

Research Article

Open Access

Ping Xu, Martina Billmeier, Irina Mohorianu, Darrell Green, William D Fraser, Tamas Dalmay

An improved protocol for small RNA library construction using High Definition adapters

Abstract: Next generation sequencing of small RNA (sRNA) libraries is widely used for studying sRNAs in various biological systems. However, cDNA libraries of sRNAs are biased for molecules that are ligated to adapters more or less efficiently than other molecules. One approach to reduce this ligation bias is to use a pool of adapters instead of a single adapter sequence, which allows many sRNAs to be ligated efficiently. We previously developed High Definition (HD) adapters for the Illumina sequencing platform, which contain degenerate nucleotides at the ligating ends of the adapters. However, the current commercial kits produced a large amount of 5' adapter – 3' adapter ligation product without the cDNA insert when HD adapters were used to replace the kit adapters. Here, we report a protocol to generate sRNA libraries using HD adapters with greatly reduced proportion of adapter-adapter products due to the degradation of non-ligated 3' adapters. The libraries can be completed within two days and can be used for various biological and clinical samples. As examples for using this protocol, we constructed sRNA libraries using total RNA extracted from cultured mammalian cells and plant leaf tissue.

Keywords: RNA-seq, reduction of ligation bias, *Medicago truncatula*, SW1353 chondrosarcoma cell line

DOI 10.1515/mngs-2015-0001

Received September 10, 2014 accepted December 1, 2014

***Corresponding author: Tamas Dalmay:** School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK, tel.: 0044 1603 593221, fax: 0044 1603 592250, E-mail: t.dalmay@uea.ac.uk
Ping Xu, Martina Billmeier, Irina Mohorianu, Darrell Green: School of Biological Sciences, University of East Anglia, Norwich, United Kingdom
Darrell Green, William D Fraser: Norwich Medical School, University of East Anglia, Norwich, United Kingdom

1 Introduction

Small RNAs (sRNAs) ranging from 20 to 30 nucleotides (nts), such as small interfering RNAs (siRNAs) and microRNAs (miRNAs), are important regulators of gene expression. They play important roles in development and responses to biotic and abiotic stress [1, 2]. Thus, it is of great interest to investigate diverse sRNAs and their expression levels in a given sample. There are several methods for miRNA expression analysis such as hybridization-based miRNA microarrays, quantitative reverse transcription-PCR and sRNA library sequencing [3]. However, only sRNA library sequencing provides sequence information such as known and new sRNAs and corresponding sequence frequencies. Next generation sequencing (NGS) of sRNA libraries enhances the power of new sRNA discovery and largely widens the quantification range. It is now the most common approach used to discover new regulatory sRNAs and for sRNA quantification. There are several commercially available NGS platforms such as Illumina, Applied Biosystems (ABI) SOLiD and 454 Life Sciences; while their corresponding kits for library construction are also available. Among them, the Illumina sequencing platform can perform massively parallel sequencing of 50-100bp DNA and is the most widely used platform for sequencing sRNA libraries.

Different NGS platforms (or even the same platform when different versions of kits are used for library preparation) produce different results. The results from NGS are often not consistent with the results from other quantitative analysis [4-9]. One major source of the problem is the RNA ligases used in sRNA library construction [10-13]. Both T4 RNA ligase 1 and 2 (T4Rnl1 and T4Rnl2), and more recently the truncated T4Rnl2 mutant with reduced side reactions [14], are used in ligating sRNAs to customized adapters during library construction. The ligation efficiency of a sRNA to an adapter is influenced by the ability of the sRNA to anneal to the adapter [10, 12, 13]. If a sRNA can anneal to an adapter, that sRNA has a higher chance to get ligated to the adapter than a sRNA without such ability

(probably because those molecules would only get close to an adapter by chance). Therefore sRNAs with different sequences have different ligation efficiencies to certain adapters, which lead to over- or under-representation of sRNAs in a library. Also, the sequences of adapters in the kits for different sequencing platforms (or in the kits of different versions for the same platform) are different; therefore the sRNA profiles obtained are different and incomparable. To reduce the bias, mixed adapter pools were used for sRNA cloning so that sRNAs have options to be annealed to different sequences instead of one single sequence. This type of modification has improved the discovery of sRNA sequences and quantitative profiling of sRNAs [11-13, 15, 16].

In order to reduce ligation bias we developed High Definition (HD) adapters, by adding four degenerate nucleotides at each ligating end of the Illumina adapters [12]. The Illumina version 1.5 kit can be used by directly replacing the Illumina adapters with HD adapters. However, when HD adapters were used along with the most recent multiplexing Illumina small RNA preparation kit Truseq 2.0, the library was overwhelmed by the 5' adapter-3' adapter ligation product without a cDNA of sRNA insert (adapter-adapter product) (see Results). The contaminating adapter-adapter products decrease the number of usable sequence reads. This happened more often when only small amount of RNA was available for ligation or when ligation efficiency was low due to lower RNA quality. It was a common problem for sRNA cloning that adapter-adapter ligation products were reverse transcribed and amplified because excessive adapters were used to ensure sufficient desired ligation products. Thus, it was often recommended to purify the sRNAs with 3' end being ligated to or tagged by the customized adapter through gel extraction. This method was successful in generating the sRNA libraries with degenerate or mixed adapters [11, 13, 17]. However, the protocol includes one extra day of gel extraction with the possibility of sRNA loss during the extra gel extraction. In the meantime, Epicenter developed the Scriptminer multiplexing sRNA preparation kit where a degradase, RecJ exonuclease, was introduced to degrade the excessive 3' adapter before ligating the 5' end of sRNA to the 5' adapter [18, 19] (<http://www.google.com/patents/WO2011056866A2?cl=en>). The kit worked well with many RNA samples [20] while it failed with some others, due to the varying degrees of adapter-adapter products (Xu, unpublished data). This kit has now been discontinued since the end of 2013. Consequently, it was important to develop a method to minimize the amplification of adapter-adapter product. Based on the invention by Vaidyanathan et al [19], we developed a protocol to generate high quality sRNA libraries by using HD adapters without purifying ligated sRNAs through gel extraction.

