Automated size selection of NEBNext® Small RNA libraries with the Sage Pippin Prep™

Daniela Munafo[1], Sadaf Hoda[2], Laurence Ettwiller[1], Brad Langhorst[1], Eileen Dimalanta[1], Fiona Stewart[1], Chris Boles[2]

[1] New England Biolabs, Ipswich, MA; [2] Sage Science, Beverly, MA.

Introduction

One of the fastest growing areas of biological research is regulatory small RNA structure, processing and function. Next generation sequencing (NGS) is the method of choice for studying the variety and expression of small RNAs.

A common problem in NGS methods for small RNAs is contamination from adapter-dimer artifacts, because these artifacts are very close in size to the small RNA library elements. To address this problem, the NEBNext Small RNA Library Library Prep Kits from New England Biolabs use specially engineered RNA ligases, optimized workflows and novel technology (patent pending) to dramatically reduce the formation of adapter-dimer artifacts during library construction. Since the workflow uses total RNA as the starting material, it is beneficial to perform a final size selection step on the amplified libraries. NEB has previously validated size selection methods using AMPure XP beads and manual preparative gel electrophoresis on a 6% polyacrylamide gel. Here, we validate the use of Pippin Prep 3% agarose gel cassettes for size selection of NEBNext Small RNA libraries.

General Protocol

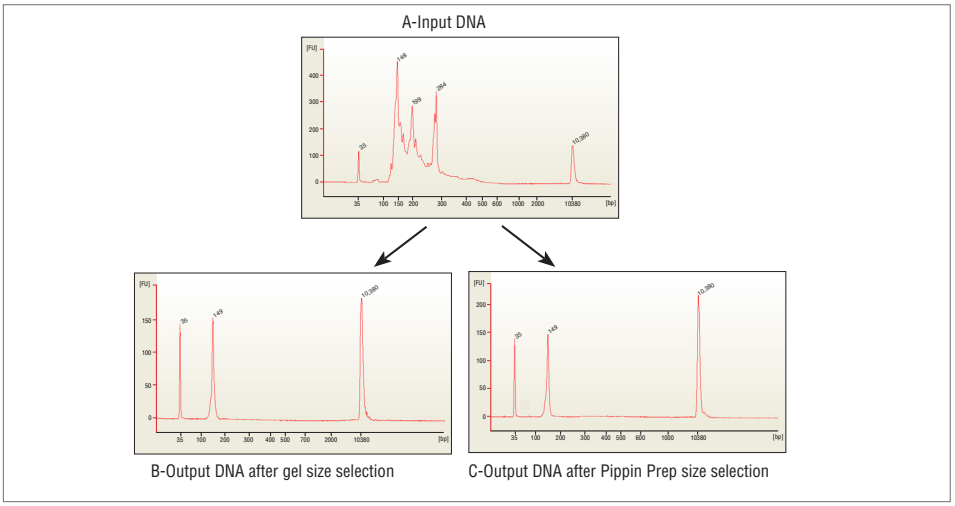
Small RNA Library Generation

1. Construct Small RNA libraries using the NEBNext Small RNA Library Prep Kit for Illumina.
2. Following PCR amplification, QC the libraries on an Agilent Bioanalyzer, using 1 µl of the purified PCR reaction on a DNA 1000 chip (according to manufacturer's instructions).

   Note: The miRNA library should appear as a peak at 147–149 bp (for a 21 nt insert; Figure 1A).
3. Perform size selection either manually, via polyacrylamide gel electrophoresis (PAGE) according to the NEBNext Small RNA Instruction Manual, or on an automated agarose gel elecrophoresis platform, such as Pippin Prep (Sage Science).

