

Simplified DNA Library Preparation: Simultaneous DNA Fragmentation and Adaptor Tagging by *In Vitro* Transposition

Nicholas Caruccio, Haiying Grunenwald, Dilara Begum, Scott Kuersten, and Mark Maffitt
EPICENTRE Biotechnologies, Madison, WI

Introduction

DNA library preparation is a common entry point for next-generation sequencing. Current methods generally consist of distinct DNA fragmentation, end-polishing, and adaptor-ligation steps; often with significant sample loss and hands-on time.

EPICENTRE Biotechnologies has developed Nextera™ technology, a streamlined, scalable, and efficient method that generates adaptor-tagged libraries for multiple next-generation sequencing platforms, whole-genome amplification, and other applications.

In vitro transposition with Nextera Transposomes™ fragments and covalently tags the target DNA in a single step. Di-tagged sequencing libraries compatible with the Roche 454 and Illumina platforms can be enriched and bar-coded using limited-cycle PCR.

Deep sequencing of the transposon-fragmented libraries produces accuracy, coverage, and bias comparable to control libraries produced by physical shearing methods.

Methods and Results

DNA Fragmentation and Tagging with Free Transposon Ends

During *in vitro* transposition with hyperactive Transposomes, strand-transfer occurs via random, staggered, dsDNA breaks in the target DNA and covalent attachment of the 3' end of the transferred (top) transposon strand to the 5' end of the target DNA. When free transposon ends are used in the reaction, the target DNA is fragmented and the transferred strand of the transposon end oligonucleotide is covalently attached to the 5' end of the fragment (Fig. 1A). The size distribution of the fragments can be controlled by changing the amount of Transposomes (data not shown) and reaction buffer conditions (Fig. 1B and 1C). Genomic DNA can be fragmented to <1 kb in 5 minutes with enhanced reaction conditions.

