Published in final edited form as:

Methods Mol Biol. 2021 ; 2200: 225–254. doi:10.1007/978-1-0716-0880-7_11.

Identification and Quantification of Small RNAs

Di Sun^{1,2,3}, Zeyang Ma^{1,2}, Jiaying Zhu^{1,2}, Xiuren Zhang^{4,5}

¹Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, USA

²Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX, USA

³Graduate Program for Molecular and Environmental Plant Science, Texas A&M University, College Station, TX, USA

⁴Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, USA

⁵Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX, USA

Abstract

RNA silencing plays a critical role in diverse biological processes in plants including growth, development, and responses to abiotic and biotic stresses. RNA silencing is guided by small non-coding RNAs (sRNAs) with the length of 21–24 nucleotides (nt) that are loaded into Argonaute (AGO) to repress expression of target loci and transcripts through transcriptional or posttranscriptional gene silencing mechanisms. Identification and quantitative characterization of sRNAs are crucial steps toward appreciation of their functions in biology. Here, we developed a step-by-step protocol to precisely illustrate the process of cloning of sRNA libraries and correspondingly computational analysis of the recovered sRNAs. This protocol can be used in all kinds of organisms, including Arabidopsis, and is compatible with various high-throughput sequence technologies such as Illumina Hiseq. Thus, we wish that this protocol represents an accurate way to identify and quantify sRNAs in vivo.

Keywords

Small RNA; RNA silencing; Library construction; High-throughput sequencing; Computational analysis

1 Introduction

RNA silencing is a fundamental mechanism for regulating gene expression in diverse biological contexts in eukaryotic organisms. RNA silencing is implemented through a ribonucleoprotein complex, also known as RNA-inducing silencing complexes (RISC) that is composed of sRNAs and AGOs proteins. sRNAs can be separated into different classes depending on their originalities and biogenesis processes. Among sRNAs, microRNAs (miRNAs) and trans-acting small interfering RNAs (ta-siRNAs) are typically 21 nt long and loaded into AGO proteins to cleave complementary transcripts and/or inhibit their

translation in cytoplasm [1]. Some species of sRNAs, typically 24 nt long, guide AGOs to execute transcriptional gene silencing in nucleus [1].

Understanding of function and mechanisms of sRNAs entails precise identification and quantification of sRNA populations in vivo. In earlier studies in the sRNA field, computational prediction [2], sRNA blot [3], and quantitative RT-PCR [4, 5] are prevailing techniques to identify sRNAs and detect their expression. These methods are easy and still widely used in many laboratories that are modern-resource inaccessible. However, the limitations of these approaches are the relatively lower recovery of sRNA species and that they restricted to only small numbers of sRNAs [6]. Next-generation sequencing (NGS) has been developed as a powerful technique to detect and quantify sRNAs [7–10]. NGS technique has significant advantages in systematically recovering sRNAs with high efficiency, especially the ones with low abundance, or the ones that are expressed in specific tissue niches, developmental stages, and physiological processes. Numerous commercial kits exemplified by the ones provided by Illumina and NEB vendors are at hand for this purpose and wide arrays of NGS protocols have also been customized/published. All the methods are based on the fact that most classes of sRNAs harbor 5' phosphate and 3' hydroxyl termini that result from RNA processing. This feature distinguishes *bona fide* sRNAs from RNA turnover products and RNase degradation products that rather contain a 5' hydroxyl and a 2' or 3' phosphate group. Thus, the *bona fide* sRNAs are easily captured through ligations with a pair of oligo adapters that provide 5' phosphate and 3' hydroxyl groups, respectively. One problem to clone sRNA library is circularization of 5' phosphate/3' hydroxyl sRNAs and even adapters themselves during the adapter ligation process. To prevent this issue, chemically pre-adenylated 3' adapter deoxyoligonucleotides, which are also blocked at their 3' ends through an amino group or a dideoxy nucleotide, are widely used to avoid their circularization [7]. The use of 5'-pre-adenylated adapters eliminates the need for ATP during ligation, and thus solves the problem of adenylation of the pool RNA 5' phosphate that causes circularization. Another strategy to address circularization is to use a truncated and mutant form of T4 RNA ligase 2, Rnl2(1–249)K227Q. This T4 RNA ligase mutant is able to prevent adenylyl transfer from the 5' phosphate of 3' adapter to the 5' phosphate of the sRNA pool and subsequent pool RNA circularization [11].

