diagenode

Innovating Shearing Solutions

Wouter Coppieters
Head of the genomics core facility
GIGA center, University of Liège

Bioruptor® NGS: Unbiased DNA shearing for Next-Generation Sequencing

Introduction

Wouter Coppieters, PhD, and his colleagues operate the genomics core facility at the GIGA center, University of Liège (Belgium). The center routinely performs a variety of Next-Generation Sequencing (NGS) applications including de novo sequencing, whole genome and amplicon sequencing, and targeted resequencing on an Illumina $GAII_x$ analyzer. Coppieters recognizes that optimal sequencing results rely heavily on the quality and preparation of DNA libraries. Library preparation can be complex with multiple steps that require optimization. As Coppieters notes, "A prerequisite to obtaining high quality Next-Generation Sequencing data is to have libraries that have the desired insert size (November 2011)". In addition, obtaining good NGS data depends on a large number of factors such as the employed platforms, protocols, sample types, and experimental conditions in the library preparation. Each sample preparation step, including DNA shearing, must be as optimal as possible to avoid uneven read coverage. Also, libraries from genomes that are difficult to sequence (e.g. AT rich) may require additional optimization such that they provide an unbiased representation of the whole genome.

To reduce the complexity of library preparation for NGS, Diagenode has launched the Bioruptor® NGS fully optimized for NGS applications. The Bioruptor® NGS sonicator is proven as the shearing device of choice for library preparation and provides superior sample yields, fragment size, and consistency, which are all essential to cost-effective Next-Generation Sequencing workflows.

The GIGA center, in collaboration with Diagenode, incorporated the Bioruptor® NGS into its Next-Generation Sequencing experiments and illustrated its power in optimal library preparation for exceptional NGS results.

Deep sequencing of *E. coli* and *S. epidermis* genomes

Coppieters' team sequenced two bacterial genomes with dissimilar GC characteristics: 1) the well-characterized *Escherichia coli* (reference ATCC 8739), which has a fairly even GC-AT distribution (though some local regions contain outstanding GC content) and 2) the skin commensal *Staphylococcus epidermis* (reference ATCC 12228) whose genome is extremely AT rich.

The reads from whole genome sequencing are used for reference assembly, where high coverage is particularly important for uncovering rare genetic variations. Coppieters' team wanted to achieve ultra-high coverage with equal representation of every part of the genome. To accomplish this challenging goal, they chose the Bioruptor® NGS for reliable sample preparation. The following section details the relevant laboratory protocols used for sample preparations. Table 1 shows the characteristics of the sequencing data and the genomes.

Sample	Genome size (Mb)	GC content	Read length	Number of reads	Read coverage
E. coli	4,75	50,9%	2x75 bp	7963513	255x
S. epidermis	2,5	32,1%	2x75 bp	14446582	878x

Table 1: Summary of figures of the *E. coli* and *S. epidermis* genomes and the sequencing performed on an Illumina Genome Analyzer.

Workflow

Bacterial genomic DNA was prepared using the DNeasy Blood & Tissue kit (Qiagen) according to manufacturer's instructions and concentrated using the PrepEase PCR purification 96-well Plate kit (USB).

In the first step of library preparation, the Bioruptor® NGS was used to sonicate 5 µg of sample DNA (100 µl final sonication volume in TE buffer) for 8 minutes of 30 seconds on/off cycles at 4°C. The sheared DNA was analyzed on a Bioanalyzer 2100 using the DNA High Sensitivity chip.

The remaining sample preparations and sequencing were performed with standard Illumina procedures required for generating 75 base long reads from the paired end libraries using a single flow lane for the two bacterial strains.

Results

The shearing performance was closely monitored during the whole workflow using the Bioanalyzer 2100 and its DNA High Sensitivity chips, which are the gold standard for quickly and accurately checking DNA fragment sizes after shearing. The Bioruptor® NGS produced a more than adequate number of fragments, which fit tightly in the specified size range. The default Bioanalyzer output in log scale is shown in Figure 1 for both bacterial strains. Insets represent the linear transformation of the data.

