Streamlined Express DNA Library Preparation Methods Meet Requirements of High-throughput Library Construction

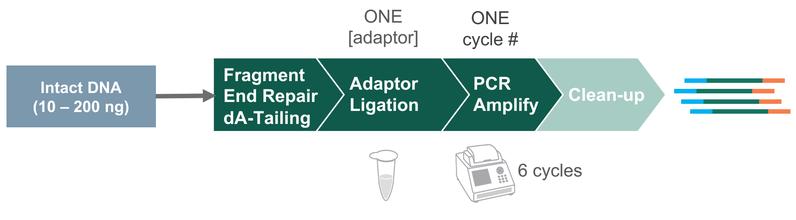
Karen R. McKay, Chen Song, Kaylinnette Pinet, Matthew Campbell, Bradley W. Langhorst, V K Chaithanya Ponnaluri, Pingfang Liu, and Keerthana Krishnan

New England Biolabs, Inc. 240 County Road, Ipswich, MA 01938, USA.

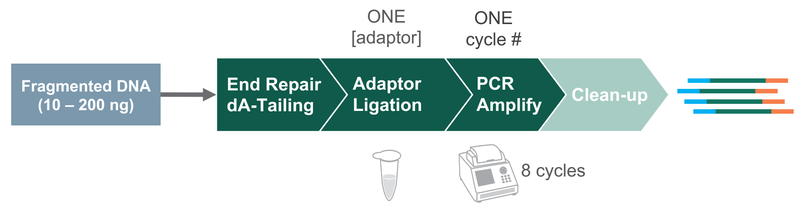
INTRODUCTION

As next generation sequencing technologies improve and capacities expand, library construction has increasingly become the bottleneck for fast data turnaround and overall cost. There is a growing need for faster, easier, and automatable protocols that perform reliably and do not compromise the quality of the libraries produced. We have developed two new NEBNext UltraExpress™ workflows with mechanical or enzymatic shearing approaches, allowing for a single tube, standardized protocol for fast library preparation. The streamlined, user-friendly workflows reduce hands-on time by combining reaction steps, removing clean-ups, and simplifying the protocols with a single adaptor concentration and one PCR cycling condition. They allow many DNA samples with varying sources and input amount (10 ng to 200 ng) to be processed in one setting without normalization or adjusting reaction conditions. We have applied these two robust workflows to generate high quality sequencing data. We expect these library preparation methods will serve a myriad of NGS applications, including clinical and diagnostic settings.

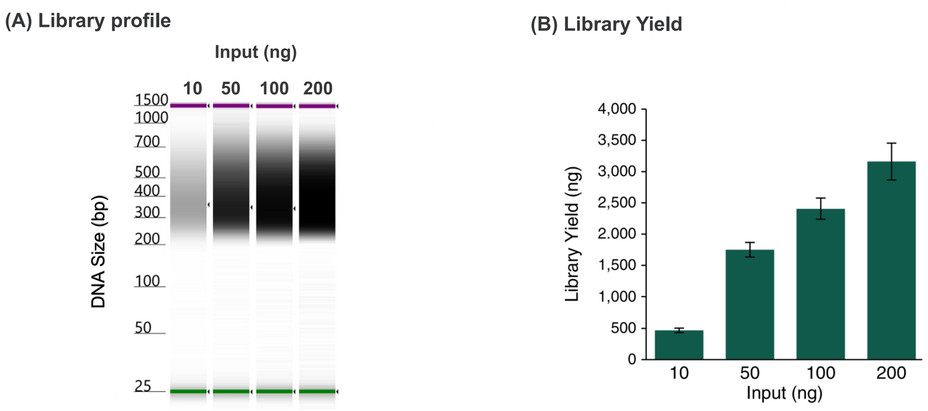
NEBNext UltraExpress FS DNA library preparation



