Integrating the DNA Integrity Number (DIN) to Assess Genomic DNA (gDNA) Quality Control Using the Agilent 2200 TapeStation System

Application Note

Authors

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

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

Abstract

Next Generation Sequencing (NGS) requires the input of high molecular weight genomic DNA (gDNA) to construct quality libraries for large scale sequencing projects, such as the 100K Pathogen Genome Project. The assessment of DNA integrity is a critical first step in obtaining meaningful data, and intact DNA is a key element for successful library construction. The Agilent 2200 TapeStation System plays an important role in the determination of the DNA quality using the DNA genomic assay. Profiles generated on the 2200 TapeStation System yield information on concentration, allow a visual inspection of the DNA quality, and generate a DNA Integrity Number (DIN), which is a value automatically assigned by the software that provides an indication of integrity (that is, lack of degradation). This application note describes a new software algorithm that has been developed to extract information about DNA sample integrity from the 2200 TapeStation System electrophoretic trace.

The Agilent 2200 TapeStation System

Agilent Technologies

Introduction

Reduced costs and high-throughput methods have rendered microbial whole genome sequencing (WGS) accessible to many applications in infectious disease, food safety, and public health. The production of thousands of genomes 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/ is sequencing 100,000 bacterial pathogens from around the globe. This large scale next-generation sequencing project requires high-throughput procedures for DNA extraction before library construction and sequencing [1].

Genomic DNA (gDNA) extracts are often evaluated on agarose gels, but this approach is not suitable for a high-throughput workflow and automation. Size estimation against a ladder coupled with densitometry to determine concentration often results in low-resolution images, and cannot be automated. Assessment of gDNA quality is crucial, because the next step in library preparation for automated sequencing is DNA shearing, which requires high molecular weight gDNA [2,3]. The Agilent 2200 TapeStation System and associated Agilent Genomic DNA ScreenTape assay has the potential to become the standard in DNA quality assessment and quantification as well as provide the remaining QC checks for the entire work flow [4].

The 2200 TapeStation Analysis Software generates an electropherogram that provides a detailed visual assessment of the DNA size distribution and fragments, virtual gel images, and sample concentration. In addition, the software automatically generates a value referred to as the DNA Integrity Number (DIN) that determines the level of sample degradation as opposed to the classical gel electrophoresis method that inadequately determines sample integrity. These advantages provide a quantitative basis for selecting gDNA samples to proceed with into the next phase of library construction for WGS.

DIN was developed to remove the manual interpretation of the DNA integrity by evaluating the entire electrophoretic trace. The DIN software algorithm allows for the classification of total DNA based on a numbering system from 1 to 10, with 1 being the most degraded and 10 being the most intact (that is, high molecular weight). This algorithm has been derived from approximately 7,000 gDNA traces provided by Genomic DNA ScreenTape users covering samples derived from whole and dried blood, saliva, and human tissues from fresh, frozen, and FFPE sources [4]. The DIN facilitates the interpretation of electropherograms, allows for the comparison of samples, and ensures the repeatability of experiments and quantitation of high-quality gDNA moving into library construction.

Table 1. Bacterial Isolates Used to Investigate DNA Integrity Estimations

Bacterium	Gram reaction	Approx. genome size (Mb)	GC content (%)	Average DIN values
Campylobacter	Negative	1.7	30	8.8
Staphylococcus	Positive	2.8	32	8.9
Listeria	Positive	2	38	8.9
Escherichia	Negative	5	51	8.3
Salmonella	Negative	5	52	8.6

Methods

As with all whole genome sequencing projects, the 100K Pathogen Genome Project sample preparation workflow begins with isolation of high molecular weight gDNA followed by quality control metrics (intact gDNA, A260/230, and A260/280 ratios) prior to production of sheared DNA for library construction. Specific bacterial isolates with a range of different GC content and genome sizes were chosen to validate the DNA integrity using a 2200 TapeStation System. After lysis, gDNA was isolated using the Qiagen QIAamp DNA Mini Kit (51306) using the manufacturer's instructions [5,6]. The isolated gDNA was analyzed using the 2200 TapeStation System for high molecular weight prior to shearing and library construction to obtain the DIN value [7-10].