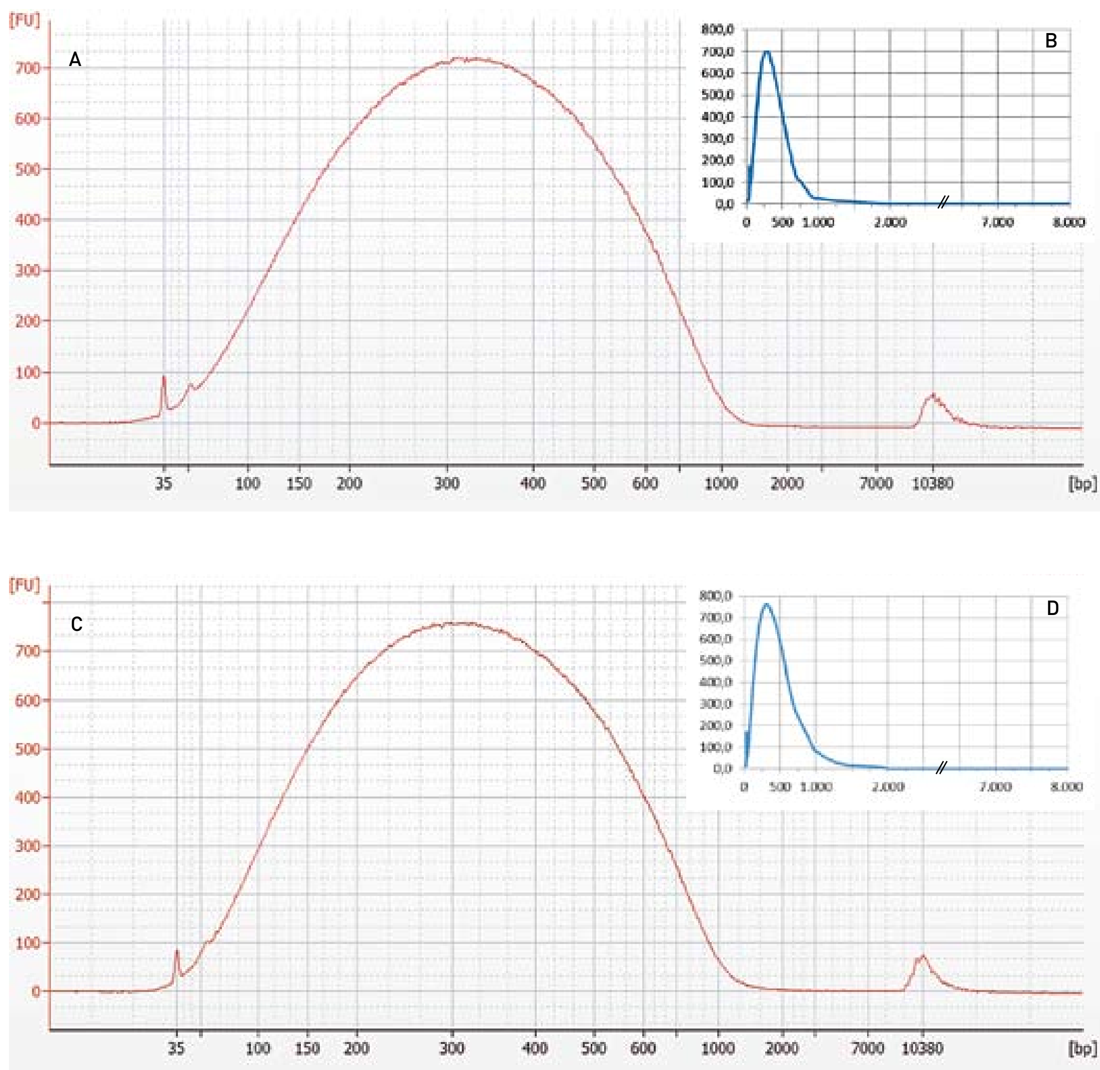


Figure 1: The Bioanalyzer images (A and C) and their linear transformation (B and D) showing the fragment distribution from the *E. coli* (top) and *S. epidermis* experiments (bottom).

Figure 2 allows the visual verification of the *E. coli* and *S. epidermis* fragments. The consistency of the shearing results was also controlled with standard gel electrophoresis between the different samples. Together, these results show that the shearing produces a tight and reproducible distribution along the expected size range.

