MuA Transposase Enzyme Enables Fast And Easy DNA Library Preparation For Next Generation Sequencing

Authors: **Gediminas Alzbutas, Sanna Askolin, Julius Gagilas, Simona Gliebutė, Heli Haakana, Juuso Juhila, Ian Kavanagh, Laura-Leena M. Kiiskinen, Arvydas Lubys, Justas Morkūnas, Minna Päiväsaari, Edita Povilaitienė, Jurgita Rubekina, Ossian Saris, Remigijus Skirgaila, Romas Tamoševičius, Mindaugas Ukanis**

Thermo Fisher Scientific, Vantaa, Finland & Vilnius, Lithuania

Abstract

Currently many different methods are used for Next Generation Sequencing (NGS) library preparation. Two major steps in the NGS library preparation workflow are the DNA fragmentation and the addition of adaptors for sequencing. While DNA can be fragmented using either enzymatic or physical shearing, the latter method requires additional equipment or has reproducibility issues. Following enzymatic or physical shearing, DNA fragments later in the workflow have their ends repaired and sequencing platform specific adaptors ligated. In contrast to conventional methods, library preparation using transposases can significantly simplify the process by combining the two steps into one. Transpososome treated DNA is cleaved into fragments containing uniform ends which are then easily modified to adapt the fragment library for various sequencing platforms. We demonstrate a library preparation protocol using the MuA transposase enzyme, which enables fast and easy NGS library preparation. In our workflow, MuA transposase enzyme simultaneously catalyzes fragmentation of double-stranded target DNA and tagging of the fragment ends with transposon DNA. A subsequent PCR step adds the platform-specific adaptors generating high yield of DNA fragments with the inserts ranging from 100 bp to 1000 bp.

Introduction

In this work a preformed MuA transposase/transposon complex was used to prepare libraries from genomic DNA of three different microorganisms: *Staphylococcus aureus* (GC% 32.9), *Escherichia coli* (*GC% 50.8)* and *Rhodopseudomonas palustris* (GC% 65.0)*. S. aureus* and *R. palustris* genomes have extreme GC content which may impact uniformity of resulting fragment distribution during library preparation. Here we show that MuA transpososome-based fragmentation is well suited for library preparation of *E. coli* and for both high- and low-GC content genomes starting from as little as 100 ng of input gDNA. MuA-generated libraries were subsequently used in template preparation reactions and PGM-sequenced in order to evaluate the length of sequencing reads. These reads should correspond with the length of size-selected fragments. Moreover, sequencing data of all three organisms were used to determine MuA target site selectivity, which is known to only weakly deviate from random (1), making MuA transposase ideally suited for NGS library preparation.