NEBNext UltraExpress DNA library preparation

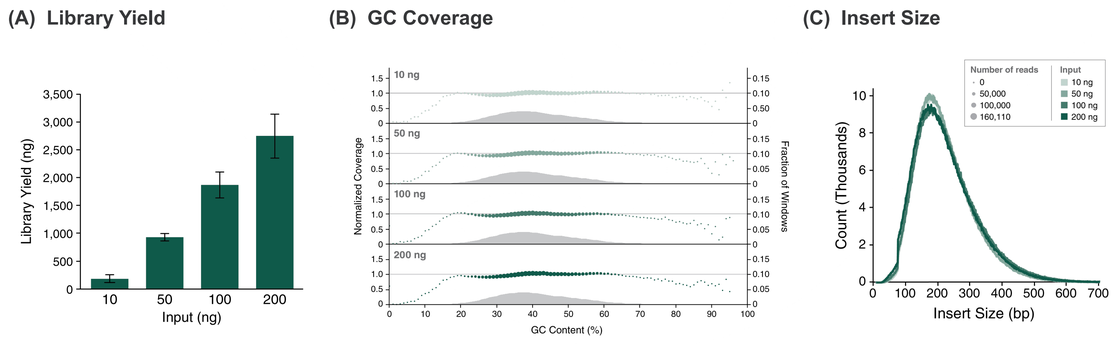


NEBNext UltraExpress FS DNA Library Prep Kit provides consistent fragmentation and robust library yields over a wide input range



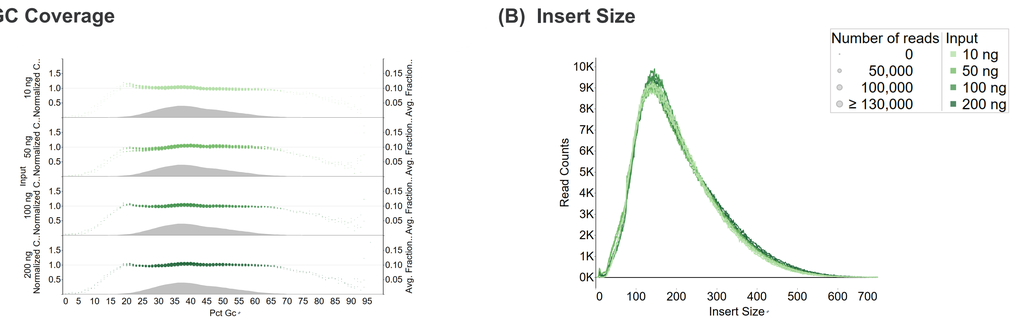
(A) Libraries made using 10 ng to 200 ng a 9:1 Human NA19240 genomic DNA (Coriell Institute for Medical Research) and *Escherichia coli* gDNA (Lofstrand Labs Limited) mixed sample, using the NEBNext UltraExpress FS DNA single-protocol workflow (e.g., same adaptor amount and 6 PCR cycles for all input amounts) showed uniform library profiles. (B) Libraries prepared in triplicate from 10, 50, 100 and 200 ng of a 9:1 Human NA19240 genomic DNA and *Escherichia coli* gDNA mixed sample, using the NEBNext UltraExpress FS DNA single-protocol workflow. Yields exceeded the minimum requirement (40 ng) for a single Illumina® NovaSeq® 6000 run to achieve whole genome sequencing with at least 30X coverage.

NEBNext UltraExpress DNA Library Prep Kit produces robust library yields and superior data quality