FIGURE 1:
Library size distribution before and after size selection

Bioanalyzer traces from NEBNext Human Brain miRNA libraries before size selection (A) and after size selection on a 6% polyacrylamide gel (B) or a 3% agarose (dye-free) cassette for Pippin Prep (C). Instrument Program Mode = Range; Start (bp) = 105 and End (bp) = 155 (C)



Applications

Small RNA library preparation for next-generation sequencing

Materials

NEBNext Multiplex Small RNA Library Prep Kits for Illumina® Set 1 (NEB #E7300) or Set 2 (NEB #E7580), Index Primers 1-48 (NEB #E7560) or Multiplex Compatible (NEB #E7330)

Pippin Prep Instrument with 3% Agarose Dye Free gel cassette with internal standards (Sage Science #CDP3010)

Bioanalyzer® (Agilent®) with DNA 1000 Chip and High Sensitivity Chip

Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB #T1030)

6% PAGE gel (for manual PAGE)

Purify and Concentrate the PCR-amplified NEBNext Small RNA Library (100 µl)

1. Purify and concentrate the PCR amplified library (100 µl) using a Monarch PCR & DNA Cleanup Kit following the protocol recommendations.

   IMPORTANT: Use the 7:1 ratio of binding buffer:sample. Discard the flow-through after each centrifugation step.

2. Elute purified and concentrated library in 32 µl nuclease-free water.

Size Selection of the Concentrated Small RNA Library Using a Pippin Prep 3% Agarose, Dye Free Gel Cassette

Create a protocol:

1. In the Pippin Prep software, go to the "Protocol Editor" tab.
2. Click "Cassette" folder, and select "3% DF Marker P". (Note: This is an internal standard cassette. Markers are provided premixed with sample loading solution.)
3. Select "Range" as the collection mode, and enter the size-selection parameters as follows:
   BP Start = 105 bp
   BP End = 155 bp
   The BP Range Flag should indicate "broad".
4. Click the "Use of Internal Standards" button.
5. Make sure the "Ref Lane" values match the lane numbers.
6. Press "Save As" and name and save the protocol.

Prepare a 3% Agarose, Dye Free Pippin Prep gel cassette:

1. Follow the instructions in the Pippin Prep Operations Manual (Chapter 6) for preparing and testing a cassette.

Load and run samples on Pippen Prep:

1. Bring loading solution to room temperature.
2. For each sample, combine 30 µl of library sample with 10 µl of DNA Marker P.
3. Vortex briefly to mix samples thoroughly. Briefly centrifuge to collect.
4. Load 40 µl (library sample plus marker) in each sample well of the 3% agarose cassette.
5. Run the program with the settings indicated above.
6. After the sample has been eluted, collect the size-selected sample (40 µl) from the elution module.

QC the size-selected products:

1. Run 1 µl of the size-selected library in an Agilent Bioanalyzer using the High Sensitivity Chip. (Note: Purification of the size-selected material is not required for QC by the Bioanalyzer.)

Data analysis:

1. Map the trimmed and length-filtered reads to miRBase (Release 20, human) using bowtie2 (option – sensitive-local – non-deterministic).
2. Map remaining reads, that do not map to miRBase, to the human genome (hg19) using Star (option – outFilterMatchNmin 20 – outFilterMistmatchNoverLmax 0.05 – outFilterMultimap-Nmax100000).
3. Compare the position of the mapped reads to genomic features described in gencode annotation (v17).

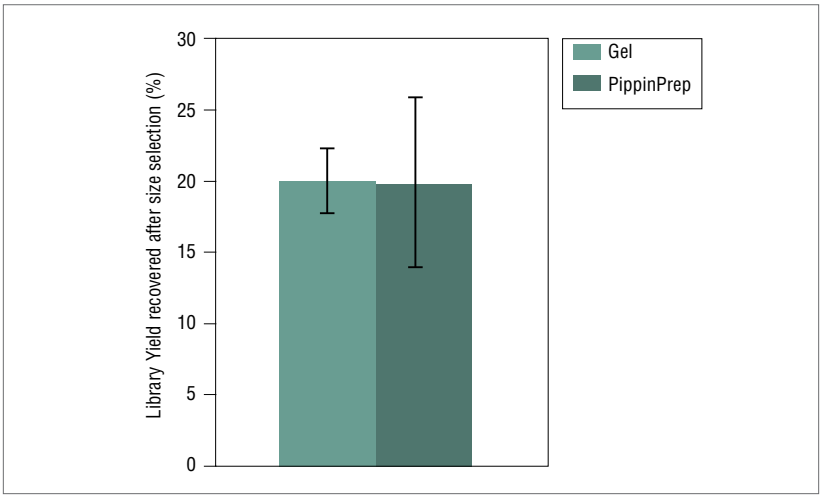
Results

Library profile before and after size selection

The experiments were designed to compare NEBNext Small RNA library quality using two gel-based protocols: manual preparative electrophoresis on 6% PAGE gels, and automated agarose gel electrophoresis on the Pippin Prep.

