NEBNext UltraShear® Long Read: Novel and time-dependent enzymatic DNA fragmentation for long read sequencing

Brittany S. Sexton, Matthew Angel, Karen McKay, Jonathan Sanford, Ruby Moulton, Louise Williams, Bradley W. Langhorst, and V.K. Chaithanya Ponnulari | New England Biolabs, Inc.

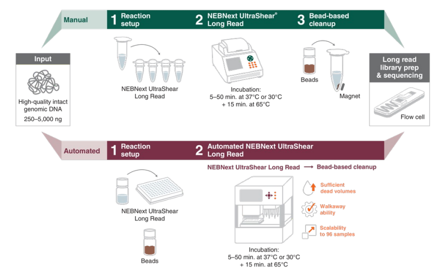
Introduction

Ultra-long DNA, if not fragmented, is typically not made into libraries or sequenced on long read platforms efficiently. DNA fragmentation upstream of long read sequencing typically can result in comparable fragment sizes across samples and experiments. DNA fragmentation results in higher and consistent N50 lengths as well as improved sequencing yields compared to ultra-long DNA. Current methods for DNA fragmentation include mechanical shearing and enzymatic fragmentation. Mechanical shearing (e.g., Covaris® g-TUBE and Megaruptor®) is the gold standard method for DNA fragmentation upstream of long read library preparation, however it does require expensive consumables and/or instruments, is not automation-friendly, and results in sample loss. In comparison, enzymatic fragmentation methods do not require expensive instruments and are automation friendly; unfortunately, the fragment sizes generated may not be compatible with long read sequencing.

To address these constraints, we developed a novel enzymatic fragmentation solution, NEBNext UltraShear® Long Read (UltraShear LR), that is quick, tunable, and automation friendly. Enzymatic fragmentation is time-dependent and can be used to generate a wide-range of DNA fragment sizes (2 to 30 kb) suitable for different sequencing platforms and applications. UltraShear LR is robust across a wide range of gDNA input amounts (250 to 5,000 ng) as well as different gDNA samples and species (e.g., animal, plant and human). Here we demonstrate that UltraShear LR fragmentation generates high quality libraries with tunable read lengths that are sequenced using Oxford Nanopore™ Technologies (ONT) and PacBio® platforms. UltraShear LR libraries retain base modifications (including CpG methylation) and identify more CpGs than mechanically sheared libraries at the same read count.

The time-tunable enzymatic fragmentation of DNA using the UltraShear LR system is robust and cost effective when compared to mechanical fragmentation. UltraShear LR overcomes many limitations of mechanical shearing methods by simplifying sample processing, increasing throughput, and preserving base modifications. These advantages combine to enhance usability and quality when preparing long read sequencing libraries.

Methods



UltraShear LR is compatible with manual and automated workflows. The UltraShear LR workflow accommodates a wide input range, from 250 ng to 5,000 ng of gDNA, and can be completed in as little as 30 minutes. UltraShear LR is automation friendly with minimal sample transfers, sample tracking flexibility, and sufficient reagent overages to account for automation dead volumes. UltraShear LR fragmentation integrates seamlessly with ONT and PacBio® library preparation and sequencing.

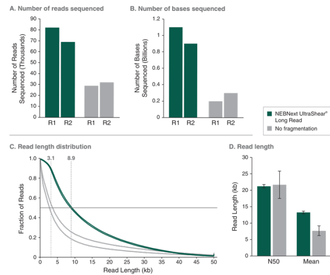
Sample Input: 1,000 ng of gDNA (DQN = 9.9), extracted from GM12878 human cells using the Monarch® HMW DNA Extraction Kit was fragmented using UltraShear LR, unless noted.

Library Prep and Sequencing: Libraries were generated with ONT Native Barcoding Kit 96 V14 and sequenced on a P2 Solo instrument, or with PacBio SMRTbell® prep kit 3.0 and sequenced on a PacBio Sequel® II Instrument unless noted.

Bioinformatics Analysis: ONT sequencing metrics were aggregated from Nanostats. S-shaped curves were generated by plotting the fraction of reads at a particular read length or greater (in 100 bp bins). Reads longer than 50 kb are represented as 50 kb. The vertical dash lines represent the read length at 0.50 fraction of reads. GC Coverage plots: The GC coverage plots were analyzed using Picard and the distribution of normalized coverage plotted across different GC contents of the genomes (0–100%).