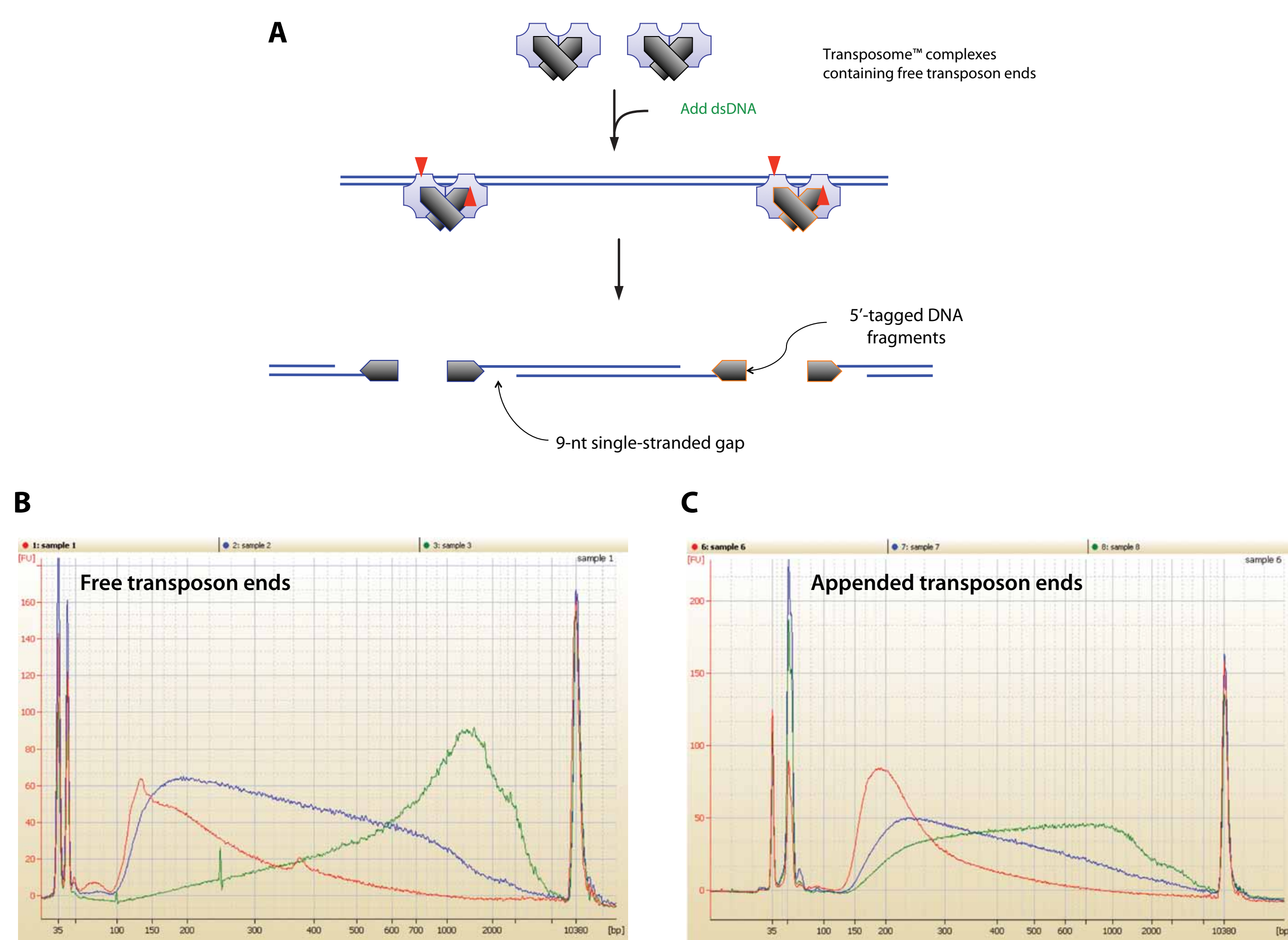


Figure 1. Simultaneous DNA fragmentation and tagging by *in vitro* cut-and-paste transposition. A) When free transposon ends are used in the insertion reaction, the target DNA is cleaved and tagged at the 5' end with the transposon sequence. The resulting fragments have single-stranded gaps. B, C) Control of molecular weight distribution by altering buffer conditions. *E. coli* genomic DNA was tagged with Transposomes containing free transposon ends (B) or transposon ends appended with sequencing tags (C) in LMW (red trace), HMW (blue trace), or TA (green trace) Buffers, for 5 minutes at 55°C. The resulting tagged DNA was examined using a BioAnalyzer (Agilent) with a High-Sensitivity Chip.

Production of Bar-Coded Libraries Compatible with Roche and Illumina Platforms

Bar-coded adaptors compatible with Roche 454 or Illumina sequencing platforms were added by PCR. The primer design uses limited-cycle PCR to enrich di-tagged fragments. PCR products are compatible with emulsion PCR (Roche 454) or bridge PCR (Illumina).