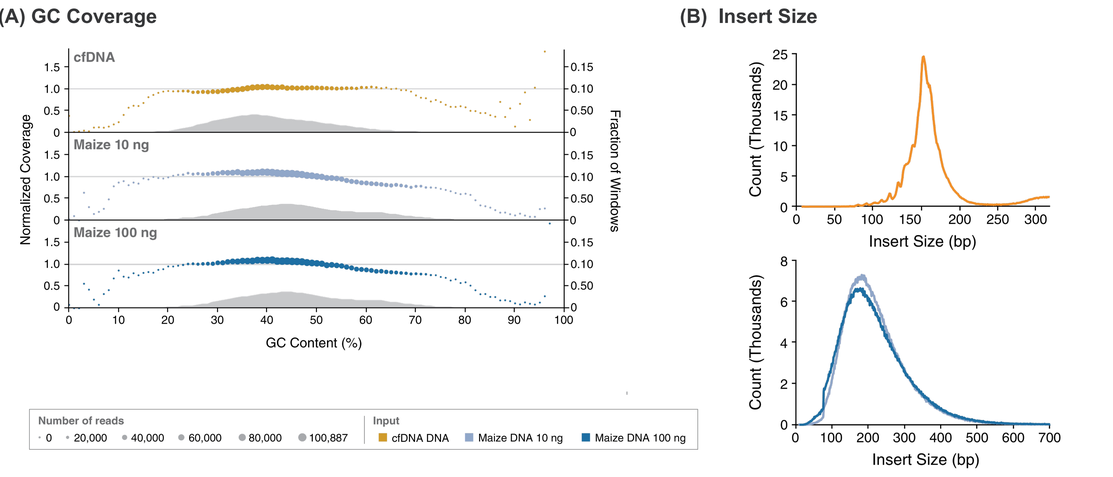
Libraries were prepared from 10, 50, 100 or 200 ng of Human NA19240 genomic DNA (Coriell Institute for Medical Research) using the same adaptor amount and 8 PCR cycles. (A) Yields exceeded the minimum requirement (40 ng) for a single Ilumina NovaSeq 6000 run to achieve whole genome sequencing with at least 30X coverage. Libraries were pooled and sequenced on an Illumina MiSeq (2 x 75 bases). Data showed consistent GC coverage (B) and insert size (C). 2 million paired-end reads from each library were sampled (seqtk v1.3), adapter-trimmed (seqprep v0.1) and mapped to the GRCh38 reference (bowtie2 v2.4.5), and GC coverage information was calculated using Picard's CollectGCBiasMetrics (v 1.56.0). In (B), the horizontal grey line indicates the expected normalized coverage of 1.0, and the dots in shades of green represent read numbers at each GC%. The grey area plot is a histogram representing the distribution of GC content in 100 bp windows of the reference genome.

NEBNext UltraExpress FS DNA Library Prep Kit produces libraries with uniform GC coverage and insert size from a range of input amounts



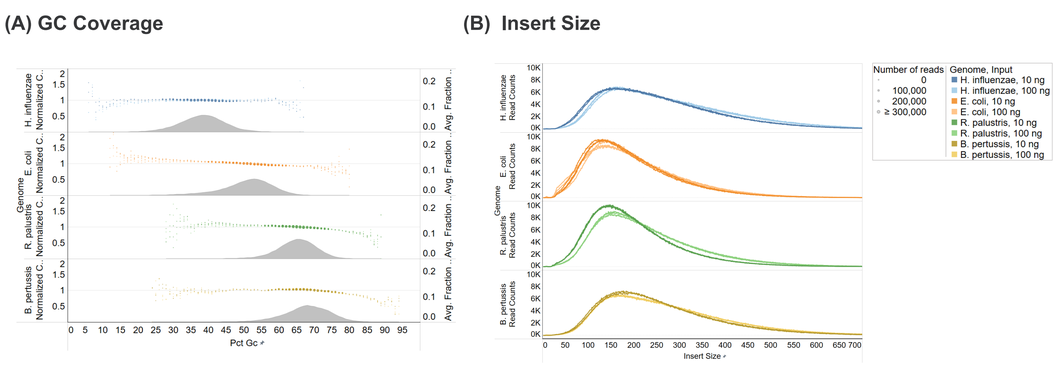
Libraries were prepared from 10, 50, 100 and 200 ng of a 9:1 Human NA19240 genomic DNA (Coriell Institute for Medical Research) and *Escherichia coli* gDNA (Lofstrand Labs Limited) mixed sample, using the NEBNext UltraExpress FS DNA single-protocol workflow (e.g., same adaptor amount and 6 PCR cycles for all input amounts). Libraries were pooled and sequenced on an Illumina NextSeq® 500/550 (2 x 75 bases). Data showed consistent (A) GC coverage and (B) insert size. 2 million paired-end reads from each library were sampled (seqtk v1.0), adaptor-trimmed (seqprep v0.1) and mapped to a composite reference containing *GRCh38* and *E. coli* MG1655 contigs (bowtie2 v2.5.0). GC coverage and insert size distributions were calculated using Picard's CollectGCBiasMetrics and Picard CollectInsertSizeMetrics (v1.56.0); Picard CollectGCBiasMetrics. (1.56) was run on human autosomes only due to the even copy number assumption of the tool. In (A), the horizontal grey line indicates the expected normalized coverage of 1.0, and the dots in shades of green represent read numbers at each GC%. The grey area plot is a histogram representing the distribution of GC content in 100 bp windows of the reference genome.

