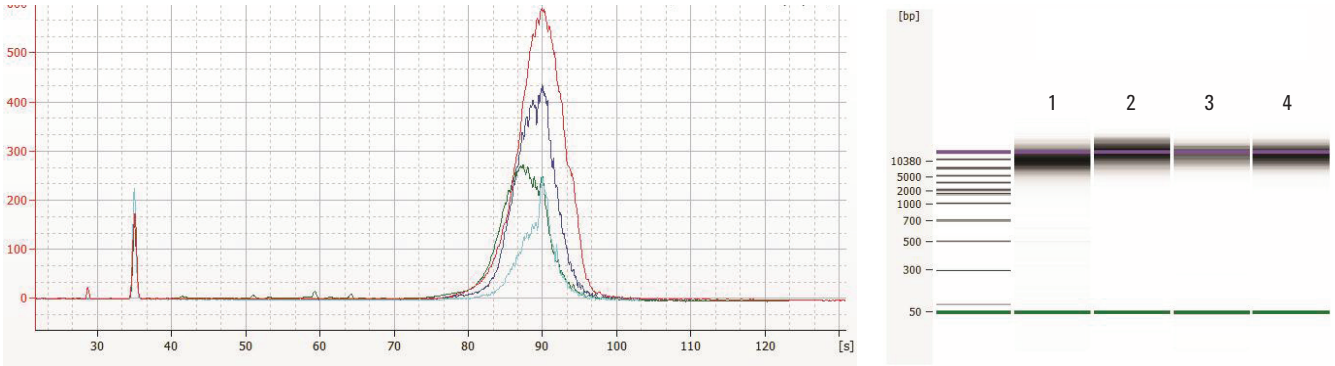


*Figure 5. Representative electropherograms and virtual gels are used for visual inspection (generated with the Agilent 2100 Bioanalyzer system with the DNA 12000 Kit) of sheared bacterial genomic DNA with average shearing size for Campylobacter (green, 10 kb), Listeria (blue, 13.5 kb), Vibrio (aqua, 11.6 kb), and Salmonella (red, 17 kb). Peaks near 35 are the lower marker internal standard for the DNA 12000 kit. A typical electropherogram using the Agilent Bioanalzyer 2100 DNA 12000 kit shows the lower marker at 35 seconds and the upper marker at 90 seconds. The sheared DNA and the upper marker coelute together.*

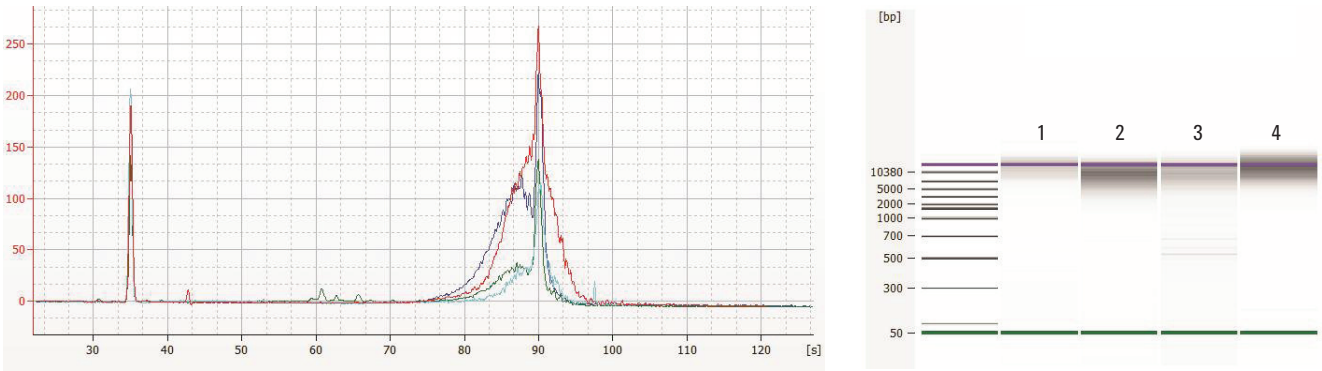


*Figure 6. Representative electropherograms and virtual gel used for visual inspection (generated with the Agilent 2100 Bioanalyzer system with the DNA 12000 Kit) of DNA libraries sizes prepared for sequencing with the PacBio SMRTbell 10 kb Template Preparation Kit on the Agilent NGS Workstation. A typical electropherogram using the Agilent Bioanalzyer 2100 DNA 12000 kit shows the lower marker at 35 seconds and the upper marker at 90 seconds. The DNA libraries and the upper marker coelutes with each other, the sharper peak is the upper marker. The average library sizes are: Campylobacter (green, 9.1 kb), Listeria (blue, 9.5 kb), Vibrio (aqua, 10 kb), and Salmonella (red, 15 kb).*