Results and Discussion

Bacteria samples with a range of % GC content were obtained using the Qiagen QIAamp Mini Kit. Genomic DNA was analyzed on the 2200 TapeStation System with the Genomic DNA ScreenTape assay to obtain electropherograms, and resulted in high molecular weight gDNA gel image similar to an agarose gel (Figures 1 and 2). The gDNA data files used were from already constructed libraries and were re-analyzed with the 2200 TapeStation Analysis Software (version A.01.05) to obtain the DIN value. This version of the software results in the display of an electropherogram, virtual gel as well as a DIN value calculated from the electropherogram of the gDNA. The DIN value indicates the intactness of the DNA, giving a qualitative measure of the integrity, which can be used to compare across samples (Figure 2). This value can be used before proceeding with the next steps in library construction.

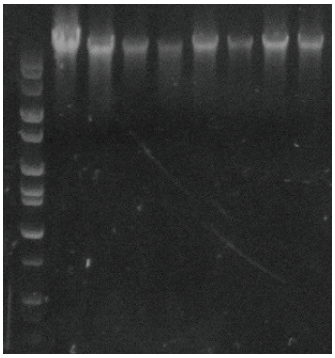

Figure 1. Classical agarose gel with the upper marker of 10 kb.

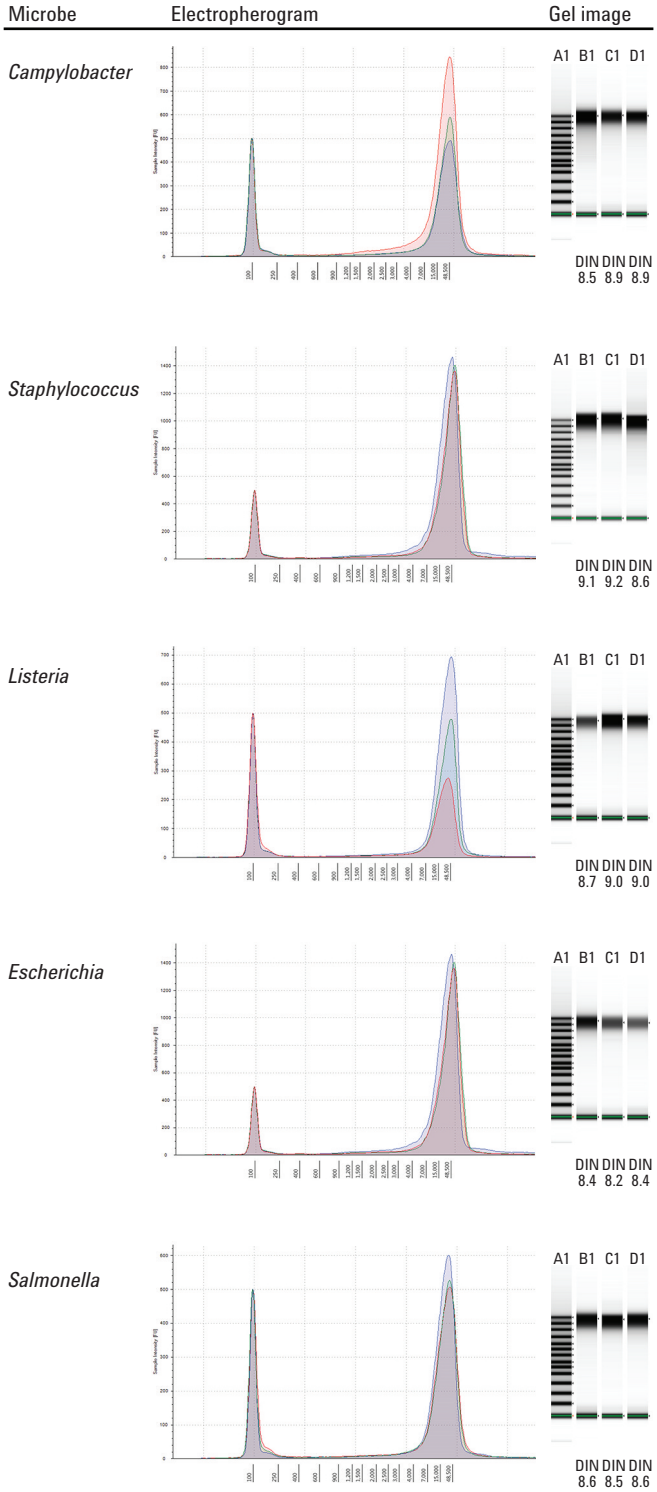

Figure 2. Electropherograms and gel images of Genomic DNA from Agilent 2200 TapeStation with DIN values.

Three independent isolates of each bacteria were chosen to be re-analyzed with the software, with an average DIN of 8.3 to 8.9 (Table 1), which showed that the gDNA input data were acceptable. The input data include samples with a predefined numeric system from 1 to 10. The gDNA input is shown with the electropherograms to illustrate the DIN in the software ranging from intact (DIN 9.2), to degraded (DIN 1.1) in Figures 3-4. The specifications for the Genomic DNA ScreenTape System indicates that the linear concentration range for samples is 10–100 ng/µL, and that the DIN functional range is from 5–300 ng/µL [8]. Within the Analysis software, the concentration of gDNA is shown under the samples (data not shown). The electropherogram and gel for the DIN 1.1 sample shows that this degraded sample is too dilute to be within useful range, while the concentrations of the better quality DIN gDNA samples were comfortably with these working ranges.

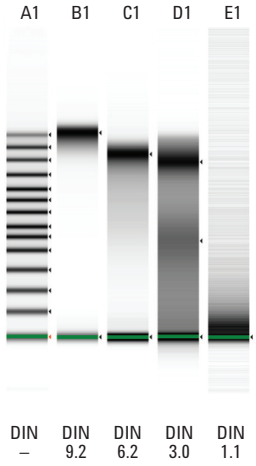

Figure 4. Sample bacteria gel image that correspond to the electropherogram of the DIN ranges from Figure 3.