2 Methods

2.1 Materials:

2.1.1 *Medicago truncatula* leaves

The seedlings were grown in a growth room with 16 hours of daylight under constant 22°C. The leaves were harvested from the plants grown for 4 and 7 weeks. They were frozen in liquid nitrogen and stored at -80°C until total RNA extraction.

2.1.2 SW1353 chondrosarcoma cells

Cells were cultured in Dulbecco's modified Eagle's medium (Life Technologies) containing 10% (volume/volume) fetal bovine serum (Sigma Aldrich), 2 mM glutamine and 1% (volume/volume) penicillin-streptomycin. Cells were fed every other day and maintained at 37 °C in an atmosphere of 5% CO₂. After 21 days of culture, cells were harvested for total RNA extraction.

2.2 RNA isolation

For *Medicago truncatula* leaves, total RNA was isolated with TRIzol reagent (Invitrogen). The total RNA or the sRNA fraction were further purified by using the mirVana miRNA isolation kit™ (Ambion). We followed the protocols provided by both manufacturers. After the RNA was eluted, it was concentrated by ethanol precipitation, where 0.1 volume of 3M NaOAc, 3 volume of absolute ethanol and 25 ug/mL glycogen (Ambion) were mixed and precipitated at -20 °C overnight. The precipitate was spun down by centrifugation at 20,000 x g for 20 minutes (min) at 4 °C and washed with 80% ethanol. The pellet was air-dried for 5-8 min at room temperature and re-suspended with nuclease free H₂O. The concentration of RNA was determined by nanodrop. The RNA was stored at -80°C.

For the SW1353 chondrosarcoma cell line, total RNA was isolated using the miRCURY™ RNA Isolation kit (Exiqon). The RNA concentration was determined by Nanodrop and RNA was stored at -80°C.

2.3 Small RNA library construction

sRNA libraries were constructed with HD adapters [12] using Illumina Truseq2.0, Epicentre Scriptminer™ Small RNA-seq library preparation and modified Scriptminer method. For the commercial kits, we followed the manuals

provided by the manufacturers but replaced the adapters with HD adapters. The new protocol takes two days to complete. During the first day the sRNAs are ligated to the adapters and the ligation products are turned into DNA by RT-PCR and the PCR product is gel purified on the second day (Figure 1). The detailed protocol is provided at the end of the manuscript, here we give a short summary and some details that are not in the protocol.

5' HD adapter (5'-/5AmMC6/GUUCAGAGUUCUACAGUCCGACGAUCNNNN-3'; where all

nucleotides are ribonucleotides) was synthesized through IDT, re-suspended in nuclease free H₂O at 100 μM as a stock solution and stored at -80°C. 3' HD adapter [5'-NNN NTGGAATTCTCGGGTGCCAAGG(2'3'ddC)-3'] (where all nucleotides are deoxy-ribonucleotides) was synthesized through Sigma, re-suspended in nuclease free H₂O at 100 μM as a stock solution and stored at -20°C. 3' HD adapter was phosphorylated with T4 polynucleotides kinase (NEB). The phosphorylated product was concentrated by ethanol precipitation and adenylated with 5' DNA Adenylation

