A New Protocol for Next Gen Library Construction Increases Yield / Complexity and Simplifies Parallel Sample Handling

Adam N. Harris, Kyusung Park, Xiaoping Duan, Rhonda Meredith, Jennifer M. Kilzer, Eric White, Cisilya Duncan, Delia Ye, Steven Roman, Tanya Sokolsky, and Rob Bennett
Life Technologies, 5791 Van Allen Wy, Carlsbad, CA 92008, U.S.A.

ABSTRACT AND INTRODUCTION

With the advent of sample indexing schemes and as throughputs increase on second generation sequencers, the number of samples which can be combined and analyzed in a single run has risen dramatically. However, the practicality of generating multiple libraries simultaneously is only beginning to be addressed. For high throughput scenarios, automated systems have become tenable, but for those still prepping libraries by hand, the process remains tedious.

Fragment library generation for single- and paired-end reads is typically performed by a series of enzymatic steps following gDNA shearing by ultrasonication. These may include end-repair, A-tailing, and ligation, each followed by bead- or column-based purification. The clean-ups are necessary to prevent enzyme activity carryover from one step to another.

We have developed a method to take gDNA through shearing, end-repair, A-tailing, and ligation with only one intervening bead-based purification that doubles as a size selection step. The A-tailing is done by a thermophilic polymerase. Ligation is performed by adding ligase and adaptors after cooling the reaction such that little polymerase activity remains. The temperature is shifted back up to reactivate the polymerase to perform a nick translation step. A bead- or column-based purification is then performed after the ligation reaction. In many cases, subsequent library amplification is not needed, allowing high complexity libraries to be brought directly into the clonal amplification step.

With the removal of the extra clean-up steps and library amplification, the workflow is simplified, allowing faster processing of samples with less hands-on time. We also increase library yields by using a protective buffer during shearing and optimizing all enzymatic reactions. Our method is well suited for labs which are producing multiple libraries each week but have not scaled up to full automation.

MATERIALS AND METHODS

Library construction

DNA was sheared by ultrasonication on the Covaris® S2 instrument (4387833). Libraries designated as "SOLiD™ 4" were built using reagents from the SOLiD™ Fragment Library Construction Kit (4443473) and SOLiD™ Fragment Library Barcode kit 1-96 (4449637) except as noted. Libraries designated as "SOLiD™ 5500" were constructed with reagents from the 5500 SOLiD™ Fragment Library Core Kit (4464412) and 5500 SOLiD™ Fragment Library Barcode Adaptors 1-96 (4464404). Where indicated, size selection and purification steps were performed using Agencourt® AMPure® XP beads (Beckman Coulter, A63882). Amplifiable library yields were determined by using the SOLiD™ Library TaqMan® Quantitation Kit (4449639) on nick-translated material according to the users' manual. Size distribution was determined using an Agilent® 2100 Bioanalyzer™ with the Agilent® High Sensitivity DNA Kit (Agilent Technologies, G2938C and 5067-4626, respectively).

ePCR

Emulsion PCR was carried out using the SOLiD™ EZ Bead™ System (4448417) or manually with the SOLiD™ ePCR Kit V2 (4400834), Ultra-Turrax® Tube Drive from IKA® (4400335), and SOLiD™ Bead Enrichment Kit (4387894) as described in the Templated Bead Preparation Guide.

Sequencing

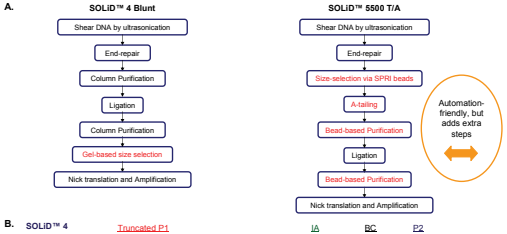
All sequencing was done on the SOLiD™ 4 System (4452773) according to the Instrument Operations Guide.