Table 2 shows that microbial genomes with a different GC content had minimal bias from libraries made on the Agilent NGS Workstation. The GC content observed by sequencing matches closely with the known GC content on each genome and no significant difference was observed. The results demonstrate that the PacBio 10 kb Library Preparation protocol is able to generate SMRTbell libraries from high complexity genomes with no significant coverage or GC bias in a reproducible fashion on a NGS Workstation [13,14].

*Table 2. Library Construction Metrics for Libraries Prepared with Each Bacterium*

| Bacterium | Gram reaction | Approx. genome size (MB) | GC content (%) | Avg. library size (kb) |
|---|---|---|---|---|
| *Campylobacter jejuni* | Negative | 2 | 30 | 9.1 |
| *Listeria monocytogenes* | Positive | 2 | 38 | 9.5 |
| *Vibrio fluvialis* | Negative | 5 (2 chromosomes) | 41 | 10 |
| *Salmonella enterica serovar enteritidis* | Negative | 5 | 52 | 15 |

## Conclusion

Automated protocols of PacBio 10 kb Library Preparation generated similar technical performance to those produced manually. With automation of the 10 kb SMRTbell library preparation, up to 96 samples can be processed simultaneously. This is a marked improvement in throughput over the manual method without comprising the integrity of the sequence coverage or quality.

## Acknowledgement

We gratefully acknowledge the technical assistance provided by Carol Huang, Vivian Lee, Alan Truong, Lucy Cai, Christina Kong, Alex Weimer, and Surene Foutouhi from the laboratory of Dr. Bart Weimer in the School of Veterinary Medicine at the University of California, Davis.

## References


1. Qiagen QIAamp DNA Mini Kit: http://www.qiagen.com/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/qiaamp-dna-mini-kit#resourc
2. N. Kong, *et al*. Production and Analysis of High Molecular Weight Genomic DNA for NGS Pipelines Using Agilent DNA Extraction Kit. *Agilent Technologies Application Note* publication number 5991-3722EN (p/n 200600).
3. B. Ganesan, *et al.* Identification of the Leucine-to-2-Methylbutyric Acid Catabolic Pathway of *Lactococcus lactis*. *Appl. Environ. Microbiol.* **72**:4264-73 (2006).
4. Y. Xie, *et al.* Expression profile of *Lactococcus lactis* ssp. *lactis* IL1403 during environmental stress with a DNA macroarray. *Appl. Environ. Microbiol.* **70**:6738-6747 (2004).
5. Agilent 2200 TapeStation User Manual, Agilent Technologies (p/n G2964-90001).
6. Agilent Genomic DNA ScreenTape System Quick Guide, Agilent Technologies (p/n G2964-90040 rev.B).
7. Covaris User Manual g-TUBE, http://covarisinc.com/wp-content/uploads/pn_010154.pdf
8. G2946-90004 the Agilent 2100 Bioanalyzer 2100 Expert User's Guide.
9. Agilent DNA 7500 and DNA 12000 Kit Guide (Rev. B), Agilent Technologies (p/n G2938-90024).
10. Qubit 2.0 Fluorometer User Manual, Life Technologies: http://tools.lifetechnologies.com/content/sfs/manuals/mp32866.pdf
11. PacBio Procedure & Checklist -Low-Input 10 kb Library Preparation and Sequencing, http://www.pacificbiosciences.com/samplenet/ProcedureChecklistLowInput10kbLibraryPreparationandSequencingMagBeadStation.pdf
12. Agilent NGS Workstation (G5522A)
13. D. Aird, *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**: R18 (2011).
14. E.E. Schadt, *et al.* A window into third-generation sequencing. *Hum. Mol. Genet*. **19**: R227–240 (2010).


## For More Informatiom

For more information on our products and services visit our Website at www.agilent.com/chem.

www.agilent.com/chem


Published in the USA
July 2, 2014
5991-4482EN

Agilent Technologies