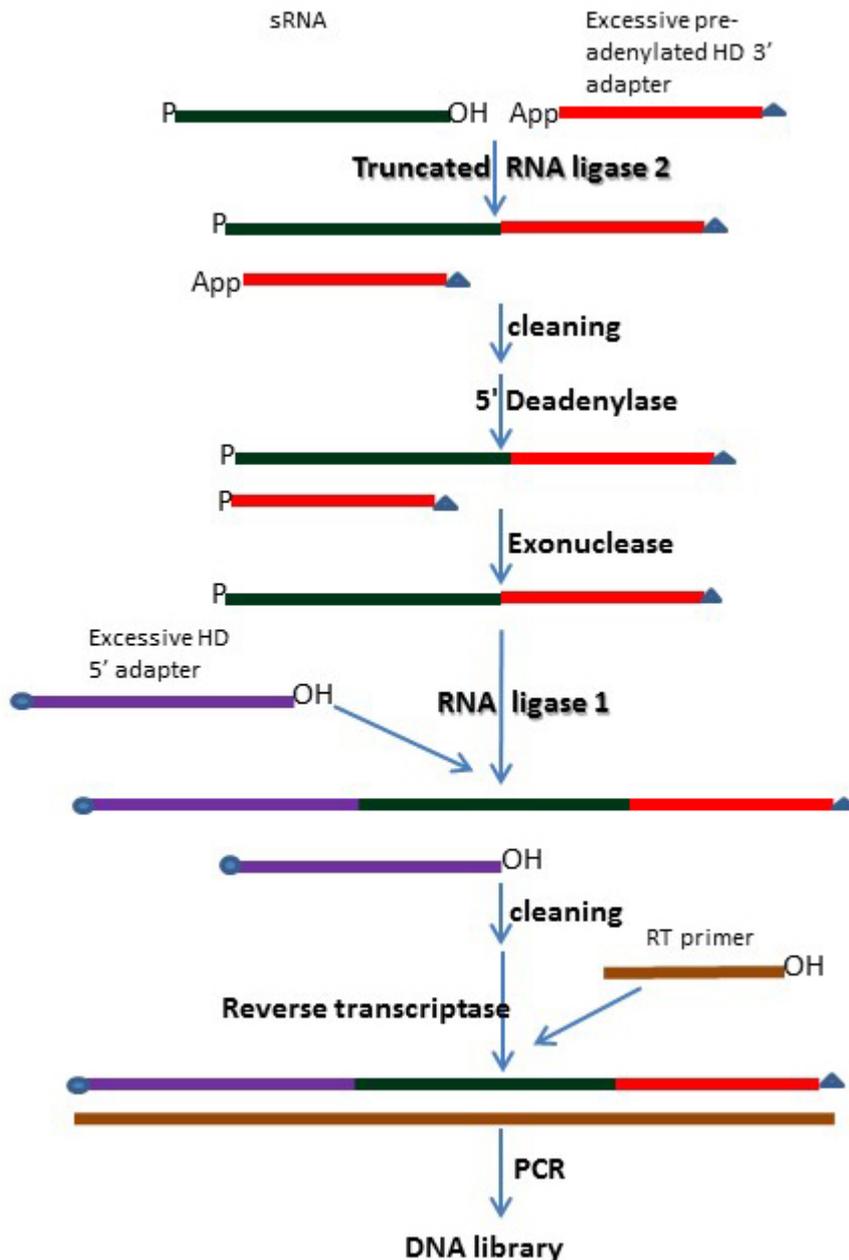


Figure 1. Workflow of the protocol. The steps shown in the workflow can be completed within one day. If there are many libraries to be constructed, one may consider to stop at the RT step. Following PCR amplification, the products are separated with PAGE and the bands around 145-150bp are extracted for illumina sequencing, which takes another 1-2 days depending on the number of libraries.

kit (NEB) following the protocols recommended by the manufacturer. The reaction mix was cleaned and concentrated by phenol:chloroform extraction and ethanol precipitation. The precipitate was re-suspended with nuclease free H₂O at final concentration of 10 µM.

Library construction started with 3' end ligation using truncated T4 RNA ligase 2 (T4Rnl2, NEB) followed by cleaning with an RNA clean and concentrator kit™ (Zymo Research) and eluting in water. After the sRNAs were ligated to the 3' HD adapter, the remaining adapter was degraded with RecJ Exonuclease after de-adenylation. After the degradation of excessive 3' adapter the 5' adapter was ligated to the sRNA-3' adapter molecules, the product was cleaned using RNA clean and concentrator kit™ (Zymo research) and eluted. After the sRNAs were ligated to both HD adapters, reverse transcription (RT) was carried out to generate cDNA, using the RT primer (5' GCCTGGCACCCGAGAATTCCA 3'). The RT product was PCR amplified using the Illumina RP-1, 5' AATGATACGGC-GACCACCGAGATCTACAGTTCAGAGTTCTACAGTCCGA 3') forward primer and reverse primers containing different index sequences (underlined) (5' CAAGCAGAAGACGGCA-TACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCCGAG-AATTCCA 3'; 5' CAAGCAGAAGACGGCATAACGAGATGG-CCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA 3'; 5' CAAGCAGAAGACGGCATAACGAGATCGTACGGTACTG-GAGTTCCTTGGCACCCGAGAATTCCA 3'). To determine the optimal number of reaction cycles, multiple reactions at 20 µL were set up and different number of cycles were tried (usually between 10 and 15). In the present study, several samples were amplified well with 12-14 cycles. After PCR amplification, the products were separated on 8% polyacrylamide gels and the bands at sizes 145-150 bp were excised from the gel. The DNA was purified from the gel, the concentration was determined by Nanodrop and sent for sequencing at The Genome Analysis Centre (TGAC) in Norwich, UK using the HiSeq2500 platform.

2.4 Bioinformatics analysis

The analysis was conducted on version 3.5 v5 of the *Medicago truncatula* genome and related annotations. The analysis was also conducted on version 38 of the Human genome and related annotations. The *Medicago truncatula* chloroplast genome, accession NC_003119 was downloaded from <http://chloroplast.ocean.washington.edu/>.

First, the fastq files from NGS were converted to fasta files and sequence reads with no "Ns" were kept for further analysis. Next, the first 6 nts (TGGAAT) of the 3' adapter were identified and trimmed off. The four nts on the 5'

and 3' ends of the reads (which corresponded to the four degenerate nts on each HD adapter) were also trimmed using the UEA sRNA Workbench [21]. The sequence reads were mapped to the reference genomes and the *Medicago truncatula* chloroplast genome with 0 mis-matches, in non-redundant format, using PatMaN [22]. The expression levels of sRNAs for the *Medicago truncatula* samples were normalized using the scaling method proposed by Mortazavi et al [23].

The conserved miRNAs were identified using miRProf within the UEA sRNA Workbench [21]. The location of hairpins and secondary structures were determined using Vienna RNA package [24]. The new miRNAs were predicted using miRCat with default parameters for plant and animal samples.

The data presented in this paper is available in fastq format on Gene Expression Omnibus (GEO) [25] under the accessions GSE59330 (GSM1435111 and GSM1435112) for *Medicago truncatula* and GSE59331 (GSM1435113) for the SW1353 chondrosarcoma cell line.