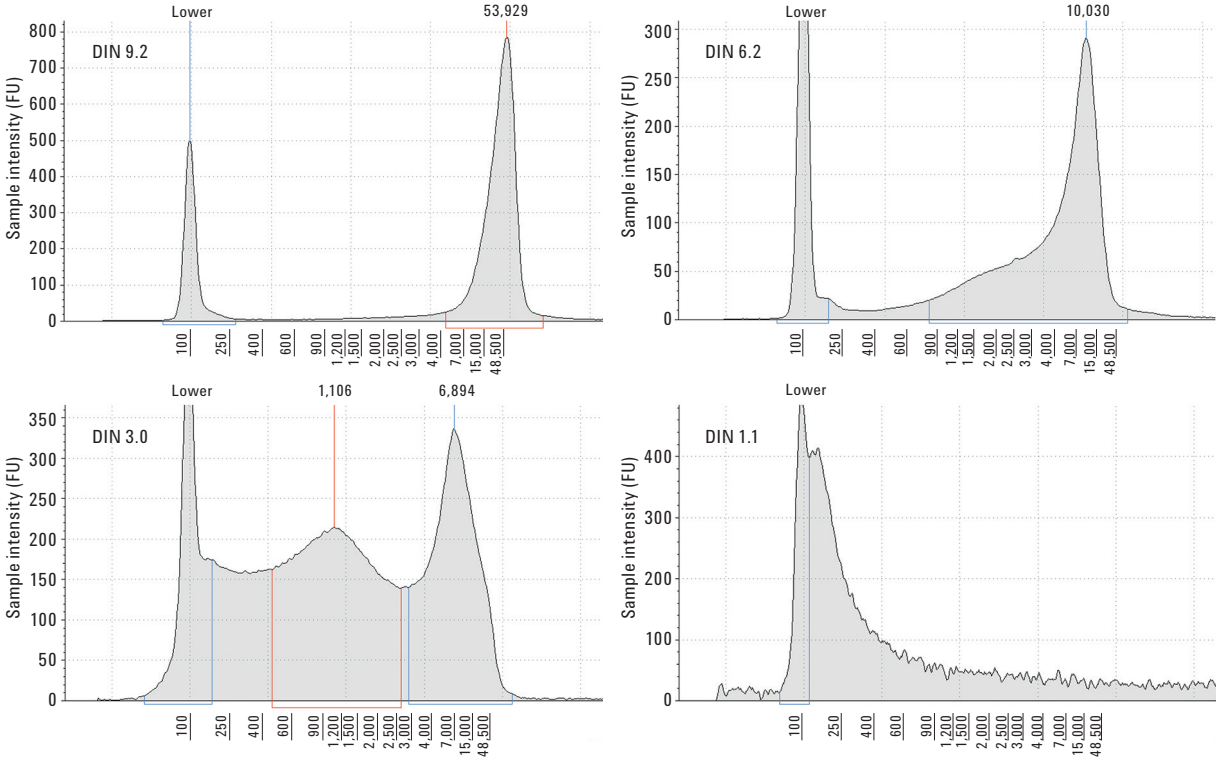

Figure 3. Sample electropherograms to show the DIN in the software. Samples range from intact (DIN 9.2), to degraded (DIN 1.1).

It was determined that samples with a DIN of > 7 were acceptable to progress into the next step of library construction (Table 1). Figure 5 shows a gDNA image with the average DIN of 8.6 that produced quality final libraries with an average size of 267 bp. However, a degraded gDNA sample (DIN of 6) produced a final library with an average size of 198 bp, which is out of the acceptable range of the typical 250–500 bp final library requirement for WGS sequencing (Figure 6). The Agilent 2200 TapeStation with the Genomic DNA ScreenTape assay in the new software update automatically determines the DIN value using their new algorithm for each gDNA sample. In this assay, the libraries produced from gDNA with a higher DIN were better quality than those produced from gDNA in the lower range. The DIN number was successfully used to assess the quality of gDNA.