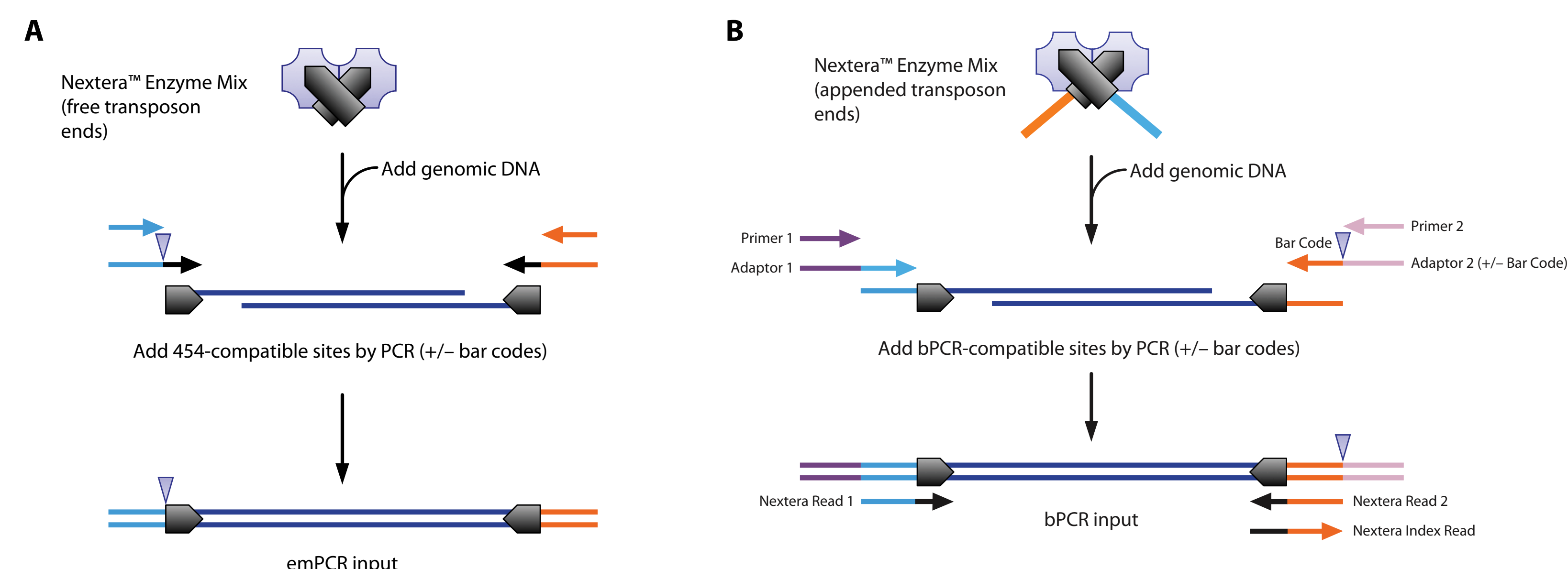


Figure 2. Addition of bar-coded Roche 454-compatible and Illumina-compatible adaptors for multiplexed sequencing. A) Roche 454-compatible libraries. Target DNA is fragmented and tagged with the core Transposomes containing MEs. Limited-cycle PCR with a four-primer reaction adds Roche 454-compatible adaptor sequences (blue and orange). Optional bar coding (triangle) is added between the upstream emPCR adaptor (blue) and the transposon ME (gray). B) Illumina-compatible libraries. Target DNA is fragmented and tagged with the core Transposomes containing MEs appended with sequencing primer sites (blue and orange). Limited-cycle PCR with a four-primer reaction adds bPCR-compatible adaptors (purple and pink) to the core sequencing library. Optional bar coding (triangle) is added between the downstream bPCR adaptor (pink) and the core sequencing library (orange). Alternative sequencing primers are required for the Illumina-compatible libraries; Read 1 (blue/gray arrow), Read 2 (orange/gray arrow), Index Read (gray/orange arrow).

Contact Information:

Nick Caruccio • nick.caruccio@epibio.com • 1-800-284-8474, x6162

Effect of Clean-up Method on MW Distribution of Sequencing Libraries

Sample DNA was fragmented and 5'-end-tagged with Transposomes. Roche 454-compatible adaptor sequences were added and di-tagged fragments were enriched by limited-cycle PCR.