3 Results

3.1 An improved method for generating cDNA libraries for sRNAs using High Definition adapters

A large amount of adapter-adapter cDNA was produced with HD adapters using Illumina Truseq 2.0 kit (Figure 2A) when 2 µg *M. truncatula* sRNA was used for library construction. Although there was a band at the size of about 150 bp, further sequencing analysis (both Sanger and NGS) indicated that 95% of the product was adapter-adapter product (data not shown). When less sRNA or total RNA was used, the results were even worse because the band around 150 bp was not visible (data not shown).

When Epicentre Scriptminer multiplexing sRNA preparation kit was used, the band around 150 bp was clearly visible, while there was still some accumulation of adapter-adapter product (Figure 2B). However, this kit was discontinued at the end of 2013 therefore there was a need to develop an alternative solution. We followed the same approach but optimised the buffers for 3' adapter ligation and the subsequent de-adenylation and DNase steps. When this modified method was used, the accumulation of adapter-adapter product was drastically reduced while the library band appeared much stronger (Figure 2C). This improvement was further confirmed in other total RNA or sRNA fraction samples from plant, mammalian, insect and fungal samples which failed to produce the expected products using the

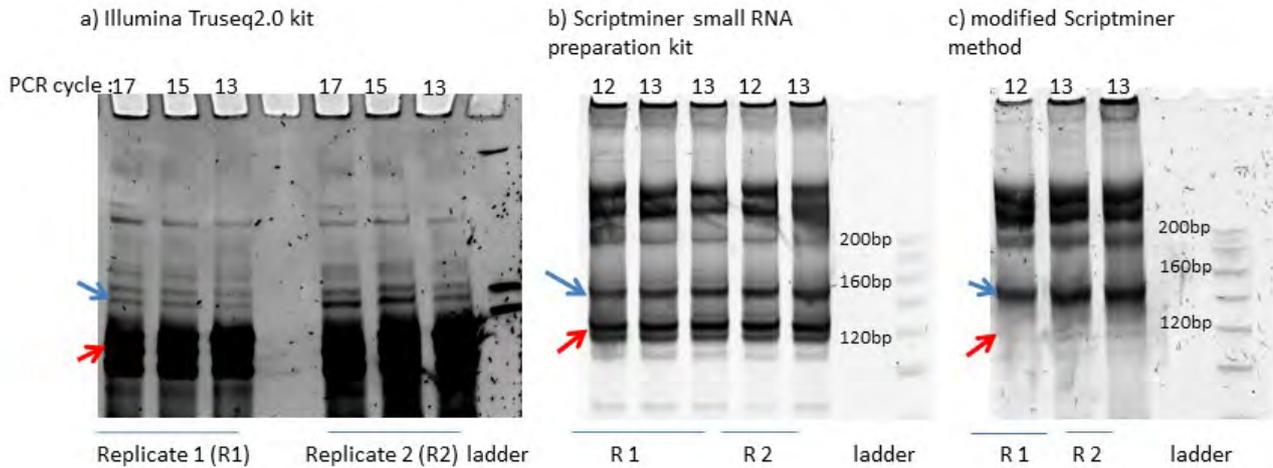


Figure 2. Less adapter-adapter ligate products were produced when the modified method was used for constructing medicago leaf sRNA libraries. The blue arrows point to the band with 20-30 nt ds cDNA inserts and the red arrows point to the band of adapter-adapter products.

original Epicentre Scriptminer kit, but yielded good results with the modified method (data not shown).

In addition, the amount of RNA required for library construction was reduced with the modified method for library construction using HD adapters. When 1-5 μg *Medicago truncatula* leaf total RNA or less than 500 ng of purified sRNA fraction was used using the original Scriptminer kit, the adapter-adapter product was too strong (data not shown). However, with the improved protocol 1 μg of total RNA and 300 ng of purified sRNA fraction can produce a usable library band; although more RNA tends to produce stronger library bands and

less adapter-adapter product (Figure 3). To validate the new method and demonstrate that the bands at the desired size contained the PCR products of sRNA cDNA, the libraries from the two replicates of 1 μg *Medicago truncatula* leaf total RNA and the one library from 500 ng total RNA from mammalian SW1353 chondrosarcoma cells were submitted for sequencing.

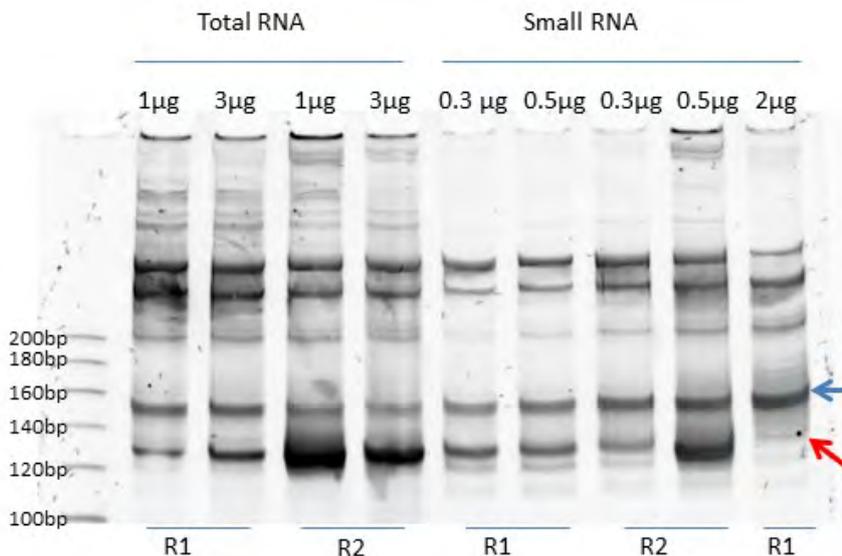


Figure 3. Using the modified method, we were able to generate good sRNA libraries from small amount of total or small RNA from *Medicago truncatula* leaves.