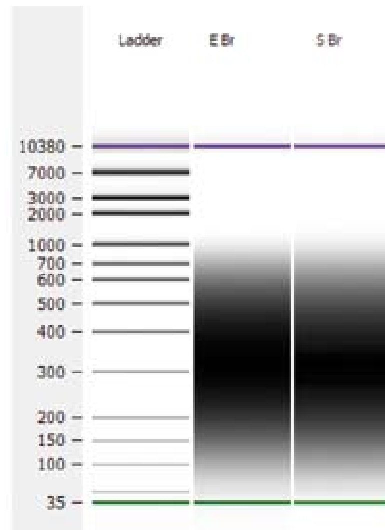


Figure 2: The sheared fragments clearly show a reproducible distribution along the expected size range regardless of sample base content. (left: *E. coli*, right: *S. epidermis*).

With such optimal shearing results, it was not surprising that the subsequent sequencing run resulted in high quality data with extensive coverage and minimal bias.

Figure 3 presents the even and ultra-high coverage along the whole genome of the bacteria. Proper shearing accomplished by the Bioruptor® NGS prevented undesirable gaps or large peaks (under- or overrepresentations) in the dataset.

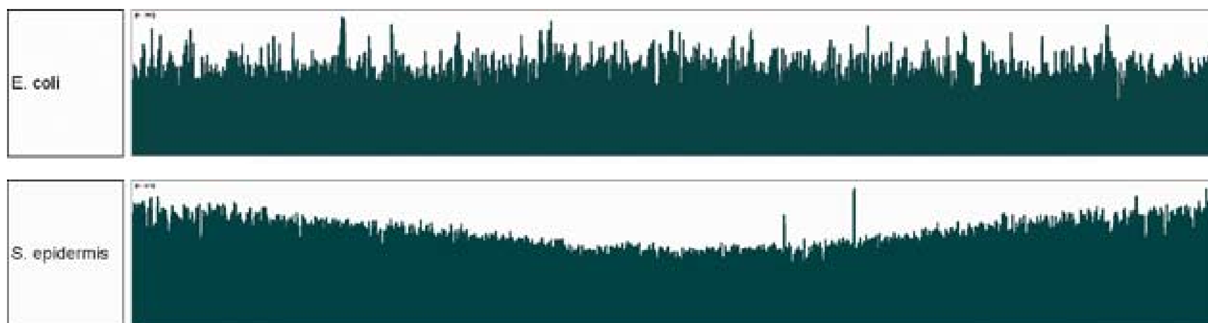


Figure 3: The *E. coli* (top) and the *S. epidermis* (bottom) genomes have uniform and high coverage. The bar charts show read numbers per base.

The sequencing depth frequency was not biased towards lower or higher depths, as shown in the histograms in Figure 4. The coverage frequency of the whole genomes sheared with a Bioruptor® NGS shows a normal distribution.