To simplify the cloning process and minimize the workload for sRNA library construction and sequencing, essentially all commercial kits and a majority of the customized NGS methods do not separate sRNAs from other RNA species in total RNA through a size-fractionation process in a polyacrylamide gel. Instead, total RNA is initially used through the adapter ligation processes and final separation of potential sRNA-containing cDNA fragments is conducted according to their estimated sizes in an Agarose gel. However, the pre-adenylated 3' adapter provided in most of the kits has very low concentration (i.e., 5 µM in NEBNext® Multiplex Small RNA Library Prep Set for Illumina®), likely due to the relative technique difficulty in chemical synthesis. On the other hand, many species of RNAs including the abundant rRNAs and mRNAs in total RNA contain 5' phosphate and 3' hydroxyl termini and could easily compete out sRNAs for the pre-adenylated 3' adapter, leading to bias toward the recovery of the most abundant sRNAs in the sRNA library construction. Another issue in the NGS including sRNA-seq that prevents the precise quantification is that the ligation of adapters might have bias toward certain RNAs probably

due to inherent RNA secondary structure or chemical reaction preference as previously reported [12, 13].

Based on the previous methods [7, 11] and our extensive experience in sRNA studies [14–19], we have optimized our protocol (Fig. 1). Compared to commercial kits and most published protocols, ours has several unique features: (1) It provides friendly applied and cost-effective technique to synthesize a large amount of 3' pre-adenylated adapters so that sufficient amount of adapters can be used to saturate all species of sRNAs, allowing recovery of low abundant sRNAs; (2) it uses spiked ^{32}P labeled positive controls to separate sRNAs so that we could track the entire process of sRNA libraries, allowing precisely identifying the *bona fide* sRNAs. Alternative method of separation of sRNAs is also provided for the radioisotope-free laboratories; (3) we have now introduced degenerated adapters (3' pre-adenylated oligos) to minimize the bias in the cloning and sequencing steps; (4) finally, we provide homegenerated scripts to quantify the sRNA reads. Our protocol is compatible with the prevailing sequencing platforms in the market. We hope that our protocol can advance plant research community to clone and sequence sRNAs in a cost-effective and more precise way.

2 Materials

2.1 Synthesis of the 5' Pre-adenylated 3' Adapter Oligonucleotides

2.1.1 Synthesis of the Adenosine-5' - phosphoimidazolidine (ImpA)

1. Flasks and beakers.
2. Balloon.
3. Nitrogen.
4. Syringes and needles.
5. Rubber septum.
6. Magnetic stirrer and magnetic bars.
7. Corex tubes.
8. Lyophilizer.
9. Centrifuge.
10. 5' AMP free acid.
11. Dimethylformamide (DMF).
12. Triphenylphosphine.
13. 2,2' - Dipyridyldisulfide.
14. Imidazole (molecular biology grade).
15. Sodium perchlorate.
16. Triethylamine.

17. Acetone.
18. Diethyl ether.

2.1.2 Quality Control of ImpA

1. UV lamp and silica gel plates.
2. 20 cm × 20 cm TLC plate (Silicycle).
3. DMF.
4. Ammonium hydroxide solution.
5. Iso-propanol.
6. Autoclaved double-distilled water.

2.1.3 Adenylation of 3' Adapter Oligonucleotides

1. SpeedVac.
2. Glasses, combs, and clips used for gel preparation.
3. Gel electrophoresis tank and power device.
4. 1.5 mL Eppendorf tubes and all size tips.
5. Plastic film.
6. UV lamp and silica gel plates.
7. Blades.
8. Nano drop.
9. 50 mM MgCl₂ dissolved in sterilized deionized water.
10. 5' Phosphorylated 3' NH₂-linker modified 3' adapter (Table 1) oligonucleotide dry powder.
11. Autoclaved double-distilled water.
12. Diethylpyrocarbonate (DEPC).
13. RNase-free water (1 L water mix well with 1 mL DEPC. Incubate the mix at room temperature overnight in a hood. Autoclave treated water at 121 °C for 20 min to destroy DEPC).
14. Urea-polyacrylamide gel reagents: UreaGel concentrate solution (National diagnostics), UreaGel diluent solution (National diagnostics), UreaGel buffer (National diagnostics), TEMED, 10% ammonium persulfate (APS) (10 g in 100 mL water) (Table 2).
15. 1× TBE buffer diluted from 10× TBE buffer: 1 M Tris base, 1 M boric acid, 0.02 M EDTA (disodium salt) prepared with RNase-free water.
16. Loading buffer: 8 M urea, 50 mM EDTA pH 8.0, 0.5 mg/mL bromophenol blue.