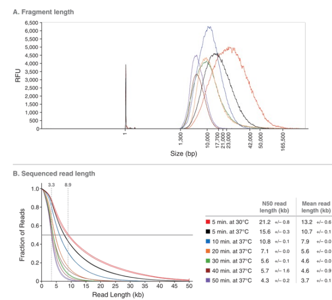
Results

UltraShear Long Read fragmentation improves sequencing yield while maintaining N50 values comparable to no fragmentation



gDNA was fragmented for 5 minutes at 30°C followed by 15 minutes at 65°C with UltraShear LR. 100 ng of UltraShear LR fragmented gDNA and non-fragmented gDNA were used to generate libraries for Oxford Nanopore Technologies (ONT) sequencing. (A & B) The UltraShear LR libraries generated higher sequencing yield (number of reads and bases sequenced) compared to the non-fragmented gDNA libraries. Technical replicates are shown across all conditions. (C) The libraries generated using UltraShear LR fragmentation result in a higher proportion of data at longer read lengths compared to non-fragmented libraries. (D) The UltraShear LR fragmented libraries have similar N50 read lengths and increased mean read lengths compared to non-fragmented libraries. The average of two technical replicates and standard error are shown for each condition.

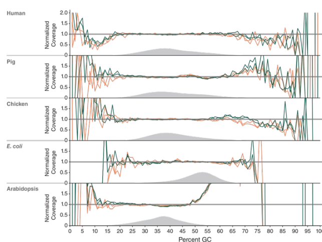
UltraShear Long Read is a time-dependent enzymatic fragmentation method



	N50 read length (kb)		Mean read length (kb)	
5 min. at 30°C	21.2	+/– 0.8	13.2	+/– 0.6
5 min. at 37°C	15.6	+/– 0.3	10.7	+/– 0.1
10 min. at 37°C	10.8	+/– 0.1	7.9	+/– 0.0
20 min. at 37°C	7.1	+/– 0.0	5.6	+/– 0.0
30 min. at 37°C	5.6	+/– 0.1	4.6	+/– 0.0
40 min. at 37°C	5.7	+/– 1.6	4.6	+/– 0.9
50 min. at 37°C	4.3	+/– 0.2	3.7	+/– 0.1

gDNA was fragmented for 5 minutes at 30°C or 5 to 50 minutes at 37°C followed by 15 minutes at 65°C with UltraShear LR. (A) Agilent® Femto Pulse® trace (165kb protocol) showing the time-dependent shifts in fragmentation size and profile of gDNA fragmented using UltraShear LR. Following fragmentation, 100 ng of fragmented DNA was used to generate libraries for ONT Sequencing. (B) S-shaped curve plotting the read length distributions, generated using gDNA fragmented from 5 minutes at 30°C to 50 minutes at 37°C, showcasing the tunability of UltraShear LR. The table summarizes the sequenced N50 and mean read lengths across different fragmentation conditions. The average of two technical replicates and standard error are shown for each condition.

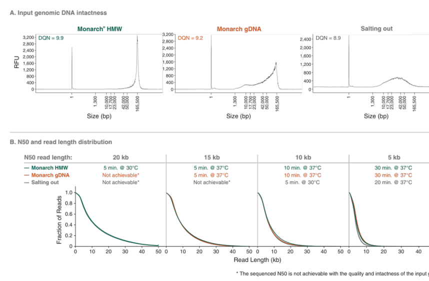
UltraShear Long Read generates even GC Coverage for different gDNAs



1,000 ng of gDNA from different species (human, pig, chicken, E. coli, and Arabidopsis) was fragmented for 5 minutes at 37°C followed by 15 minutes at 65°C with UltraShear LR. Following fragmentation, 100 ng of fragmented DNA was used to generate libraries with ONT Native Barcoding Kit 96 V14 and sequenced on a P2 Solo instrument (green), or with PacBio SMRTbell® prep kit 3.0 and sequenced on a PacBio Sequel® II Instrument (orange). Dorado 0.8.3. was used for ONT base calling and approximately 100,000 reads for ONT libraries and 60,000 reads for the PacBio libraries were aligned to the appropriate genomes (Human T2T, s_scrofa_v11.1, Gallus_gallus, eschColi_BL21, and ColCEN respectively) using Minimap 2. The libraries generated using UltraShear LR fragmented DNA have uniform GC coverage across different gDNA samples for both ONT and PacBio platforms.