3.2 Evaluation of the libraries by bioinformatics analysis

We obtained 12-13 million raw reads from the three libraries and ~95%, 86% and 76% of the reads in these three libraries, respectively, contained both adapters and cDNA inserts, thus coming from the desired amplified products (Table 1

and Supplementary Table S1). These libraries contained only 5%-7% of the reads originating from adapter-adapter products. More than 72% of the reads in *M. truncatula* libraries and 64% of the reads in the chondrosarcoma cell line library were mapped to the corresponding genome sequence (Table 1 and Supplementary Table S1). In the chondrosarcoma cell line library the most abundant class

Table 1. General information of *M. truncatula* leaf libraries

| | mtr1* | | | | | mtr2* | | | | |
|-------------------------------|-------------|------------|--------------|------------|--------------|-------------|------------|--------------|------------|--------------|
| | total reads | proportion | unique reads | proportion | complexity** | total reads | proportion | unique reads | proportion | complexity** |
| original total reads | 13,572,778 | | | | | 12,118,554 | | | | |
| adapter-adapter product | 784,912 | 0.05 | | | | 56,103 | 0.004 | | | |
| total reads after cleaning*** | 11,743,358 | 0.86 | 4,933,08 | 0.42 | 0.42 | 11,543,965 | 0.95 | 5,412,196 | 0.46 | 0.46 |
| chloroplast genome | 607,141 | 0.05 | 51,34 | 0.01 | 0.08 | 720,030 | 0.06 | 59,726 | 0.01 | 0.08 |
| genome match | 8,493,508 | 0.72 | 3,233,01 | 0.65 | 0.38 | 8,499,234 | 0.73 | 3,541,120 | 0.65 | 0.41 |
| coding region of mRNAs | 1,219,284 | 0.14 | 292,71 | 0.09 | 0.24 | 1,205,567 | 0.14 | 308,349 | 0.08 | 0.25 |
| transposable elements | 1,407,441 | 0.16 | 449,48 | 0.13 | 0.31 | 1,320,197 | 0.15 | 475,417 | 0.13 | 0.36 |
| known microRNAs | 1,651,507 | 0.19 | 19,42 | 0.00 | 0.01 | 1,536,323 | 0.18 | 20,760 | 0.006 | 0.01 |

* *Medicago truncatula* leaf tissue replicate 1 and 2.

** complexity is the ratio of unique reads to total reads

*** all the sequences containing N, the sequences without adapter sequences, and the cDNA insert smaller than 17nts were removed.

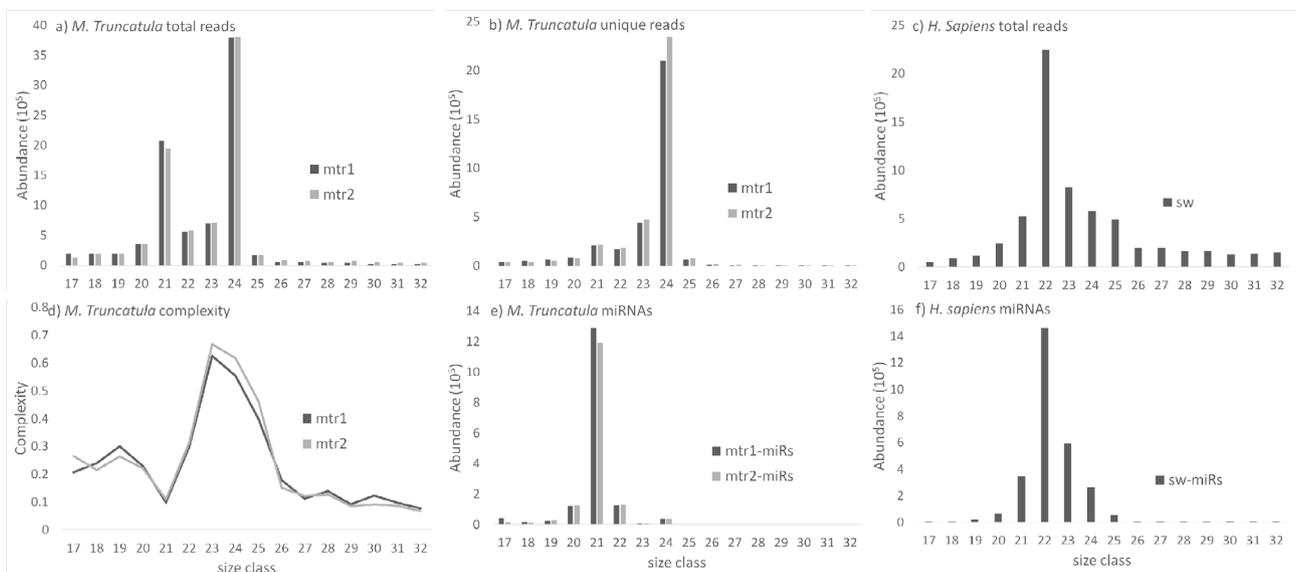


Figure 4. Results from initial bioinformatics analysis showed the expected characteristics of animal and plant sRNA libraries when they were constructed using the modified method.

(45%) was the known miRNAs (Supplementary Table S1), which were mainly 22 nts (Figure 4E).