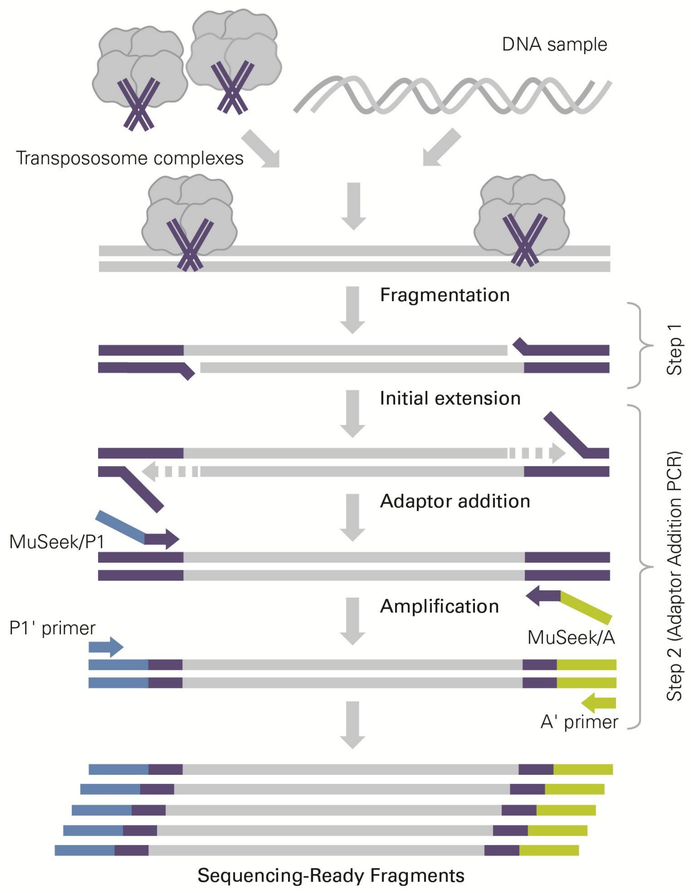
Methods

Escherichia coli str. K-12 substr. DH10B (Thermo Scientific), *Staphylococcus aureus* str. USA300_TCH959 and *Rhodopseudomonas palustris* str. CGA009 (LGC standards) genomic DNAs were used for library preparation and subsequent sequencing. Library preparation was performed using the Thermo Scientific MuSeek Library Preparation Kit for Ion Torrent™, which provides already preformed MuA transpososome and other reagents for fragmentation and PCR-assisted adaptor addition (AA-PCR). Size-selection was performed using E-Gel™ Agarose Gel Electrophoresis System on E-Gel SizeSelect™ 2% Agarose Gel (Life Technologies). DNA fragment size distribution after fragmentation, AA-PCR and size-selection was analyzed by Agilent 2100 Bioanalyzer instrument using the High Sensitivity DNA Kit (Agilent Technologies). Sequencing templates were prepared using Ion OneTouch™ 200 Template Kit v2 (Life Technologies). Libraries were sequenced on Ion Torrent Personal Genome Machine™ using Ion PGM™ 200 Sequencing Kit and Ion 316 Chips (Life Technologies).

Technology Overview

MuA transposase enzyme is a 75-kDa protein originating from the bacteriophage Mu. Under normal conditions, MuA transposase forms a homotetrameric complex with two 50 bp double-stranded transposon DNA that contain a specific MuA binding sequence. Following fragmentation, the ends of target DNA are tagged with transposon sequences. PGM-specific adaptors are added to the fragments in a subsequent adaptor-addition reaction. First, the 3'-ends of the fragmented target DNA are elongated over the 5-nt gaps that are generated during the transposition reaction and further extended into transposon sequences. In the initial cycles of consecutive PCR the first 16 nt of the adaptor 3'-ends hybridize to transposon sequences in tagged DNA fragments. In later PCR cycles, the fragments are amplified using a pair of external primers (see Fig. 1).

FIGURE 1. NGS library preparation workflow. MuA transposase complex simultaneously fragments input DNA and tags the fragments. In a subsequent PCR adaptor sequences are added.