Six human brain small RNA libraries were prepared from total RNA and indexed using different barcodes. Three libraries were PAGE-size-selected (Figure 1B) and the other three libraries were size-selected on the Pippin Prep using a 3% dye-free agarose cassette (Figure 1C). Both size selected libraries have only a single peak at ~149 bp, with minimal contamination from smaller or larger species. Library yields from the size-selected libraries using the manual gel and Pippin Prep were similar (~20% of the input material is recovered; Figure 2).

FIGURE 2: **Library yield after size selection.**
The percentage of library yield recovered after size selection is comparable for both size selection methods.



MiSeq® sequencing and data analysis

The 6 barcoded libraries were pooled in equimolar concentrations, loaded onto a MiSeq reagent v2 kit at 8 picomolar final concentration, and sequenced on a MiSeq instrument (SE; 1X 36 bp; 2.5 million of reads/library).

Adapter trimming and filtering reads by length

Reads were adapter-trimmed and filtered by length. Reads shorter than 15 nucleotides were discarded. A high percentage of reads (> 90% total reads for both size selection methods) passed the length filtering. The NEBNext Small RNA library prep optimized workflow prevents adapter-dimer formation, therefore only a minimal percentage of reads did not contain insert.

Mapping rate

A high percentage of trimmed reads (> 66%) mapped to human miRNAs present in miRBase for both gel and Pippin Prep size selection methods (Figure 3). The vast majority of trimmed reads (98%) mapped either to the human genome or to miRBase. From the reads that mapped to human, an average of 28% overlapped at least one exonic feature (miRNA, lincRNA, pseudogene, snRNA, protein coding RNA, snoRNA, rRNA, sense_intronic, mt-tRNA, antisense RNAs) (Figure 4). Some of the reads that did not map to miRBase overlapped with a putative microRNA, increasing the number of reads mapping to various microRNAs (from miRBase or gencode annotations) to more than 68% of total trimmed reads.

FIGURE 3: **Mapping Rate.**
Libraries have a high percentage of reads mapping to miRBase.

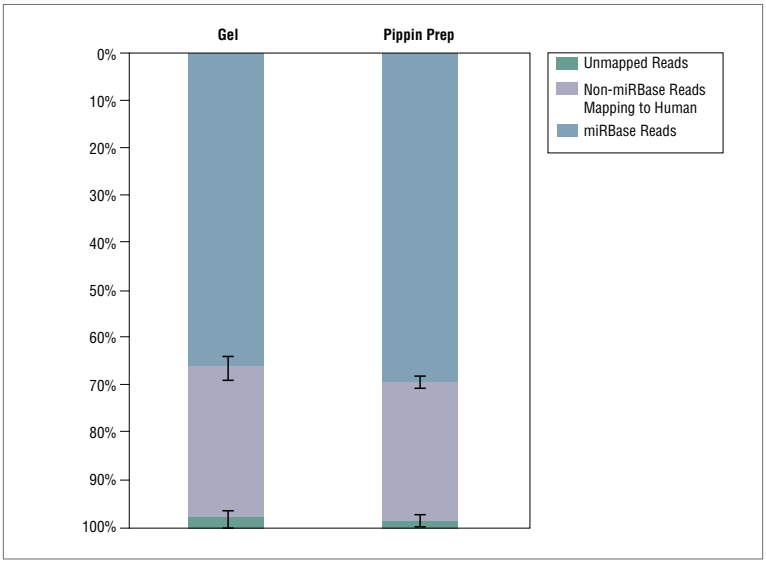
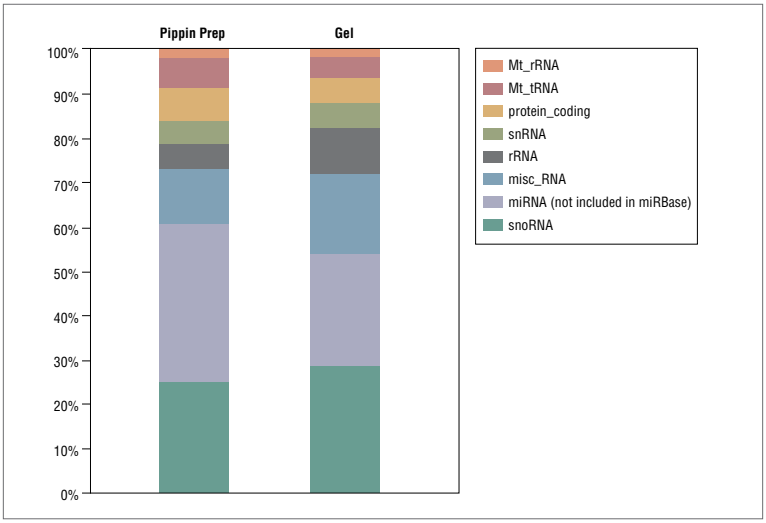