Conclusion

Agilent Technologies has designed a software algorithm that is capable of assessing DNA quality to produce a quantitative measure of quality. The DIN algorithm was developed to remove user dependent interpretation of DNA quality and to provide a standardized assessment. However, successful library construction is dependent on several variables. It is essential that other quality assessments are made in addition to the new DIN software to achieve optimum results. Characterization of gDNA samples with DIN is independent of the instrument, sample concentration, and the operator allowing for unbiased comparison of the samples. The researcher is no longer tied to arbitrary classification of total DNA, and it can be used to ensure the consistency of library construction.

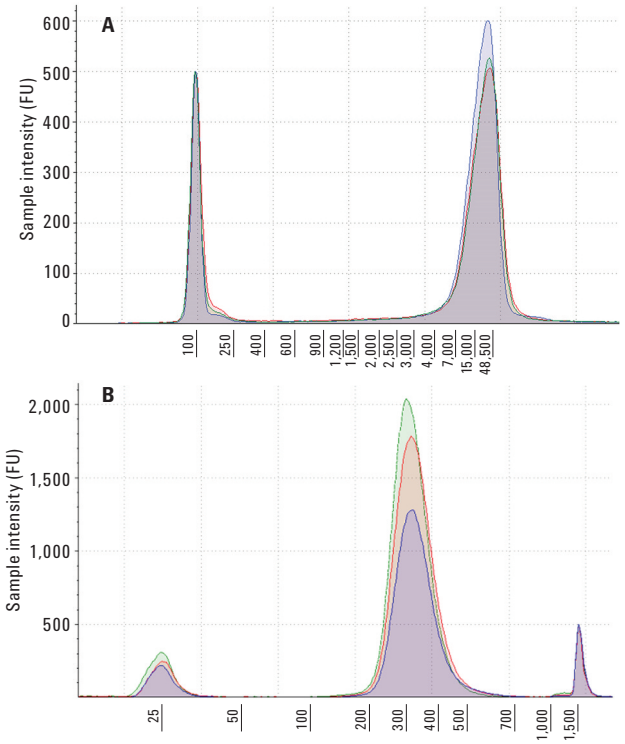

Figure 5. Three electropherograms of input gDNA from Salmonella (A) that produced the examples of quality final libraries (B), shown assayed with the Agilent D1000 ScreenTape assay.