Due to the more diverse sRNA population in plants, we analysed the sRNA profiles from leaf in more detail. The data from the two replicates were consistent (Figure 3 and Supplemental Figure S1). First we analysed the redundant reads where all reads were used for the analysis (i.e. if a certain sequence was found 5000 times, all 5000 reads were included in the analysis). The majority (94-95%) of the redundant reads mapped to the nuclear genome and about 5-6% mapped to the chloroplast genome. For the nuclear genome matching reads about 18-19% of sRNAs were known miRNAs, 14% of the sRNAs mapped to the coding region and 16% to transposable element sequences (Table 1). Both leaf libraries contained more than 3 million unique sequences where we collapsed the read numbers of identical reads (i.e. if a certain sequence was found 5000 times, that sequence was considered as one in the analysis). While 13%-14% of unique sequences were mapped to transposable elements and 9% to coding region, only 0.6% of the unique sequences were mapped to known miRNAs. The size distribution of cloned sRNAs was bimodal in both leaf libraries with the most abundant sRNAs at 24 nt, followed by 21 nts (Figure 4A). This is similar to previous reports [26, 27]. The most diverse unique reads were 24 nts sRNA (Figure 4B) and the least complex sRNAs were 21 nts (Figure 4C). The majority of miRNAs in *Medicago truncatula* leaf libraries were 21 nts (Figure 4D; Supplementary Table S2). These initial analyses showed that both the human and the plant libraries showed the expected characteristics therefore the new method did not introduce any unexpected problem.

4 Discussion

One of the potential difficulties with cloning sRNAs is the generation of a large amount of adapter-adapter PCR products (Figure 2). An excessive amount of adapters are normally used to ensure sufficient ligation because of the low efficiency of RNA ligases but this leads to a huge amount of adapter-adapter molecules that can dominate the product [28]. Before the availability of Illumina 1.5 and Truseq 2.0, it was standard practice to purify the sRNAs via gel extraction after each ligation step. The Illumina 1.5 kit was the first where a truncated form of T4Rnl2 was used that was able to ligate a 5' adenylated 3' adapter. Using T4Rnl1 for the subsequent 5' adapter ligation reduced the adapter-adapter product because T4Rnl1 cannot ligate 5' adenylated end. However, this approach did not lead to a complete abolition of adapter-adapter product and other

methods were later introduced to solve this problem. One such method is to use a locked nucleotide oligo that was complementary to the junction region of the two adapters to stop the reverse transcription of the dimer [29]. The second approach uses an RT primer to anneal the remaining 3' adapter and form a double stranded DNA oligo before 5' adapter ligation because dsDNA is not a preferred substrate of T4 RNA ligase 1 [30]. However, none of these methods were successful with HD adapters that contain degenerate oligonucleotides at the ligating ends of the adapters, probably due to the random sequences. The third option is to use a single strand specific DNA exonuclease to degrade the remaining 3' adapter before 5' ligation [18, 19], which was effective for HD adapter-based cloning (Figure 2).

Our protocol is based on the single strand specific DNA exonuclease to degrade the non-ligated 3' adapters but we improved the buffer systems during the first few steps of the protocol. The original Epicentre kit carried out 3' adapter ligation and de-adenylation in the same buffer. It simplified the protocol but compromised the buffer. Our protocol is expected to improve ligation efficiency between sRNAs and 3' adapter because we use the buffer recommended by the manufacturer followed by a purification step. The subsequent de-adenylation is then carried out in a second buffer. The cost of the protocol is about £40/library (when the manuscript was submitted in November 2014), from which the purification step after the first ligation, the de-adenylation and adapter degradation is about £10/library. Using this protocol, we successfully constructed HD adapter-based sRNA libraries from various biological samples. This procedure does not involve gel extraction of ligated products and the high quality multiplexing libraries can be obtained within two or three days, depending on the number of samples. Results from the sequence data analysis with the example libraries of a human cell line and *Medicago truncatula* leaf showed low contamination of adapter-adapter product, expected size distributions of the sequenced sRNAs and known miRNAs.

There are possibilities for further improvements to this protocol. Some recent studies have suggested that the reaction conditions for ligation can be improved for increased efficiency and reduction of ligation bias. For example, adding the crowding molecule PEG to certain concentrations, increasing the adapter concentration, increasing the amount of RNA ligases, selecting a certain type of RNA ligase and optimising elongating reaction time were all useful in reducing ligation bias to many miRNAs [17, 31]. Some conditions were also found to have a general improvement on cloning largely mixed sRNA populations in real biological samples when mixed or degenerate adapters were used [17, 31, 32].

Acknowledgement: The work was funded by a BBSRC Sparking Impact Award granted by the University of East Anglia to Tamas Dalmay.

Protocol for constructing small RNA library using HD adapters

1. 3' Adapter ligation

1.1 Warm up 50% PEG Solution (NEB) and keep it at room temperature. Set up heat blocks at 70°C and 26°C.

1.2 Mix 11.25µL of the RNA sample with 1µL of 10µM pre-adenylated 3'-HD adapter. Incubate it at 70°C for 2 minutes and place it on ice

1.3 Add the following reagents into the above denatured RNA and adapter mix. Incubate it at 26°C for 2-3hrs.

| | |
|--|---------|
| T4 RNA ligase 2 truncated 10X buffer | 2 µL |
| RNaseOUT 40U/µL (in vitrogen) | 0.75 µL |
| 50% PEG Solution | 4 µL |
| T4 RNA ligase 2 truncated (200 U/µL) (T4Rnl2, NEB) | 1 µL |

1.4 Use RNA clean and concentrator kit™ (Zymo Research) to clean up the reaction. Elute it in 12.1 µL nuclease free water. Place the elute on ice.