17. Xylene cyanol.
18. 0.4 M NaCl prepared with RNase-free water.
19. Ethanol.

2.2 Primer Purification

1. Glasses, combs, and clips used for gel preparation.
2. Gel electrophoresis tank and power device.
3. Phosphor screen and Typhoon biomolecular imager.
4. Heat block.
5. Incubator.
6. Syringes and needles.
7. Blades.
8. Tumbling machine.
9. 1.5 mL Eppendorf tubes and all size tips.
10. Centrifuge.
11. Kim wipe.
12. UV lamp and silica gel plates.
13. Nano drop.
14. Primers: these reagents can be customized (Table 1). Noticed that two popular sRNA library preparation kits (Illumina® TruSeq® Small RNA Library Prep Kit and NEBNext® Multiplex Small RNA Library Prep Set for Illumina®) have different primer sequences but share the same barcodes and 5' adapter sequence.
15. 1× TBE buffer.
16. FA buffer: 95% deionized formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, 5 mM EDTA, 0.025% SDS [20].
17. Urea-polyacrylamide gel reagents.
18. 3 M NaAc prepared with RNase-free water.
19. Autoclaved double-distilled water.
20. RNase-free water.
21. Glycogen.
22. 0.4 M NaCl prepared with RNase-free water.

2.3 Marker Labeling

2.3.1 CIP Treatment of RNA Oligomers with 5' Mono- or Tri-phosphate

1. Markers (Table 1).

2. 1.5 mL Eppendorf tubes and all size tips.
3. Centrifuge.
4. Incubator.
5. CIP (together with Buffer).
6. RNase inhibitor (RNase OUT).
7. RNase-free water.
8. Phenol:chloroform:isoamyl alcohol (25:24:1), pH 4.3.
9. Chloroform:isoamyl alcohol (24:1).
10. Ethanol.
11. 3 M NaAc prepared with RNase-free water.
12. Glycogen.

2.3.2 T4 PNK Treatment

1. Glasses, combs, and clips used for gel preparation.
2. Gel electrophoresis tank and power device.
3. Phosphor screen and Typhoon biomolecular imager.
4. Heat block.
5. Incubator.
6. Blades.
7. 1.5 mL Eppendorf tubes and all size tips (with and without barriers).
8. Centrifuge.
9. Eppendorf Thermomixer.
10. Needles and syringes.
11. Phosphor screen.
12. Kim wipe.
13. Geiger counter.
14. Markers (Table 1).
15. RNase-free water.
16. T4 PNK (together with PNK buffer).
17. [γ -³²P] ATP (3000 Ci/mM).
18. RNase inhibitor (RNase OUT).
19. Phenol:chloroform:isoamyl alcohol (25:24:1), pH 4.3.

20. Ethanol.
21. 3 M NaAc prepared with RNase-free water.
22. Glycogen.
23. 1× TBE buffer.
24. FA buffer.
25. Urea-polyacrylamide gel reagents.
26. 0.4 M NaCl prepared with RNase-free water.
27. RNA dissolve buffer: 100 mM KCl, 30 mM Tris-HCl, pH 7.5 in RNase-free water.

2.4 RNA Extraction

1. Mortar and pestle.
2. Lab spoons.
3. Scale.
4. Liquid nitrogen.
5. Heat blocks.
6. Centrifuge.
7. Gel electrophoresis casts, trays, combs, tank, and power device.
8. 1.5 mL Eppendorf tubes and all size tips.
9. Vortex.
10. Trizol.
11. Phenol:chloroform:isoamyl alcohol (25:24:1), pH 4.3.
12. Chloroform.
13. Ethanol.
14. RNase-free water.
15. Autoclaved double-distilled water.
16. 75% Ethanol prepared with RNase-free water.
17. FA buffer.
18. Agarose.
19. Ethidium bromide (EB).