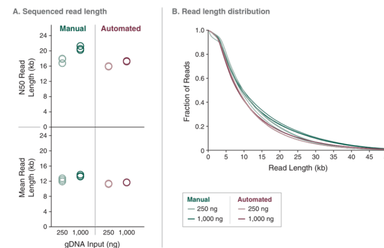
Results Continued

UltraShear Long Read is compatible with gDNA extracted from different methods



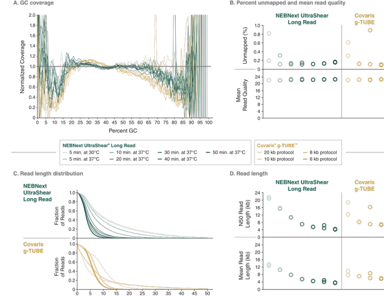
gDNA was extracted from GM12878 human cells using the Monarch® HMW DNA Extraction Kit, Monarch Spin gDNA Extract Kit, or salting out extraction methods resulting in DQNs of 9.9, 9.2 and 8.9 respectively. (A) Agilent® Femto Pulse® trace (165 kb protocol) showing the intactness of the extracted gDNA. 1,000 ng of each input was then fragmented for 5 minutes at 30°C or 5 to 30 minutes at 37°C followed by 15 minutes at 65°C with UltraShear LR. Following fragmentation, 100 ng of fragmented DNA was used to generate libraries for ONT sequencing. (B) The fragmentation conditions to generate sequenced N50 read lengths of 20 kb, 15 kb, 10 kb, and 5 kb and the read length distributions (S-shaped curves) for each extraction input are shown. If the extraction method is not listed, the sequenced N50 is not achievable with that quality and intactness of input gDNA. The S-shaped curves overlap well across the different input gDNA qualities under specific N50 conditions, demonstrating the tunability of UltraShear LR.

UltraShear Long Read is automation friendly



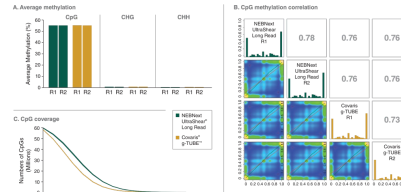
250 ng and 1,000 ng of gDNA was fragmented with UltraShear LR using either a manual workflow or an automated workflow on the SPT Labtech FireFly®. Fragmentation conditions were 5 minutes at 30°C followed by 15 minutes at 65°C. (A) Sequenced N50 and mean read lengths obtained using UltraShear LR fragmentation under manual or automated conditions. (B) S-shaped curve plotting the read length distributions. Manual fragmentation with UltraShear LR generated longer read lengths than the automated workflow, due to potential additional shearing by the instrument, which could be gDNA or platform dependent. However, automated fragmentation was more consistent across replicates. Three technical replicates are shown for each condition.

UltraShear Long Read generates high-quality libraries



gDNA was fragmented for 5 minutes at 30°C or 5 to 50 minutes at 37°C followed by 15 minutes at 65°C with UltraShear LR or with Covaris® g-TUBE™ using the 20 kb, 10 kb, 8 kb or 6 kb protocols. Following fragmentation, 100 ng of fragmented DNA was used to generate libraries for ONT sequencing. Dorado 0.8.3. was used for ONT base calling and approximately 100 thousand reads were aligned to the human T2T genome. (A) Libraries prepared using UltraShear LR fragmented DNA generate uniform GC coverage across different fragmentation conditions. (B) UltraShear LR fragmentation conditions produce high-quality ONT libraries and sequencing data (% unmapped and mean read quality). (C) S-shaped curve plotting the read length distributions. The profile observed within technical replicates for each condition is consistent for gDNA fragmented with UltraShear LR compared to Covaris g-TUBE. (D) Sequenced N50 and mean read lengths obtained using UltraShear LR fragmentation conditions are more tunable and reproducible than those achieved with Covaris g-TUBE conditions.

UltraShear Long Read maintains DNA methylation



1,000 ng of GM12878 gDNA (commercially purchased salting out extraction; DQN = 8.8) was fragmented with UltraShear LR (5 minutes at 30°C followed by 15 minutes at 65°C) or Covaris g-TUBE (20kb protocol) with two technical replicates each. Following fragmentation, 400 ng of fragmented DNA was used to generate libraries for ONT sequencing. The data was combined from six PromethION flow cells. Dorado 0.9.1 was used for base calling and approximately 4 million reads were aligned using Minimap2 to the Human T2T reference. (A) The percent of aggregated methylation in all contexts are similar for libraries prepared using UltraShear LR and Covaris g-TUBE. (B) CpG methylation correlations were generated for libraries prepared using UltraShear LR -fragmented and Covaris g-TUBE-sheared gDNA using methylKit. Approximately 52 million CpGs were common to all libraries at a 1X sequencing depth. There is high correlation between CpG methylation of UltraShear LR and Covaris g-TUBE libraries. (C) Approximately, 60 million and 58 million CpGs were identified at 1X coverage for libraries generated using UltraShear LR and Covaris g-TUBE respectively. UltraShear LR enables coverage of higher number of CpGs compared to Covaris g-TUBE at all effective coverage depth. The average coverage of ~10X for both fragmentation methods was obtained by combining data from six PromethION flow cells sequenced for 72 hours.

Conclusions

- Provides time-dependent enzymatic fragmentation
- Generates high-quality libraries
- Supports automated workflows
- Compatible with ONT® and PacBio®
- Is robust & reproducible
- Improved yields compared to no fragmentation
- Generates even GC coverage
- Maintains DNA methylation

Thank you to the NEB Sequencing Core, Evan Janzen, Kelly Zatopek, Anagha Kadam and Giron Koetsier for your support.