FIGURE 4:
Mapping to the human genome
After aligning to miRBase, remaining reads were mapped to the human genome. The majority of the reads (72%) mapped to un-annotated regions of the genome. The remaining 28% mapped to small nucleolar RNA (snoRNA), miRNA, misc_RNA, ribosomal RNA (rRNA), small nuclear RNAs (snRNA), protein _coding RNA, Mitochondrial tRNA (Mt_tRNA) and Mitochondrial ribosomal RNA (Mt_rRNA). Very few reads mapped to pseudogenes, long intronic RNA (lincRNA), processed_transcripts and antisense RNA.

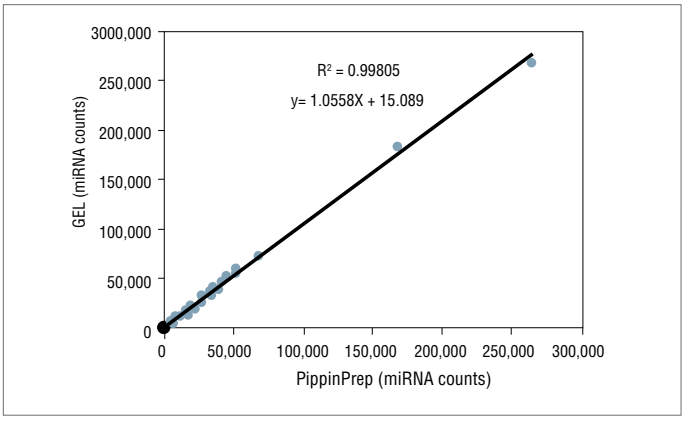


miRNA Expression Analysis

miRNA expression correlation of the 500 most abundant miRNAs between PAGE and Pippin prep size-selected libraries was excellent [(R^2 = 0.99805) (y = 1.0558x + 15.089)] (Figure 5). This data indicates that small RNA expression levels were not biased due to different size selection methods.

FIGURE 5:
miRNA correlation

miRNA expression correlation was excellent between both size selection methods (N = 500 most abundant miRNAs).



Conclusions

The experiments reported here show that NEBNext Small RNA libraries produced using the Sage Pippin Prep equal or exceed the quality of libraries size selected using the manual gel procedure. The benefits are seen in library purity.

Two other key features of the Pippin Prep that are not shown in the present experiments are reproducibility and ease-of-use. The manual gel procedure has many individual manual steps that are time-consuming and extremely difficult to perform reproducibly. In contrast, the Pippin Prep procedure requires no manual manipulations except for gel loading. The entire separation and size-selection process is controlled by the onboard computer. This removes all opportunity for the user to introduce variability into the process.

Hands-on effort and time for the Pippin Prep procedure is 15–20 minutes to set up and load a cassette, and less than 5 minutes to remove samples at the end of the run. Run time for the NEBNext Small RNA libraries in the 3% Agarose, Dye Free Gel Cassette with internal standard is a little over one hour. Up to five samples can be run per cassette.

In summary, we have developed an optimized Pippin Prep protocol for use with the NEBNext Small RNA Library Prep Kits. The Pippin Prep protocol provides all of the sequence quality benefits of the standard manual gel protocol, but with greatly enhanced ease-of-use and reproducibility.