TruSeq® Synthetic Long-Read DNA Library Prep Kit for Genome Assembly

Combining synthetic long-read technology with simple analysis to deliver accurate genomes.

Highlights

- **Highly Accurate Assembly**
  Assembling shorter reads into long fragments empowers *de novo* sequencing and genome finishing applications
- **Inclusive Library Preparation**
  Master-mixed reagents and primed 384-well plates increase protocol efficiency by minimizing manual pipetting steps
- **Instrument Flexibility**
  Reads are generated on a single instrument and long reads are constructed without requiring additional equipment
- **Integrated Analysis and Assembly**
  Inclusive solution spans the sequencing workflow, from library preparation to one-click assembly of long fragments

Figure 1: TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits

The TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits are designed for preparing DNA libraries to generate synthetically long reads. The TruSeq Long-Read Assembly App constructs long fragments from shorter reads for accurate genome assembly and genome finishing.

Introduction

Long sequence reads can be used to complement shorter, paired-end reads for genomic analysis. These long fragments facilitate alignment and improve the accuracy of genome assembly by providing insight into traditionally challenging regions, such as repetitive content. This capability is especially useful for assembling complex, polyploid plant genomes or for metagenomics studies, such as characterizing complex environmental samples.

The TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits (Figure 1) are designed for preparing DNA libraries that generate highly accurate sequencing reads. The TruSeq Long-Read Assembly App in the BaseSpace® environment constructs long fragments from the shorter reads, enabling more accurate assemblies than conventional methods. This method combines widely adopted TruSeq technology with user-friendly software to deliver highly accurate data, empowering *de novo* assembly and genome finishing applications.

Unique Library Preparation Chemistry

The TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits combine TruSeq and Nextera® chemistries with synthetic long-read technology to prepare DNA libraries for sequencing (Figure 2). The TruSeq Synthetic Long-Read DNA Barcode Kit includes 384 indexes for labeling the samples in each well. After sequencing, these indexes are used to construct synthetic long reads accurately. It also includes master-mixed reagents and pre-plated index primers in 384-well plates. This design minimizes manual pipetting steps so researchers can perform barcoding by PCR through simple centrifugation. The optional TruSeq Synthetic Long-Read DNA Accessory Kit contains alignment rings for arranging and centrifuging the plates.

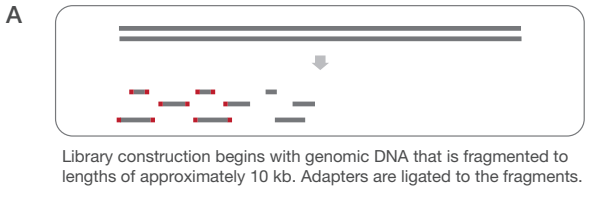
Library preparation begins by fragmenting 500 ng of genomic DNA to approximately 10 kb. Adapters are then ligated and 10 kb fragments are size selected (Figure 2A). A single Nextera “tagmentation” reaction simultaneously fragments and tags the DNA in each well (Figure 2B). PCR amplification then adds unique indexes to the samples. The fragments from all 384 wells are pooled, purified, and size selected (Figure 2C). Library preparation can be completed in 3 days, with 6 hours of hands-on time. Libraries are sequenced (Figure 2D), and the TruSeq Long-Read Assembly App uses the short reads to construct long fragments.

Highly Accurate Genome Assembly

The TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits use proven TruSeq technology[1] to prepare libraries for sequencing. The TruSeq Long-Read Assembly App constructs long fragments from the resulting sequence reads. TruSeq Synthetic Long-Read technology eliminates reliance on conventional methods that require greater sequencing depth and investment to achieve similar accuracy.

The TruSeq method constructs contiguous fragments (contigs) approximately 6–10 kb in length, which can be useful for differentiating between highly similar repetitive regions. These long contigs increase genomic coverage, supporting *de novo* genome assembly (Tables 1–2)[2]. The size-selected long fragments contain specific tags—or end markers—on both ends so that the number of fully assembled long reads can be viewed using the TruSeq Long-Read Assembly App (Figures 3–4). TruSeq Synthetic Long-Read technology increases the coverage of challenging regions without compromising accuracy, providing researchers with confidence in the resulting assemblies.