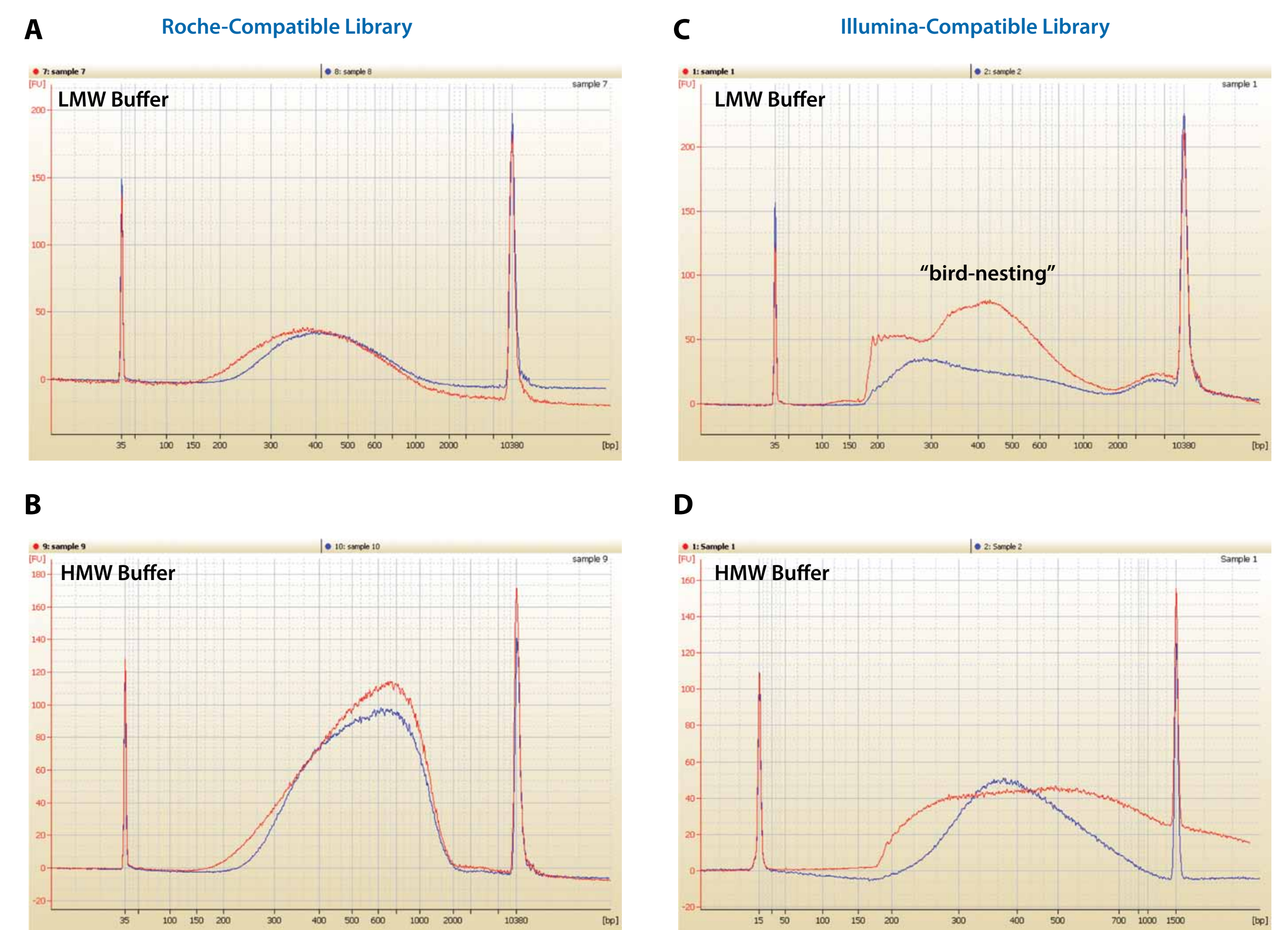


Figure 3. Comparison of MW Distribution of sequencing libraries purified by Zymo and AMPure methods. Roche-compatible (A and B) or Illumina-compatible (C and D) sequencing libraries were prepared as described in Fig. 2 using LMW buffer (A and C) or HMW buffer (B and D). PCR products were purified using either Zymo (red trace) or AMPure beads (blue trace) per the manufacturer's instructions. The resulting sequencing libraries were examined using a BioAnalyzer (Agilent). Figure 3C shows an example of "bird nesting" where noncovalent concatemers result in higher apparent molecular weight under native conditions.

Deep Sequencing of 454-Compatible Libraries Produced by *In Vitro* Transposition

Sample DNA was fragmented and 5'-end-tagged with Transposomes. Roche 454-compatible adaptor sequences were added and di-tagged fragments were enriched by limited-cycle PCR. The recovered library was used directly as input for Roche 454 FLX Titanium emulsion PCR. Deep sequencing of the transposon-fragmented libraries produced coverage that was comparable to libraries prepared by physical shearing.

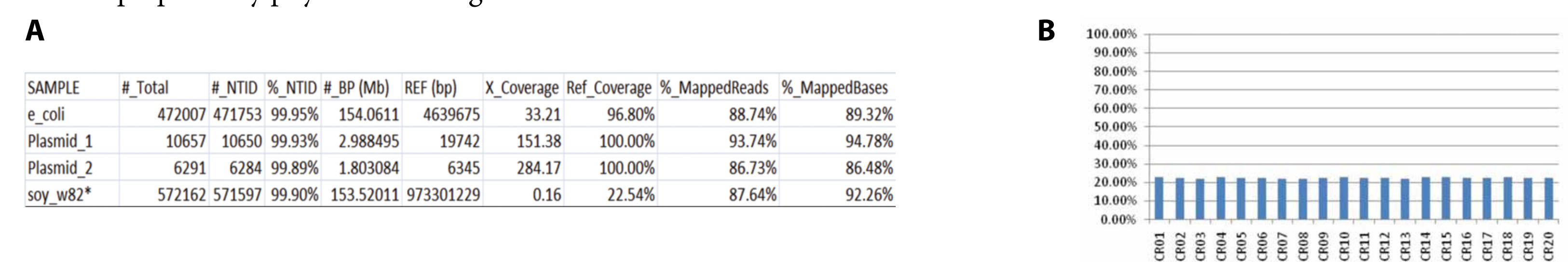


Figure 4. Deep sequencing of Transposome-generated sequencing libraries (Roche 454). Libraries prepared as outlined in Fig. 2 from *E. coli*, plasmid, or soybean genomic DNA were sequenced using Roche FLX-Titanium chemistry. A) Summary table of sequencing yield, coverage, and mapping data. B) Relative coverage of individual soybean chromosomes shows even coverage across all 20 chromosomes.

Deep Sequencing of Illumina-Compatible Libraries Produced by *In Vitro* Transposition

E. coli (B REL606) genomic DNA was fragmented and 5'-end-tagged with Transposomes appended with sequencing adaptors. Illumina-compatible adaptors were added and di-tagged fragments were enriched by limited-cycle PCR. The produced library was used as input for cluster formation without size selection. Deep sequencing of the transposon-fragmented libraries was performed with alternate sequencing primers (Fig. 2) and mapped to the reference sequence. The high coverage of LacZ and nohB promoters are reads from unrelated libraries sequenced in the same channel that mapped to the reference. The GC bias of the Nextera library is comparable to that observed with physical shearing, suggesting that downstream steps (most likely, bPCR) is the primary source of GC bias.