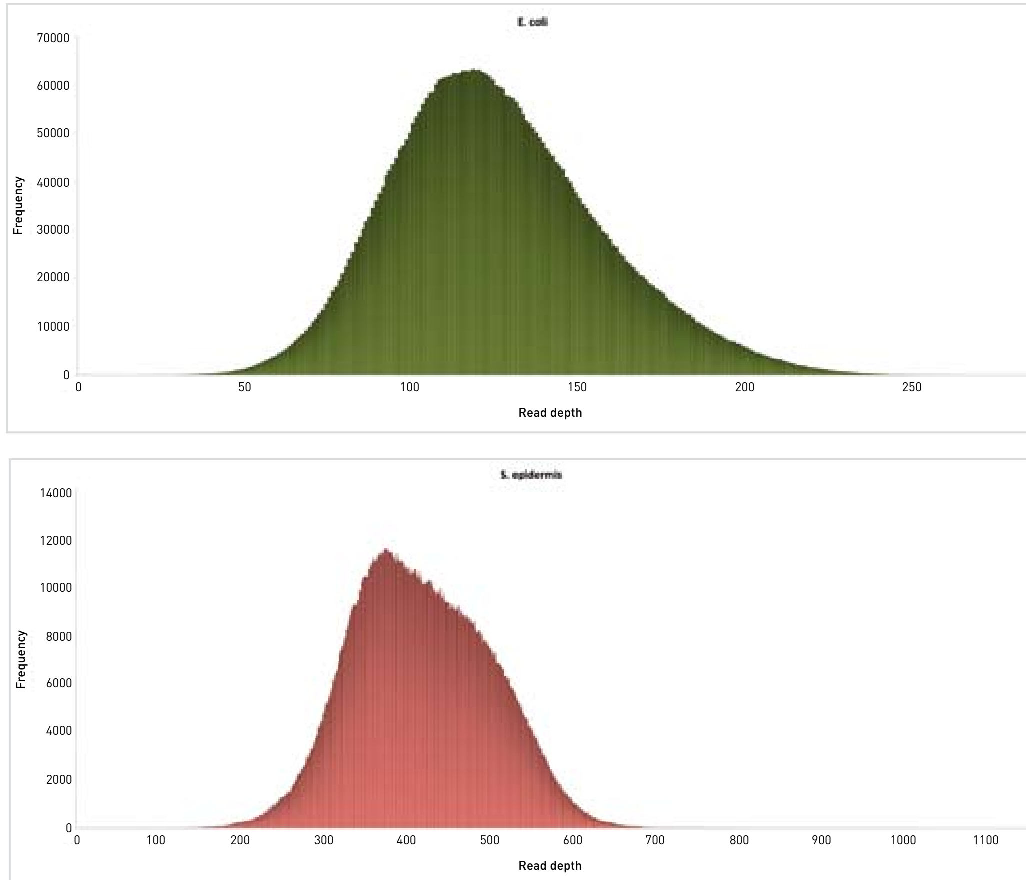
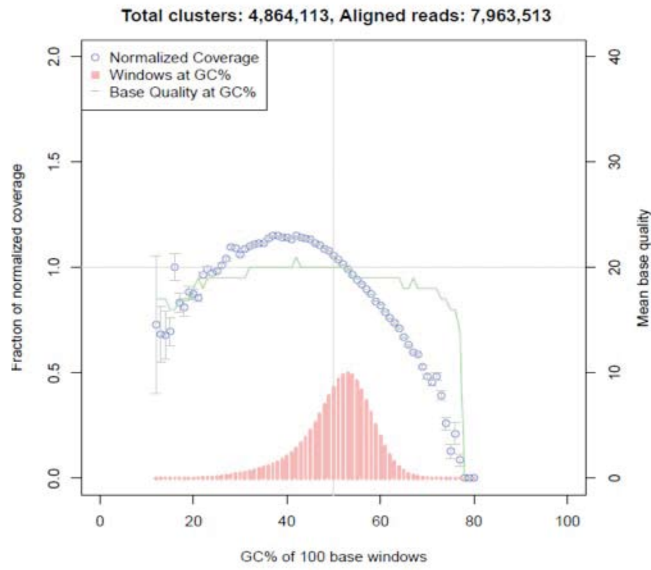


Figure 4: The read depths for every base are shown in the function of their frequency, producing a normal distribution. The X axes show the coverage values per base, and the Y axes show how many times a coverage value occurs.

The correlation between GC skewing and the sequencing data is important in understanding the level of bias in the dataset. We thus introduced different QC steps in the workflow to measure bias. Figure 5 illustrates nearly equal representation of the whole genome with coverage summits characteristic of the bacteria with consistent, high quality data in almost every GC range. The fact that coverage depth drops significantly in extreme GC sequences (> 70%) is usually caused by the PCR steps during sample preparation and the sequencing chemistry itself. (*)

A closer view in Figure 6 also depicts the even and ultra-high coverage in AT rich regions.