2.5 Preparation of RNA with Radiolabeled Marker Cocktail and Size Selection

1. The same materials as described in Subheading 2.3.2, except T4 PNK and [γ -³²P] ATP.

2. Labeled markers described in Subheading 3.3.2.
3. RNA samples as described in Subheading 3.4.

2.6 3' Ligation

1. The same materials as described in Subheading 2.3.2, except T4 PNK and [γ - 32 P] ATP.
2. 3' Adapter synthesized described in Subheading 3.1.
3. Rnl2 K227Q (together with 3' ligation buffer).
4. 10× BSA (Diluted 10 times with 100× Ac-BSA).
5. 50% DMSO prepared in RNase-free water.

2.7 5' Ligation

1. The same materials as described in Subheading 2.6, except 3' adapter and Rnl2 K227Q.
2. 5' Adapter (Table 1).
3. Rnl1 (together with 5' ligation buffer).
4. 10 mM ATP prepared with RNase-free water.
5. SR-RT primer purified in Subheading 3.2.

2.8 Reverse Transcription

1. PCR machine.
2. PCR tubes and all size tips.
3. Hydriion S/R Dispenser pH papers (pH 1.0–11.0).
4. RNase-free water.
5. Autoclaved double-distilled water.
6. 10 mM dNTP mix prepared with RNase-free water.
7. 0.1 M DTT prepared with RNase-free water.
8. SSII (together with Reverse transcription kit).
9. RNase inhibitor (RNase OUT).
10. 150 mM KOH/20 mM Tris-Base.
11. 150 mM HCl.

2.9 PCR

1. Gel electrophoresis casts, trays, combs, tank, and power device.
2. PCR tubes and all size tips.

3. Autoclaved double-distilled water.
4. 2 mM dNTPs.
5. SR primer and 3' primers purified from Subheading 3.2 (Table 1).
6. cDNA synthesized from Subheading 3.8.
7. KOD hot start DNA polymerase (together with kit).
8. Agarose.
9. 0.5× TBE buffer.
10. Ethidium bromide (EB).
11. 1 kb DNA ladder.
12. Phenol:chloroform:isoamyl alcohol (25:24:1) pH 8.0.
13. Chloroform.
14. 3 M NaAc.
15. Ethanol.

2.10 Pme I Digestion

1. Gel electrophoresis casts, trays, combs, tank, and power device.
2. Incubator.
3. Nano drop.
4. 1.5 mL Eppendorf tubes and all size tips.
5. Autoclaved double-distilled water.
6. Blades.
7. Kim wipe.
8. QiaQuick gel extraction kit.
9. Pme I (together with buffer).
10. 0.5× TBE buffer.
11. Ethidium bromide (EB).
12. 1 kb DNA ladder.
13. Low-melting agarose and regular agarose.
14. 10 mM Tris-HCl pH 8.0.

2.11 Computing Analysis Workflow

Software listed in Table 3.

3 Methods

3.1 Synthesis of the 5' Pre-adenylated 3' Adapter Oligonucleotides

This section is to provide a step-by-step protocol to synthesize a large amount of 5' pre-adenylated 3 adapters for the 3' end ligation of sRNAs. The protocol is developed from an early published method [11] with numerous modifications for simplification and friendly usage for a common laboratory.