Data Analysis

Primary analysis was performed using the default settings of the SOLiD™ 4 Analyzer. Secondary analysis was performed using SOLiD™ BioScope™ 1.3. Start point and end-bias analysis were performed using custom scripts written in perl.

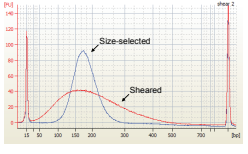
RESULTS

Figure 1. New Library Construction Workflow and Adaptor Sequence Changes



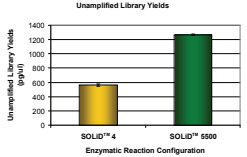
A. For both the new (SOLiD™ 5500) and existing (SOLiD™ 4) library prep methods, DNA is fragmented by ultrasonication and then enzymatically repaired to leave blunt, phosphorylated ends. In the new method, sequential binding to SPRI beads performs both a clean-up and size selection (see Fig. 2), replacing the gel-based step that appears later in the SOLiD™ 4 protocol. In the new protocol, an additional step which A-tails the inserts and a subsequent purification are included. Ligation is performed with T-overhang adaptors. Column purification steps are replaced with SPRI beads. **B.** SOLiD™ 5500 adaptors are T-tailed (arrows). This does not affect the final library sequence at the P1 junction, but adds a T/A base pair at the junction with the second adaptor, which includes Internal Adaptor (IA), Barcode (BC), and P2 sequences. Reverse read reagents for the SOLiD™ 5500 System are configured to read the F5 tag from this modified internal adaptor. The change to a full-length P1 adaptor allows for direct quantitation by Taqman® qPCR without the need for library amplification.

Figure 2. Bead-based Size Selection



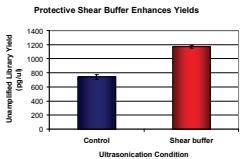
Bioanalyzer traces of *E. coli* DNA after shearing by ultrasonication (red) and subsequent end-repair and double-SPRI size selection (blue). High and low molecular weight material are effectively removed by differential binding conditions to magnetic SPRI beads.

Figure 3. Optimized Enzymatic Reactions



Libraries were created from common sheared and bead size-selected *E. coli* DNA which was then used in one of two library protocols – end-repair and blunt ligation typical of the SOLiD™ 4 protocol or the new protocol for the SOLiD™ 5500. Buffer changes and optimization of enzyme concentrations in the end-repair and ligation reactions in addition to ligation with T/A overhangs leads to an approximate doubling of unamplified SOLiD™ 5500 library yields as determined by Taqman® qPCR.

Figure 4. Protective Shear Buffer



A proprietary additive is included during ultrasonication to reduce damage to the DNA ends. After end-repair, A-tailing, and ligation, protected DNA ligates at an efficiency approximately 1.5-fold higher than DNA sheared in low TE buffer alone.

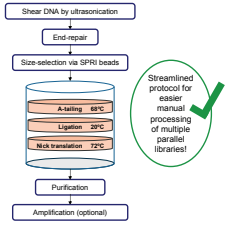
Table 1. Unamplified Library Yields

Input DNA Amount	Expected Unamplified Yield	Approx. Bead Yield after ePCR[2]
10 ng	1-2 fmol[1]	60M
500 ng	50-100 fmol[1]	3,000M
1 ug	100-150 fmol	5,000M
5 ug	225-275 fmol[1]	10,000M

[1] Yields back-calculated from 4 cycles of amplification, assuming 100% PCR efficiency.
[2] Based on typical yield from an EZ Bead™ E80 seeded with 25fmol template.

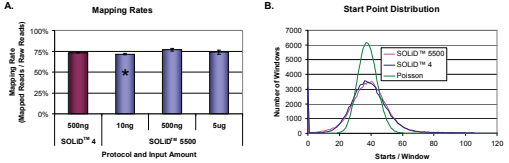
Whether amplification is required for a library depends on the size of the reference genome and the desired level of coverage. Estimated library yields and the corresponding number of templated beads which could be produced are given for inputs ranging from 10ng to 5ug.