2. 3' Adapter removal

2.1 Add the following reagents into the above 12.1µL eluted RNA-adapter ligation mixture. Incubate it at 30°C for 30min.

| | |
|-----------------------------------|--------|
| 10X deadenylase buffer | 1.6 µL |
| 100mM DTT | 0.8 µL |
| RNaseOUT 40U/µL | 0.5 µL |
| 5' deadenylase 10U/µL (Epicentre) | 1.0 µL |

2.2 Add 4µl of 25 mM EDTA to stop the reaction.

2.3 Mix the above reaction product with the following reagent and incubate it at 37°C for 30min

| | |
|--|------|
| 500 mM Tris-HCl pH 9.0 | 2 µL |
| 50 mM MgCl2 | 7 µL |
| RecJ Exonuclease 10U/µL (Epicentre or NEB) | 1 µL |

3. 5' Adapter ligation

3.1 Incubate 1µL 20µM 5' adapter at 70°C for 2min and place it on ice. Mix it with the reaction mix from step 2.3.

3.2 Add the following reagents for ligation and incubate the mixture at 26°C for 2-3hrs.

| | |
|-----------------------------|------|
| 10xT4 RNA ligase buffer | 1 µL |
| 50% PEG | 7 µL |
| 10mM ATP (Epicentre) | 1 µL |
| T4 RNA ligase (T4Rnl1, NEB) | µL |

3.3 Use RNA clean and concentrator kit™ (Zymo Research) to clean up the reaction. Elute it in 30 µL nuclease free water. Place the eluted product on ice. It can also be stored at -80 °C.

4. cDNA synthesis

4.1 Add the following reagents to the eluted ligated products from step 3.3. Incubate it at 37°C for 20 minutes.

| | |
|---|------|
| 10XMMLV Reverse Transcription Buffer | 4 µL |
| 10mM dNTP (dATP, dCTP, dTTP and dGTP) | 2 µL |
| 100mM DTT | 2 µL |
| 20µM RT Primer | 1 µL |
| high performance MMLV Reverse Transcriptase (Epicentre) | 1 µL |

4.2 Terminate the reaction by heating the tube at 85°C for 15 minutes. Keep it on ice. It can be stored at -80 °C.

5. PCR amplification

5.1 For one PCR reaction, add the following reagents into a 200 µL PCR tube.

| | |
|--|--------|
| RT product from step 4.2 | 4 µL |
| Nuclease-Free Water | 9.3 µL |
| 10mM dNTPs | 0.5 µL |
| 5X high fidelity Phusion buffer | 4 µL |
| 10µM Illumina RP-1 primer | 1 µL |
| 10µM Illumina index primer | 1 µL |
| Phusion DNA polymerase 2U/µL (thermal scientific, F-530) | 0.2 µL |

5.2 PCR conditions (12 cycles):

30 seconds at 98°C; 11 cycles of: 10 seconds at 98°C, 30 seconds at 60°C, 15 seconds at 72°C; 10mins at 72°C. Hold at 4°C.

It is important to run the reactions at different cycles to find out the best cycle for each sample.

This step concludes the first day and the PCR products can be stored at -20°C.

6. Extraction of the PCR products from PAGE
- 6.1 Separate the PCR products on native 8% polyacrylamide gels, while load a 20 bp DNA ladder (Jena Bioscience) as the size marker.
- 6.2 Stain the gel with SYBR GOLD (Invitrogen).
- 6.3 Cut out the band at sizes around 145-150 bp (cDNA library of interest) from the gel and place the gel slice into a gel breaker tube and slash it by centrifugation at top speed for 5mins.
- 6.4 Elute the cDNA library in 1xNEB restriction enzyme buffer 2 by shaking at room temperature for more than 2hrs or overnight at 4 °C.
- 6.5 Remove the gel debris using a Spin-X column (Corning, 0.45 µm).
- 6.6 Concentrate the eluted cDNA library with ethanol precipitation. For every 300 µL elute, add 2 µL of 5 mg/mL glycoBlue (Ambion), 30 µL 3M NaOAc pH5.0, and 975 µL ethanol. Incubate the mixtures at -80°C for 2-4 hrs followed by centrifugation at 20,000 x g for 20 min at 4 °C.
- 6.7 Wash the pellet with 80% of ethanol, air-dry it for 5-8 min and re-suspend in 12 µL nuclease free water.
- 6.8 Determine the concentration of the library by Nanodrop, and send it for sequencing.