NEBNext UltraExpress DNA Library Prep Kit produces even GC coverage and insert size for a range of sample types



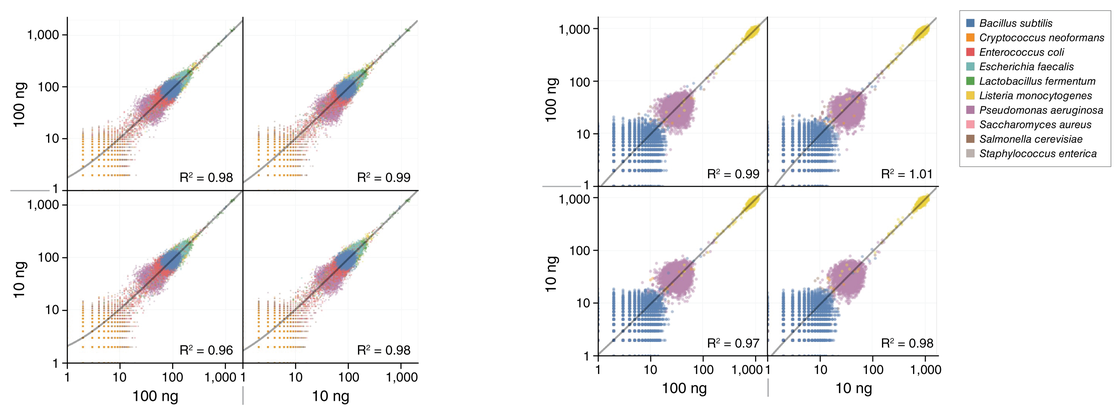
Libraries were prepared using a single protocol from cell-free DNA (cfDNA, 12 ng) without shearing and Maize DNA (10 ng and 100 ng) sheared to 200bp (Covaris ME220). Libraries were pooled and sequenced on an Illumina MiSeq (2 x 75 bases). 1.4 million paired-end reads from each library were sampled (seqtk v1.3), adapter-trimmed (seqprep v0.1), and aligned to either the GRCh38 reference (human) or Zea mays reference genome (maize) (bowtie2 v2.4.5). (A) All sample types showed representative GC coverage. High and low input Maize DNA generated consistent GC coverage. The horizontal grey line indicates the expected normalized coverage of 1.0, and the colored dots represent read numbers at each GC%. The grey area plot is a histogram representing the distribution of GC content in 100 bp windows of the reference genome. (B) cfDNA insert size had the characteristic fragmentation pattern with 167 bp peak size and periodicity feature of nucleosome-bound DNA in shorter fragments. 10 ng and 100 ng Maize DNA showed consistent insert size, as expected with the shearing protocol used.

NEBNext UltraExpress FS DNA Library Prep Kit produces representative GC coverage and insert size for a range sample types