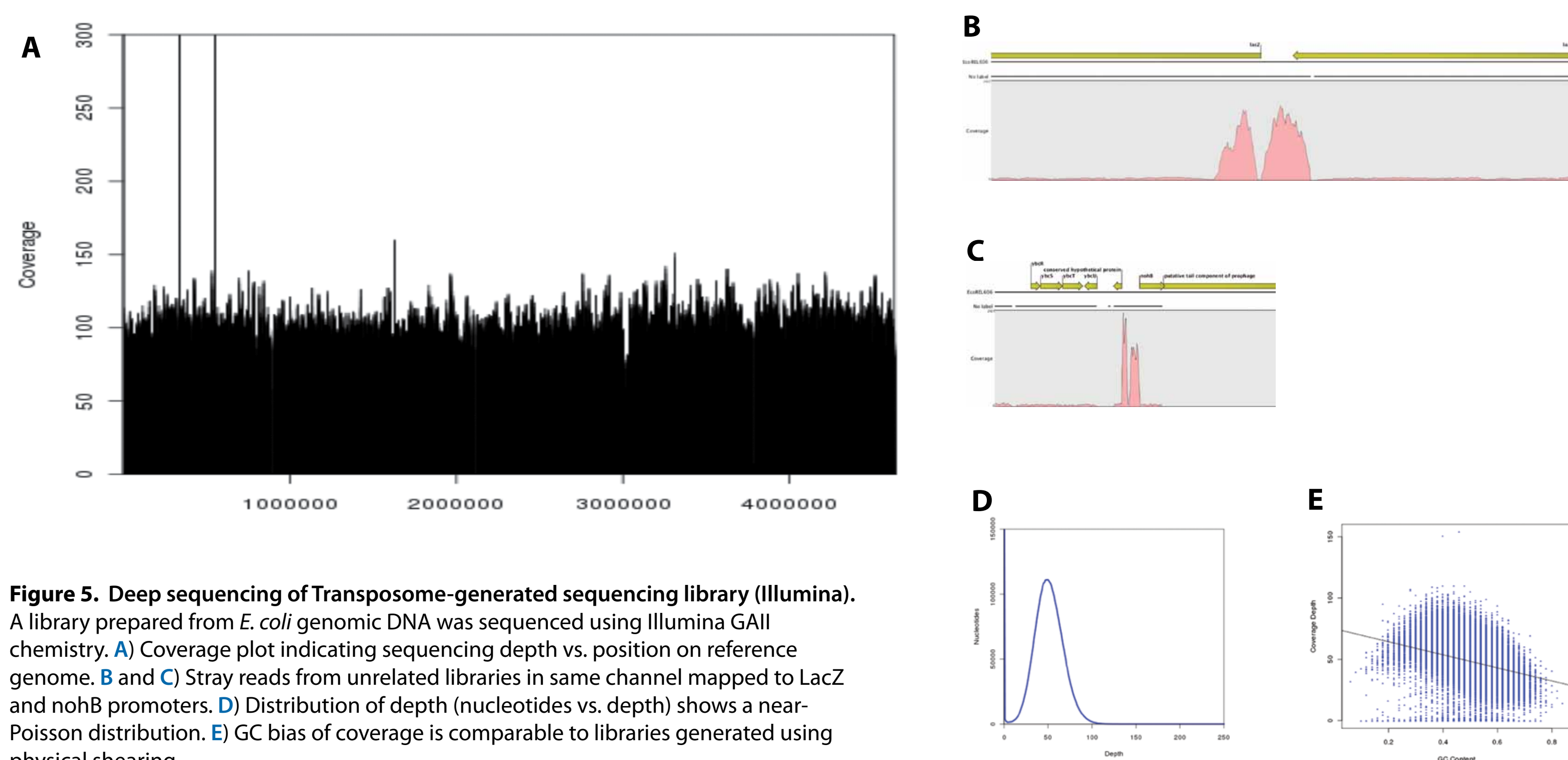


Figure 5. Deep sequencing of Transposome-generated sequencing library (Illumina). A library prepared from *E. coli* genomic DNA was sequenced using Illumina GAI chemistry. A) Coverage plot indicating sequencing depth vs. position on reference genome. B and C) Stray reads from unrelated libraries in same channel mapped to LacZ and nohB promoters. D) Distribution of depth (nucleotides vs. depth) shows a near-Poisson distribution. E) GC bias of coverage is comparable to libraries generated using physical shearing.