References

- [1] Bushati N. and Cohen S.M., *MicroRNA functions*, *Annu Rev Cell Dev Biol*, 2007, 23, 175-205.
- [2] Mallory A.C. and Vaucheret H., *Functions of microRNAs and related small RNAs in plants*, *Nat Genet*, 2006, 38, S31-S36.
- [3] Aravin A. and Tuschl T., *Identification and characterization of small RNAs involved in RNA silencing*, *FEBS Letters*, 2005, 579, 5830-5840.
- [4] Baran-Gale J., Erdos M.R., Sison C., Young A., Fannin E.E., Chines P.S. and Sethupathy P., *Massively differential bias between two widely used Illumina library preparation methods for small RNA sequencing*, *bioRxiv*, 2013, <http://dx.doi.org/10.1101/001479>.
- [5] Git A., Dvinge H., Salmon-Divon M., Osborne M., Kutter C., Hadfield J., Bertone P. and Caldas C., *Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression*, *RNA*, 2010, 16, 991-1006.
- [6] Reddy A.M., Zheng Y., Jagadeeswaran G., Macmil S.L., Graham W.B., Roe B.A., Desilva U., Zhang W. and Sunkar R., *Cloning, characterization and expression analysis of porcine microRNAs*, *BMC Genomics*, 2009, 10, 65.
- [7] Szittyta G., Moxon S., Pantaleo V., Toth G., Pilcher R.L.R., Moulton V., Burgyan J. and Dalmay T., *Structural and functional analysis of viral siRNAs*, *PLoS Pathog*, 2010, 6, e1000838.
- [8] Tian G., Yin X., Luo H., Xu X., Bolund L. and Zhang X., *Sequencing bias: comparison of different protocols of microRNA library construction*, *BMC Biotechnol*, 2010, 10, 64.
- [9] Willenbrock H., Salomon J., Søkilde R., Barken K.B., Hansen T.N., Nielsen F.C., Møller S. and Litman T., *Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing*, *RNA*, 2009, 15, 2028-2034.
- [10] Hafner M., Renwick N., Brown M., Mihailović A., Holoch D., Lin C., Pena J.T., Nusbaum J.D., Morozov P. and Ludwig J., *RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries*, *RNA*, 2011, 17, 1697-1712.
- [11] Jayaprakash A.D., Jabado O., Brown B.D. and Sachidanandam R., *Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing*, *Nucleic Acids Res*, 2011, 39, e141-e141.
- [12] Sorefan K., Pais H., Hall A.E., Kozomara A., Griffiths-Jones S., Moulton V. and Dalmay T., *Reducing ligation bias of small RNAs in libraries for next generation sequencing*, *Silence*, 2012, 3, 1-11.
- [13] Zhuang F., Fuchs R.T., Sun Z., Zheng Y. and Robb G.B., *Structural bias in T4 RNA ligase-mediated 3'-adapter ligation*, *Nucleic Acids Res*, 2012, 40, e54-e54.
- [14] Viollet S., Fuchs R.T., Munafo D.B., Zhuang F. and Robb G.B., *T4 RNA Ligase 2 truncated active site mutants: improved tools for RNA analysis*, *BMC Biotechnol*, 2011, 11, 72.
- [15] Sun G., Wu X., Wang J., Li H., Li X., Gao H., Rossi J. and Yen Y., *A bias-reducing strategy in profiling small RNAs using Solexa*, *RNA*, 2011, 17, 2256-2262.
- [16] He C.-Y., Cui K., Zhang J.-G., Duan A.-G. and Zeng Y.-F., *Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo*, *BMC Plant Biol*, 2013, 13, 119.
- [17] Zhang Z., Lee J.E., Riemondy K., Anderson E.M. and Yi R., *High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing*, *Genome Biol*, 2013, 14, R109.
- [18] Pease J., *Small-RNA sequencing libraries with greatly reduced adaptor-dimer background*, *Nat Methods*, 2011, 8, iii-iv.
- [19] Vaidyanathan R., Kuersten S. and Doyle K., *Methods and kits for 3'-end-tagging of RNA*, *US Patent 20110104785 A1*, 2011.
- [20] Xu P., Mohorianu I., Yang L., Zhao H., Gao Z. and Dalmay T., *Small RNA profile in Moso bamboo root and leaf obtained by high definition adapters*, *PLoS one*, 2014, 9, e103590.
- [21] Stocks M.B., Moxon S., Mapleson D., Woolfenden H.C., Mohorianu I., Folkes L., Schwach F., Dalmay T. and Moulton V., *The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets*, *Bioinformatics*, 2012, 28, 2059-2061.
- [22] Prüfer K., Stenzel U., Dannemann M., Green R.E., Lachmann M. and Kelso J., *PatMaN: rapid alignment of short sequences to large databases*, *Bioinformatics*, 2008, 24, 1530-1531.
- [23] Mortazavi A., Williams B.A., McCue K., Schaeffer L. and Wold B., *Mapping and quantifying mammalian transcriptomes by RNA-Seq*, *Nat Methods*, 2008, 5, 621-628.
- [24] Lorenz R., Bernhart S.H., Zu Siederdisen C.H., Tafer H., Flamm C., Stadler P.F. and Hofacker I.L., *ViennaRNA Package 2.0*, *Algorithms for Molecular Biology*, 2011, 6, 26.
- [25] Edgar R., Domrachev M. and Lash A.E., *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*, *Nucleic Acids Res*, 2002, 30, 207-210.

- [26] Jagadeeswaran G., Zheng Y., Li Y.F., Shukla L.I., Matts J., Hoyt P., Macmil S.L., Wiley G.B., Roe B.A. and Zhang W., Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families, *New Phytol*, 2009, 184, 85-98.
- [27] Szittyá G., Moxon S., Santos D.M., Jing R., Fevereiro M.P., Moulton V. and Dalmay T., High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families, *BMC Genomics*, 2008, 9, 593.
- [28] Raabe C.A., Tang T.-H., Brosius J. and Rozhdestvensky T.S., Biases in small RNA deep sequencing data, *Nucleic Acids Res*, 2014, 42, 1414-1426.
- [29] Kawano M., Kawazu C., Lizio M., Kawaji H., Carninci P., Suzuki H. and Hayashizaki Y., Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing, *BioTechniques*, 2010, 49, 751-755.
- [30] Vigneault F., Ter-Ovanesyan D., Alon S., Eminaga S., C Christodoulou D., Seidman J., Eisenberg E. and M Church G., High-Throughput Multiplex Sequencing of miRNA, *Curr Protoc Hum Genet*, 2012, 11.12. 11-11.12. 10.
- [31] Song Y., Liu K.J. and Wang T.-H., Elimination of ligation dependent artifacts in T4 RNA Ligase to achieve high efficiency and low bias microRNA capture, *PLoS One*, 2014, 9, e94619.
- [32] Munafó D.B. and Robb G.B., Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA, *RNA*, 2010, 16, 2537-2552.

Supplemental Material: The online version of this article

(DOI: 10.1515/mngs-2015-0001) offers supplementary material.