Figure 2: TruSeq Synthetic Long-Read DNA Library Preparation Workflow



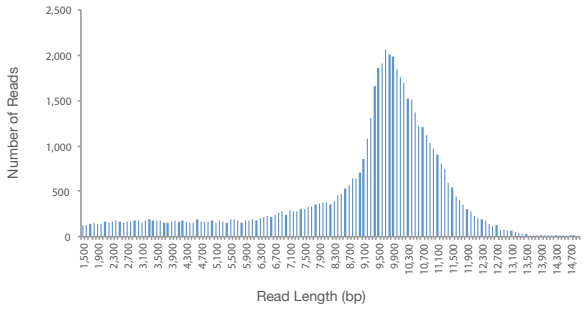
The TruSeq Synthetic Long-Read DNA Library Prep Kit prepares DNA for sequencing. The TruSeq Long-Read Assembly App assembles the sequencing reads into long fragments.

Table 1: *Caenorhabditis elegans* Genome Assembly

Parameter	Result
GC content of long reads	35.6%
Long reads aligned to reference genome	99.9%
Misassembled contigs	1.0%
N50 Length	9,125 bp
Bases fully assembled into long reads	1,182 Mb

GC content was calculated using QUAST[3]. The percentage of long reads aligned to the reference genome was determined using MUMmer[4]. Misassembled contigs were determined by calculating the percentage of total long reads that aligned with misassembly breakpoints as defined by QUAST.

Figure 3: Size Distribution of End-Marked Long Reads from the *C. elegans* Genome



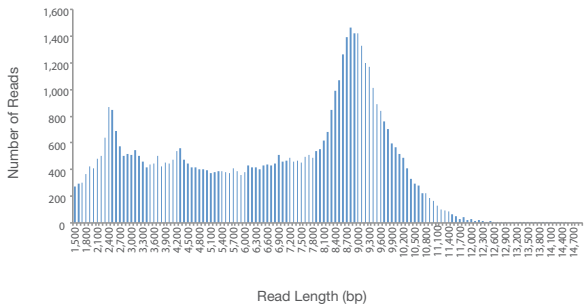
DNA libraries from *C. elegans* were prepared with the TruSeq Synthetic Long-Read DNA Library Prep Kit and long reads were assembled using the TruSeq Long-Read Assembly App. This output from the TruSeq Long-Read Assembly App shows the number of fully assembled long reads, arranged according to read length. The fragments shown denote the subset of reads that are labeled with markers on both ends and assembled from one end to another. The median read length is approximately 8–10 kb.

Table 2: *Oryza sativa* Genome Assembly

Parameter	Result
GC content of long reads	40.5%
Long reads aligned to reference genome	99.8%
Misassembled contigs	2.5%
N50 Length	7,095 bp
Bases fully assembled into long reads	1,088 Mb

GC content was calculated using QUAST. The percentage of long reads aligned to the reference genome was determined using MUMmer. Misassembled contigs were determined by calculating the percentage of total long reads that aligned with misassembly breakpoints as defined by QUAST.

Figure 4: Size Distribution of End-Marked Long Reads from the *O. sativa* Genome



DNA libraries from *O. sativa* Nipponbare were prepared with the TruSeq Synthetic Long-Read DNA Library Prep Kit and long reads were generated using the TruSeq Long-Read Assembly App. This output from the TruSeq Long-Read Assembly App shows the number of fully assembled long reads, arranged according to read length. The fragments shown denote the subset of reads that are labeled with markers on both ends and assembled from one end to another. The median read length is approximately 8–10 kb.

Instrument Flexibility

TruSeq Synthetic Long-Read DNA libraries can be sequenced on Illumina HiSeq® 2500 or HiSeq 2000 Systems. The TruSeq method enables researchers to access valuable long-read information by leveraging a single sequencing instrument for multiple applications, without requiring additional specialized equipment. This method provides greater insight into the genome at a fraction of the cost of conventional approaches.