3.1.1 Synthesis of the Adenosine-5' - Phosphoimidazolid (ImpA)—All steps should be performed under a fume hood.

1. Prepare two 25 mL clean and dry round-bottom glass flasks with two rims and one stirring bar each inside. Flasks are dried in a drying oven at 140 °C overnight.
2. Use a needle and a syringe to fetch 5 mL anhydrous DMF from a sealed bottle to a clean flask (flask A); and then resuspend 174 mg (0.5 mM) 5'-AMP free acid into the anhydrous DMF in flask A. The flask should be kept closed with a rubber septum. The AMP will not dissolve completely. Label this as solution A. Take a 20 μ L aliquot for quality control step.
3. Prepare 262 mg (1 mM) triphenylphosphine, 220 mg (1 mM) 2,2'-dipyridyldisulfide, and 170 mg (2.5 mM) imidazole in the other flask (flask B) with 10 mL DMF and 0.9 mL (2.5 mM) triethylamine. Stir vigorously. The solution shows a yellow-green color. Keep the flask closed with a rubber septum. Label this as solution B. Take a 20 μ L aliquot for quality control step.
4. Connect a nitrogen-filled balloon with the flask of solution B by a needle inserted through the rubber septum (Fig. 2a). A second needle is inserted into the septum to replace the humid air in flask B with nitrogen. Remove the second needle once the switch finishes (this process takes about 1 min).
5. Use a needle and a syringe to fetch all the solution A and insert the needle through the rubber septum of flask B. Add solution A dropwise to flask B. Remove the needle and stir the mixture to react for 1.5 h at room temperature; keep the flask connected with nitrogen balloon through the reaction (Fig. 2b).
6. Add 1.1 g (9 mM) sodium perchlorate, 110 mL acetone, and 55 mL anhydrous diethyl ether to a clean 500 mL beaker. Keep the beaker on ice and stir vigorously. Add the reaction mixture above dropwise to this beaker. As the yellow-green AMP solution is added, you will expect to observe the mixture turns cloudier.
7. After adding all the reaction solution, stop the stirrer. Keep the beaker on ice for 20 min. The cloudy precipitation starts to sink to the bottom. Decant as much as possible supernatant out.
8. Resuspend the pellet with the residual solution and transfer the mixture into two 50 mL Cortex tubes evenly. Centrifuge at $4470 \times g$ for 2 min. Wash pellet with

20 mL acetone and centrifuge again. Pour off the supernatant. Repeat the washing step until the supernatant is colorless clear. The pellet should be white.

9. Resuspend the pellet in 10 mL diethyl ether and centrifuge at $4470 \times g$ for 2 min. Pour off the supernatant, and use a lyophilizer or a vacuum oven at 40 °C to dry the pellet overnight. After dry, the ImpA powder can be stored in a sealed bottle at –80 °C up to 1 week.

3.1.2 Quality Control of ImpA

1. Dissolve 1 mg ImpA in 20 µL DMF. Draw a starting line on a TLC plate. Spot 1 µL solution A, solution B, and ImpA/DMF solution, respectively, on the starting line.
2. Use iso-propanol:water:25% ammonia (7:2:1) as solvent system and run the TLC in a chromatography chamber. Take out the plate when the solvent front is about 0.3 cm away from the top of the plate. Label the solvent front with a pencil.
3. Dry the plate and visualize the spots under 254 nm UV lamp. The material on the starting line should completely disappear, indicating the successful and efficient production of ImpA (Fig. 2c).

3.1.3 Adenylation of 3' Adapter Oligonucleotides (See Note 1)

1. Prepare 150 µL solution containing 100 mM ImpA (MW = 396.3) and 50 mM MgCl₂.
2. Add 100 µL ImpA/MgCl₂ solution into 50 nM 5' phosphorylated 3' NH₂-linker modified 3' adapter (Table 1) oligonucleotide dry powder (*see* Note 2). Incubate the solution at 50 °C for 1.5 h. Add the remaining 50 µL ImpA/MgCl₂ solution and incubate another 1.5 h.
3. After reaction, reduce the volume to about 75 µL in a SpeedVac. Then add one volume (75 µL) loading buffer (8 M urea, 50 mM EDTA pH 8.0, 0.5 mg/mL bromophenol blue).
4. Prepare 150 mL 20% urea-polyacrylamide gel (Table 2) with the size of 33 cm × 40 cm × 0.25 mm containing only one well or two wells. Load the sample in one well and load bromophenol blue and xylene cyanol loading buffer in the other if two wells are used. Run electrophoresis with 1× TBE buffer for 3.5 h at 3500 mV until xylene cyanol is about to exit the bottom of the gel.
5. Dismantle the gel and wrap it with plastic film. Place it on a fluorescence-indicator-coated silica gel plate and visualize the band with a 254 nm UV lamp. The intensity of two bands is typically similar, indicative of the pre-adenylation of approximately half amount of the initial 5' phosphorylated 3' NH₂-linker modified 3' adapter oligonucleotide (Fig. 2d).

¹This protocol is designed for 50 nM adapter oligos. Corresponding recalculation is needed for different amount of adapter oligos.