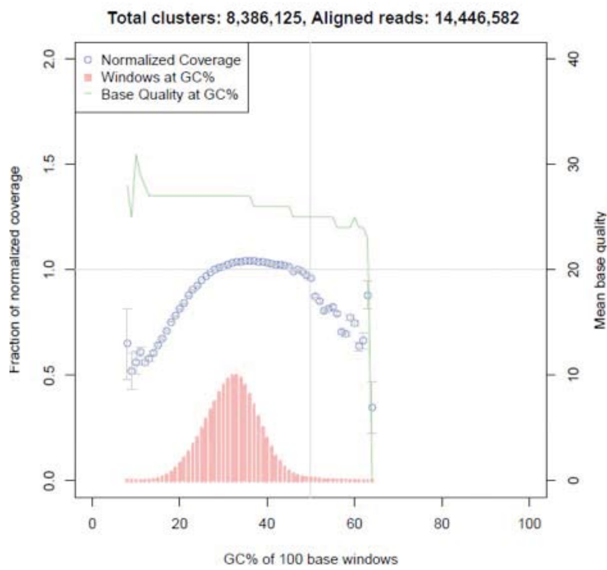


Figure 5 : The plots (top: *E. coli*, bottom: *S. epidermis*) summarize the effects of genomic GC bias on the sequencing results. The optimal shearing enable unbiased, smooth coverage for the entire GC range.

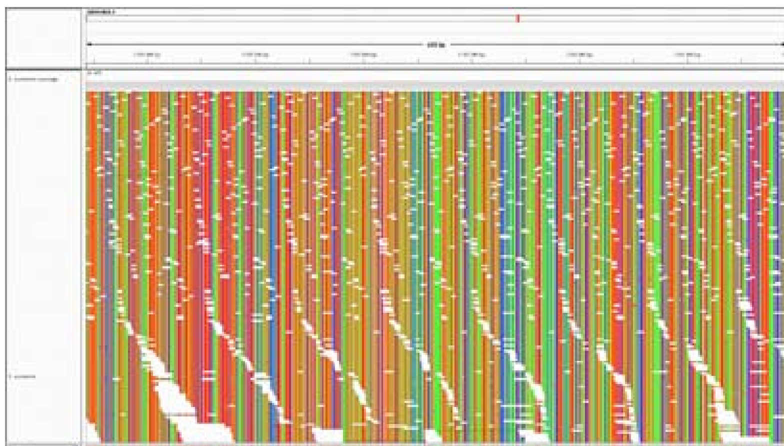


Figure 6: Reliable and even coverage of a region with high AT content (*S. epidermis* 1537-1538 kbp)

Conclusion

Critical to the success of Next-Generation Sequencing results is the preparation of high-quality material prior to the sequencing run. Library preparation requires accurate DNA fragmentation to maximize yields, quality, and cost effectiveness of sequencing experiments. Although different shearing methods and technologies exist (i.e. enzymatic shearing or AFA), this study demonstrates that mechanical shearing with the Bioruptor® NGS using Diagenode's well-established ACT technology is highly effective in creating unbiased, high-quality genomic libraries with even distribution and representation of the genome. In addition to optimal library construction, the Bioruptor® NGS enables high-throughput fragmentation by simultaneously shearing up to 12 samples. The Bioruptor® NGS was specifically designed for Next-Generation Sequencing and allows researchers to prepare NGS libraries at the fraction of the cost of other sonication technologies and to focus their valuable time analyzing data instead of optimizing sample preparation.

	Bioruptor® Sonicator	Competitor C Sonicator
Desirable fragment sizes for sequencing	Yes	Yes
Consistent fragment sizes from 100 bp – 1500 bp	Yes	Yes
Multiplexing capability	Yes	No
High-throughput (12 tubes)	Yes	No
Simple operation	Yes	Yes
Cost of operation	Yes	No

Table 2: Comparison made between the Bioruptor® NGS and Competitor C.

"Although several options exist, we routinely generate high quality libraries (using sonication) with the **Bioruptor®** Plus and more recently with the Bioruptor® NGS. Therefore, our libraries for our prospective genomic characterization of the German EHEC 2011 outbreak (PLoS ONE. 2011 [67]: e22751 [PubMed]) with the **Ion Torrent** PGM® were generated using the Bioruptor®", explained microbiologist Prof. Dr. Dag Harmsen from the Department of Periodontology of the University Münster in Germany. Dr. Harmsen is the corresponding author of the PLoS study and leads the team responsible for sequencing and conducting the bioinformatics analysis in Münster.

Acknowledgement:
Our thanks for this collaborative project to Wouter Coppieters, Benoit Hennuy and Latifa Karim of the genomics core facility at the GIGA center

(*) Kozarewa et al., Nature Methods 2009, 6: **Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes.**

Aird et al., Genome Biology 2011,12: **Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.**

CS-Wouter_09_11_11