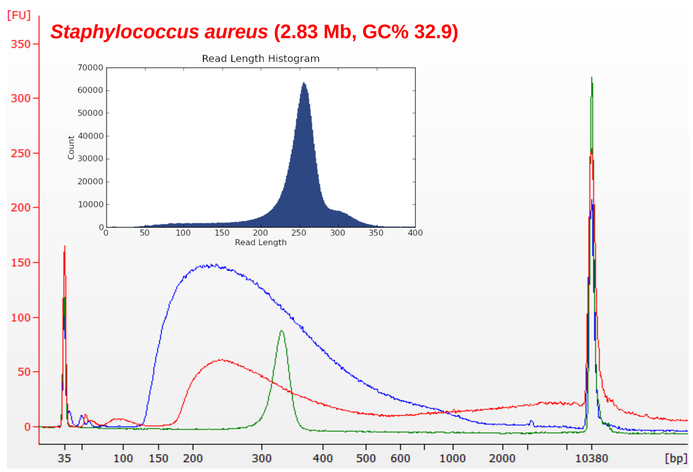
Results and discussion

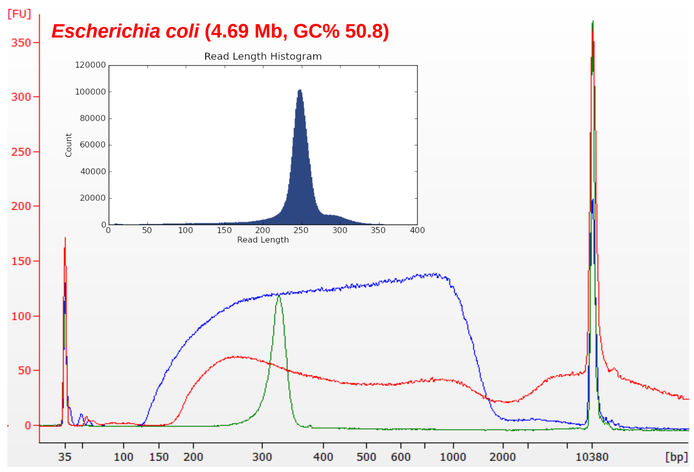
In order to analyze the yield and the size of resulting DNA fragments after individual steps in the MuA sample preparation workflow, samples of all three genomes were analyzed after gDNA fragmentation, AA-PCR and size-selection (see. Fig. 2). Results revealed that gDNA of *S. aureus* was slightly oversheared, causing an accumulation of shorter fragments after AA-PCR. Nevertheless, in all three cases the material recovered after the size selection of 330 bp fragments was enough for multiple reactions of sequencing template preparation.

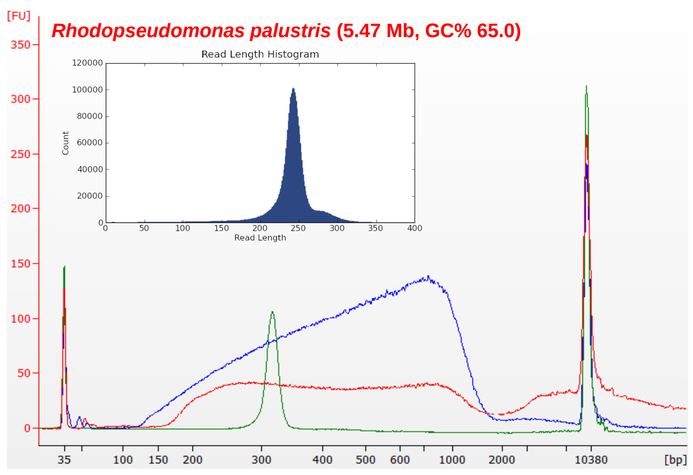
FIGURE 2. Size distribution of *S. aureus, E. coli* and *R. palustris* genomic DNA fragment libraries generated using MuA transpososome-based method:

1. Red – 2-fold diluted DNA fragments after fragmentation reaction;
2. Blue – 4-fold diluted DNA fragments after Adaptor Addition PCR;
3. Green – fragments after size-selection.

On the left top corner of each graph sequencing read length histograms are presented. Histograms were generated with Ion Torrent "CR" and "Beverly" quality filters off to avoid elimination of low quality reads and show the real length of sequenced MuA libraries.



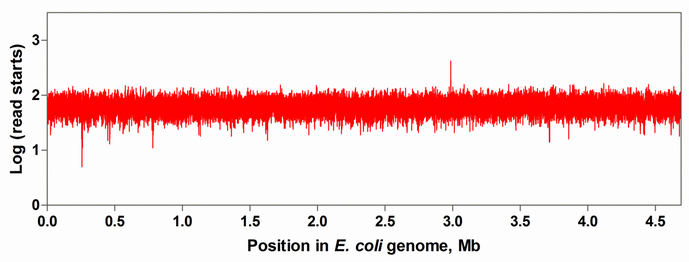




The average sequencing read length should coincide with the average library fragment length (although libraries contain adaptors, these sequences are automatically removed from sequencing data). Reducing the presence of short reads helps to maximize NGS throughput. In our case the average length of called reads (with quality filters off) were 245 nt (*S. aureus*), 244 nt (*E. coli*) and 239 nt (*R. palustris*), which perfectly match the size of 330 bp size-selected library where 93 nucleotides come from AA-PCR modified transposed ends of MuA transposon. Of note, the number of reads with unrecognized transposon-specific distal sequences was very low - in all cases <1.5 % of total reads (see sequencing read length histograms, Fig. 2).