²The 3' end modification of the 3' adapter used here is a NH₂. A dideoxy nucleotide (ddC) is also a good alternative and available in numerous vendors.

6. Mark the upper bands and excise the product band. Use a clean blade to cut the gel along the sketch. Dice the gel into small pieces and carefully transfer them to a 1.5 mL tube with 350 μ L 0.4 M NaCl.
7. Tumble in the Eppendorf thermomixer at a speed of 1200 rpm/min at 4 °C overnight to elute the adapter.
8. Centrifuge at 12,000 *g* for 5 min at 4 °C. Take supernatant out to a new tube. Add 100 μ L 0.4 M NaCl to each old tube and tumble 30 min at 4 °C.
9. Centrifuge at 12,000 *g* for 5 min at 4 °C. Mix supernatant to previous one of each sample. Add 1 mL ethanol. Mix well and incubate it at –20 °C overnight.
10. Collect the precipitation by centrifuging at 21,130 *g* for 15 min. Wash the pellet with 75% ethanol and centrifuge again at 21,130 *g* for 5 min.
11. Discard the supernatant and air-dry the pellet for 5 min.
12. Dissolve the pellet in 100 μ L RNase-free water. Measure the concentration by a nano drop. Add water to a final concentration of 50 μ M.

3.2 Primer Purification

Commercial oligonucleotides contain accumulated impurities and truncated forms during the synthesis process. The default desalting purification can only remove excess salt but not truncated sequences. While the commercial PAGE purification can be additional expense, here we show how to purify primers through a home-developed PAGE method.

1. Assemble a gel cast. Prepare 10% urea-polyacrylamide gel (Table 2) with the size of 16 cm×16 cm×1.5 mm (The gel volume is about 40 mL/gel).
2. Add water to commercial purchased primers and adapters (*see* Note 3) to a final concentration of 100 μ M. Take 50 μ L and mix very well with 50 μ L FA buffer.
3. Assemble cast with electrophoresis tank. Use clips to fix a piece of aluminum pad with thin glass plate of the cast to share heat when running electrophoresis. Fill in the top and bottom tank with 1× TBE buffer.
4. Use a needle and a syringe to completely clean urea from wells. Use a curved needle and a syringe to remove bubbles at the bottom.
5. Load samples into wells, and split one sample into two wells if necessary. Leave one empty well between different samples. Fill empty wells with FA buffer to make sure the samples run straight.
6. Run electrophoresis at 500 V or 200 mA or 35 W per gel for 1 h.
7. Dismantle the gel cast and gently transfer the gel to plastic wrap. Put the gel on the top of thick glass plate matching the labels. Use a UV lamp (254 nm) to check the signal of DNA. Draw the position along the outline of the bands. Then

³For RNA markers and adapters, and primers that are directly added to RNA, use RNase-free water to dissolve them, and the following solution (e.g., 0.4 M NaCl) should also be RNase-free. Other index primers or SR primer can be dissolved by autoclaved double-distilled water.

use a clean blade to cut the gel along the sketch. Chop that region into small pieces and carefully transfer them to a 1.5 mL tube (*see* Note 4). Use clean water to rinse and kim wipe to clean the blade between samples.

8. Add 350 μ L 0.4 M NaCl to each tube and tumble the tubes in the Eppendorf Thermomixer overnight at 4 °C.
9. Centrifuge at 12,000 *g* for 5 min at 4 °C. Take supernatant out to a new tube. Add 100 μ L 0.4 M NaCl to each old tube and tumble for additional 30 min at 4 °C.
10. Centrifuge at 12,000 *g* for 5 min at 4 °C. Take the supernatant out to the new tube and add 1 mL ethanol, mix well, and incubate at –80 °C at least 1 h, or –20 °C for an overnight.
11. Centrifuge at 21,130 *g* for 15 min at 4 °C. Discard supernatant. Use 70% ethanol to wash the pellet. Centrifuge at 21,130 *g* for 5 min. Discard supernatant and spin down again. Discard residual through pipetting and air-dry the pellet in a clear area for 5 min.
12. Add water to a final concentration of 50 μ M.