Simple Analysis and Assembly

Push-button analysis in the BaseSpace environment simplifies assembly of long reads. Data can be transferred from an Illumina sequencing instrument to the BaseSpace cloud instantly. Designed for use with the TruSeq Synthetic Long-Read DNA Library Prep and Barcode Kits, the TruSeq Long-Read Assembly App[5] constructs long sequences from shorter sequencing reads. The intuitive user interface (Figure 5) simplifies data analysis so that researchers can analyze data simply by selecting the project and file destination. The app then processes the short reads, assembles the initial contigs using overlap-based methods, and finally creates contig scaffolds to generate synthetically long reads (Figure 6). The assembled reads are exported in standard FASTQ format, which can be imported directly into downstream assembly tools for further analysis. The TruSeq Long-Read Assembly App enables one-click informatics for novice users, without requiring extensive expertise or infrastructure.

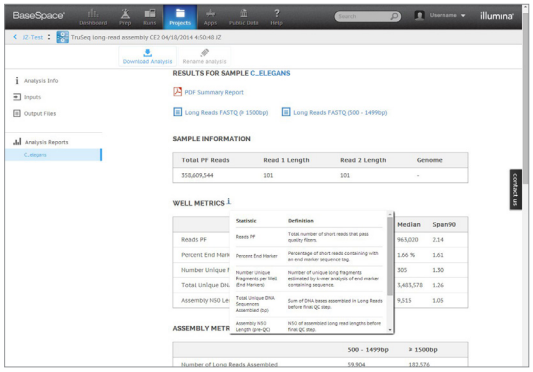
Summary

By combining shorter reads, TruSeq Synthetic Long-Read technology enables researchers to leverage a single sequencer and obtain more accurate long reads than conventional approaches. Designed with biologists in mind, the TruSeq Long-Read Assembly App simplifies bioinformatics so that researchers can spend less time analyzing data and more time focusing on their research. TruSeq Synthetic Long-Read technology provides a comprehensive solution for genome assembly and genome finishing.

References

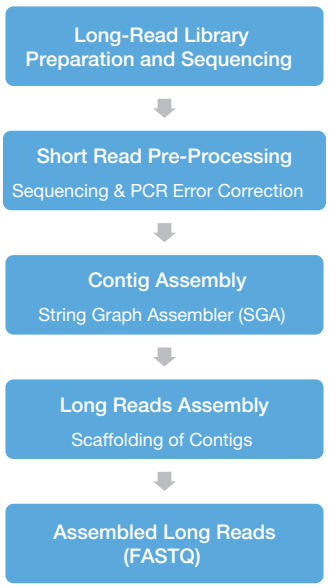
1. TruSeq | Illumina (www.illumina.com/truseq) Accessed 06 June 2014.
2. Public Data - BaseSpace (basespace.illumina.com/datacentral) Accessed 06 June 2014.
3. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 15: 1072–5.
4. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. Genome Biol 5: R12.
5. BaseSpace (basespace.illumina.com) Accessed 06 June 2014.

Figure 5: Simplified Analysis



Featuring an intuitive user interface, the TruSeq Long-Read Assembly App simplifies data analysis for any biological researcher, regardless of bioinformatics expertise. Results are displayed in straightforward tabular and graphical formats.

Figure 6: TruSeq Long-Read Assembly App Workflow



The TruSeq Long-Read Assembly App uses sequencing data to construct long reads, delivering results in a separate FASTQ file.

Ordering Information

Product	Catalog No.
TruSeq Synthetic Long-Read DNA Library Prep Kit (4 samples)	FC-126-1001
TruSeq Synthetic Long-Read DNA Barcode Kit (1 sample)	FC-126-1002
TruSeq Synthetic Long-Read DNA Barcode Kit (4 samples)	FC-126-1003
TruSeq Synthetic Long-Read DNA Accessory Kit	FC-126-1004

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2014-011 Current as of 06 June 2014

illumina®