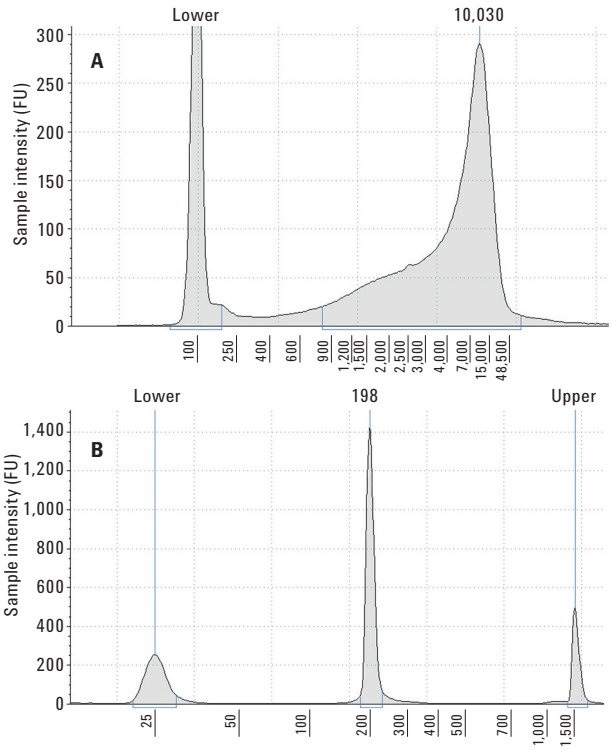

Figure 6. An example electropherogram of input gDNA with a DIN of 6 (A) that produced a low quality final library (B), shown assayed with the Agilent D1000 ScreenTape assay.

Acknowledgement

We gratefully acknowledge the technical assistance provided by Carol Huang, Regina Agulto, San Mak, Kendra Liu, Patrick Ancheta and Christina Kong from the laboratory of Dr. Bart Weimer at University of California, Davis.

References

1. 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)**.
2. M. A. Quail, *et al*. "A large genome center's improvements to the Illumina sequencing system" *Nature Methods* **5**,1005-1010 (2008).
3. S. Wilkening, *et al*. "Genotyping 1000 yeast strains by next-generation sequencing" *BMC Genomics* **14**:90 (2013).
4. M. Gassmann & B. McHoull, DNA Integrity Number (DIN) with the Agilent 2200 TapeStation System and the Agilent Genomic DNA ScreenTape Assay, *Agilent Technologies*, publication number 5991-5258EN (**2014**).
5. Qiagen QIAamp DNA Mini Kit: http://www.qiagen.com/us/products/catalog/sample-technologies/dna-sample-technologies/genomic-dna/qiaamp-dna-mini-kit/
6. 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**).
7. N. Kong, *et al*. "Quality Control of High-Throughput Library Construction Pipeline for KAPA HTP Library Using Agilent 2200 TapeStation" *Agilent Technologies*, publication number 5991-5141EN (**2014**).
8. "Agilent Genomic DNA ScreenTape System Quick Guide" *Agilent Technologies*, publication number G2964-90040 rev.B (**2013**).
9. "Agilent 2200 TapeStation User Manual" *Agilent Technologies*, publication number G2964-90002 Rev. B (**2013**).
10. "Agilent High Sensitivity D1K ScreenTape System Quick Guide" *Agilent Technologies*, publication number G2964-90131 Rev. B (**2013**).

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
December 18, 2014
5991-5442EN

Agilent Technologies