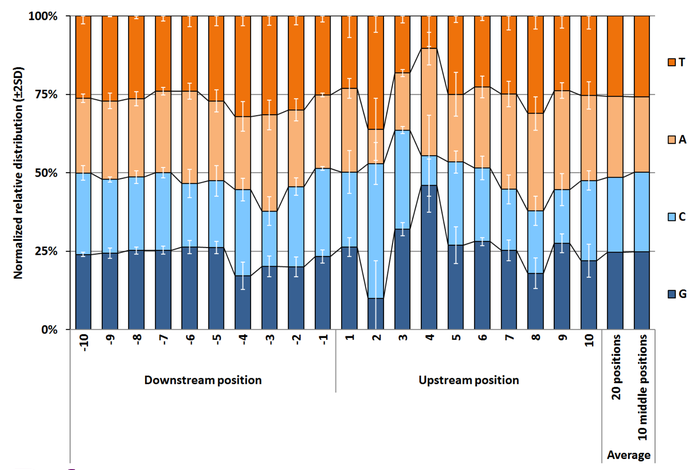
Genome coverage, which is defined as the average number of reads that align to known reference bases, is another very important sequencing characteristic. Coverage is primarily influenced by library quality. Sequencing data of libraries generated using low bias methods usually result in high coverage uniformity. Our *E. coli* sequencing coverage uniformity is 99.97% (see Fig. 3).

FIGURE 3. *E. coli* genome coverage graph. Graph represents *Escherichia coli* str. K-12 substr. DH10B genome coverage by showing logarithm of total sequencing read starts number in each 100 nt window. Average read starts – 66, average read length ~201 (quality filters ON), average coverage depth of reference ~130X (coverage at 1X = 100%, 10X = 99.993%, 20X = 99.983%, 50X 99.777%). Total number of bases aligned to the reference sequence (excludes library key, tag and 3' adapter sequences). - 608 Mb



Using sequencing data from all three genomes, the distribution of MuA-assisted insertions were analyzed to determine if the MuA specificity in a preformed transpososome is semi-random as previously reported (1). Small-scale analysis of MuA transposition events led to the hypothesis that the MuA transposition is influenced by a target which covers a region of up to 23–25 bp (2). Here, our target analysis covers 20 base pairs in total: 10 base pairs downstream from the MuA strand transfer position and 10 upstream. Analysis of more than 3×10^6 targets (see Fig. 4) revealed that MuA has nearly undetectable bias at all positions analyzed except 2 and 4 where some preference for pyrimidines (C/T) and purines (A/G), respectively, was observed. Regardless of this slight specificity of MuA insertions, we observed all 1024 (4^5) possible pentanucleotide targets in first five positions of the target. These results corroborate the ones reported earlier (3) and demonstrate that the MuA transpososome is well suited for sequencing of genomes with varying GC content.

FIGURE 4. MuA transposase target preference graph. Graph shows relative distribution of A, T, C and G nucleotides downstream and upstream from MuA cleavage/strand transfer site (error bars indicate 2SD values). Graph was generated using *S. aureus*, *E. coli* and *R. palustris* sequencing data. For analysis only sequences having original start and stop positions in particular genome were chosen, eliminating duplicates resulting from PCR. More than 3×10^6 targets were analyzed and at each position (A, T, C, G) distribution was normalized considering each genome GC content.



References

1. An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. Butterfield Y S N et. al. 2460-8 Nucl Acids Res, 2002, vol. 30, No. 11.
2. DNA transposition of Bacteriophage Mu. A quantitive analysis of target site selection in vitro. Haapa-Paananen S, Rita H and Savilahti H. 2843-51 J Biol Chem, 2002, vol. 277, No. 4.
3. Target site selection in transposition of phage Mu. Mizuuchi M and Mizuuchi K. 515-23 Cold Spring Harbor Symp. Quant. Biol. 1993, vol. 58.

Thermo SCIENTIFIC