Figure 5. Final Library Workflow



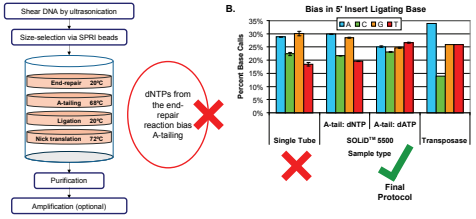
To reduce the number of purification steps and time required for parallel sample handling in the protocol, we improved the standard T/A ligation workflow by performing the A-tailing, ligation, and nick translation steps in a single tube (blue cylinder). A-tailing is carried out by a thermostable polymerase. The tube is brought to room temperature for the addition of ligase, adaptors, and dNTPs. The ligation is then performed at 20°C with minimal residual activity from the polymerase. After ligation, the polymerase is reactivated at high temperature to perform nick translation in the presence of dNTPs. For libraries which do not require amplification, there is significant time savings from including the nick translation reaction after ligation without an intervening purification step. The final purification can be performed with beads or a column.

Figure 6. Sequencing Performance



Libraries made using the new protocol were compared to those made with the existing protocol by performing ePCR and sequencing-by-ligation on the SOLiD™ 4 platform. **A.** Raw, unfiltered reads were mapped against the *E. coli* reference genome, and mapping rates were determined to be comparable to those generated using the SOLiD™ 4 procedure (n=3; *n=2 after exclusion of outlier with significant adaptor dimer carryover). **B.** Each of 3.6M randomly chosen mapped reads from the 500ng SOLiD™ 4 and SOLiD™ 5500 samples was assigned to a 100bp window on either strand of the reference genome according to its start point. The number of windows with a given number of starts / window were calculated and compared to those expected by chance ("Poisson"). Distributions were very similar for both library types.

Figure 7. A Single-tube Workflow Causes Significant Nucleotide End-Bias



A. We also tested a procedure which combines all enzymatic reactions before amplification in a single tube. End-repair enzymes are heat inactivated during the temperature ramp-up before the A-tailing procedure. **B.** However, sequencing reveals substantial bias in the nucleotide composition at the insert ends (Single Tube, left set of bars). The *E. coli* reference has ~50% GC content. This bias is caused by the presence of dNTPs during the A-tailing reaction and can be recapitulated when the A-tailing step in the SOLiD™ 5500 protocol (Fig. 5) is performed with all four dNTPs (center left) instead of dATP only (center right). For comparison, the bias from the single tube protocol approaches the level seen from transposase-based library prep (right, adapted from Adey et al. 2010 Supplementary Figure 1, *E. coli* data).

CONCLUSIONS

We improved library preparation yields and automatability by switching to a T/A-based ligation method, optimizing the enzymatic reactions, and changing from gel-based to bead-based size selection. We also incorporated a protective buffer during ultrasonication to prevent damage to the insert ends. To simplify the workflow, we combined the A-tailing, ligation, and nick translation reactions into a single tube by using a thermostable polymerase and alternating between high and low temperature incubations. We also attempted to remove the clean-up step between end-repair and A-tailing. We found that doing so introduced substantial bias at the A-tailing step because of the presence of dNTPs.

REFERENCES

Adey et al. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology* 11:R119.

ACKNOWLEDGEMENTS

The authors thank George Marnellos and Tony Xu for bioinformatics support and Tom Chen, Rachel Fish, John Bodeau, and Mark Landers for comments and suggestions during product development.

TRADEMARKS/LICENSING

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.
© 2011 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. TaqMan® is a registered trademark of Roche Molecular Systems, Inc. Agencourt® and AMPure® are registered trademarks of Beckman Coulter, Inc. Ultra-Turrax® is a registered trademark of IKA Werke GMBH & Co. Agilent® and Bioanalyzer™ are trademarks of Agilent Technologies Inc.