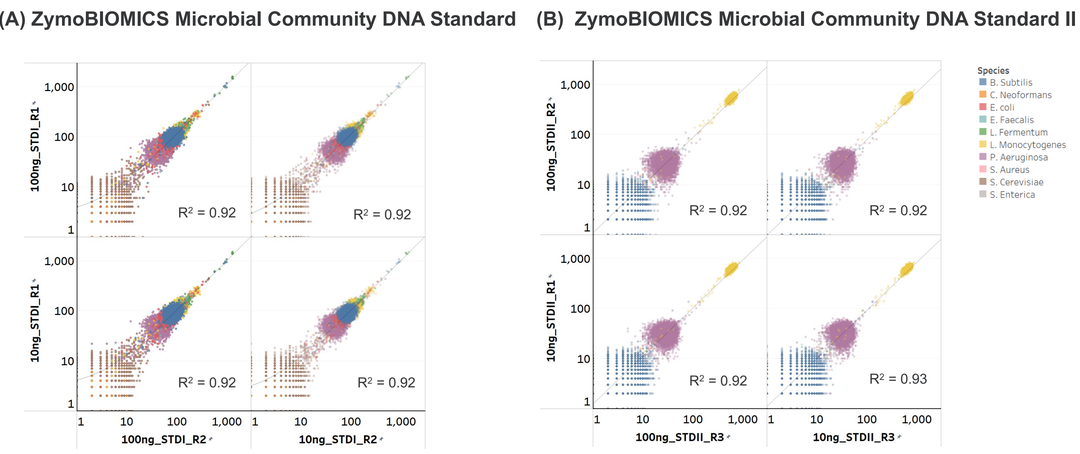
Libraries were prepared using the NEBNext UltraExpress FS DNA protocol for 10 ng and 100 ng of genomic DNA from *Haemophilus influenzae*, *Escherichia coli*, *Rhodopseudomonas palustris* and *Bordetella pertussis*. Data shown (A) representative GC coverage and (B) insert size peaks across samples with genome GC contents of 38%-68% GC. Libraries were pooled and sequenced on an Illumina NextSeq 500/550 (2 x 75 bases). 2 million paired-end reads from each library were sampled (seqtk v1.0), adaptor-trimmed (seqprep v0.1), and aligned to their respective reference genomes (bowtie2 v.2.5.0). GC coverage and insert size distributions were calculated using Picard's CollectGCBiasMetrics and Picard CollectInsertSizeMetrics (v1.56.0). The horizontal grey line indicates the expected normalized coverage of 1.0, and the colored dots represent read numbers at each GC%. The grey area plot is a histogram representing the distribution of GC content in 100 bp windows of the reference genome for the 10 and 100 ng inputs.

NEBNext UltraExpress DNA Library Prep Kit provides robust library complexity



Libraries were prepared using a single protocol from (A) 10 ng and 100 ng of the ZymoBIOMICS Microbial Community DNA Standard (Zymo Research, Catalog #D6306), and (B) 10 ng and 100 ng of the ZymoBIOMICS Microbial Community DNA Standard II (Log Distribution) (Zymo Research, Catalog #D6311). Libraries were pooled and sequenced on an Illumina MiSeq (2 x 75 bases). 1.4 million paired-end reads from each library were sampled (seqtk v1.3), adapter-trimmed (seqprep v0.1) and aligned to a composite reference genome (bowtie2 v2.4.5). 1,000 bp windows of constituent genomes were counted (bedtools 2.30.0) and compared across replicates and input levels. High correlation was observed between replicates and between inputs.

NEBNext UltraExpress FS DNA Library Prep Kit provides robust library complexity



Libraries were prepared using the NEBNext UltraExpress FS DNA protocol from (A) 10 ng and 100 ng of the ZymoBIOMICS® Microbial Community DNA Standard (Zymo Research®, Catalog #D6306), and (B) 10 ng and 100 ng of the ZymoBIOMICS Microbial Community DNA Standard II (Log Distribution) (Zymo Research, Catalog #D6311). Libraries were pooled and sequenced on an Illumina MiSeq® (2 x 75 bases). 750,000 paired-end reads from each library were sampled (seqtk v1.3), adaptor-trimmed (seqprep v0.1) and aligned to a composite reference genome (bowtie2 v2.4.5). 1,000 bp windows of constituent genomes were counted (bedtools 2.30.0) and compared across replicates and input levels. High correlation was observed between replicates and between inputs.

CONCLUSIONS

- Shorter hands-on time
- Fewer steps & consumables
- Automation friendly volumes
- Flexibility with samples types & protocols

NEBNext UltraExpress DNA and UltraExpress FS DNA library preparation workflows:

- Sample in library out in < 2 hours
- Single tube library prep reduces consumables
- One protocol fits wide input range and diverse sample types
- Produce high yields and quality sequencing metrics
- Pipetting/reaction volume automation friendly

Authors would like to acknowledge the technical assistance provided by Dora Posfai, Kristen Augulewicz, and Harry Bell at the New England Biolabs' Sequencing Core Facility.