3.3 Marker Labeling

In our sRNA library preparation, we do size selection prior to ligation. We typically make RNA oligonucleotides through in vitro T7 transcription in our laboratory. However, one could simply purchase commercially synthesized RNA oligonucleotides of certain lengths as markers to spike the sRNAs of interest. The lengths can be optimized according to the size of sRNAs to be detected. Here we use 19 nt and 24 nt RNA oligonucleotides (Table 1) as examples. In order to make them easily detectable, we label the markers with 32 P-ATP and T4 PNK. The RNA markers can also be ligated with adapters together with samples through experiments until their clearance from the samples through Pme I digestion in a final step. Thus, they can also be used to monitor the ligation efficiency and degradation scenario (*see* Note 5).

3.3.1 CIP Treatment of RNA Oligomers with 5' Mono- or Tri-phosphate (See Note 6)

1. Prepare CIP reaction system for each marker on ice as following: 2 μ L marker solution dissolved in RNase-free water (10 μ M), 2 μ L ten-fold NEB Buffer 3, 2 μ L CIP, 1 μ L RNase inhibitor (final concentration 1 U/ μ L), and RNase-free water 13 μ L. Mix well and incubate at 37 °C for 30 min.

⁴. Dicing gel into smaller pieces, the higher recycle efficiency will be obtained. But if the pieces are too small, it might be difficult to pick up. Researchers need to make a balance here.

⁵. For the laboratories which do not have permission for radioisotope usage, initial separation of sRNAs could be achieved through resolving total RNA samples align with 19- and 24-nt oligo markers that are prepared in the same volume and buffer compositions so that sRNAs and markers could migrate at the identical speed in a gel. In this scenario, one could just conduct separation of sRNA cDNAs at the final step as instructed by most of commercial kits, and does not need to monitor the 3' and 5' ligation efficiency step by step.

⁶. If oligo markers are synthesized by companies, the default synthesis is that the 5' end contains a hydroxyl group. Under this substance, no CIP treatment is needed. If the markers contain a 5' (tri-)phosphate group, such as home-transcribed RNA markers, the CIP treatment is needed.

2. Add 220 μL RNase-free water to each reaction mixture above.
3. Add 240 μL of phenol:chloroform:isoamyl alcohol (25:24:1, pH 4.3) to each mixture. Vortex for 30 s to mix well. Centrifuge for 5 min at room temperature by maximum speed.
4. Carefully take 200 μL upper phase liquid to a new labeled tube, respectively. Add 200 μL chloroform:isoamyl alcohol (24:1) to each tube. Vortex for 30 s to mix well. Centrifuge for 5 min at room temperature by maximum speed.
5. Carefully take 170 μL upper phase liquid to a new labeled tube, respectively. Add 510 μL ethanol, 17 μL 3 M NaAc (pH 5.2), and 1 μL glycogen in turn. Mix well and precipitate at $-80\text{ }^{\circ}\text{C}$ for at least 30 min, or $-20\text{ }^{\circ}\text{C}$ overnight.
6. Centrifuge at maximum speed at $4\text{ }^{\circ}\text{C}$ for 15 min. Carefully remove the supernatant by pipetting without touching the pellet.
7. Wash the pellet with 500 μL 75% ethanol. Centrifuge at $13,800\text{ }g$ at $4\text{ }^{\circ}\text{C}$ for 5 min. Discard ethanol. Spin down quickly and carefully remove the residual ethanol through pipetting.
8. Keep the tubes on a neat bench and air-dry for at most 5 min. Then resuspend the pellet with 2.5 μL RNase-free water.

3.3.2 T4 PNK Treatment

1. Prepare PNK reaction system for each marker on ice as follows: 1 μL commercially purchased or home-made marker (Table 2) solution dissolved in RNase-free water (10 μM), 1 μL ten-fold NEB PNK buffer, 5 μL of $\gamma\text{-P}^{32}\text{-ATP}$ (3000 Ci/mM or 10 mCi/mL), 0.5 μL RNase inhibitor (final concentration 1 U/ μL), 1 μL T4 polynucleotide kinase, RNase-free water 1.5 μL . Or add the following reagents to each of the resuspended markers from CIP treatment: 1 μL ten-fold NEB PNK buffer, 5 μL of $\gamma\text{-ATP}^{32}$ (3000 Ci/mM or 10 mCi/mL), 0.5 μL RNase inhibitor (final concentration 1 U/ μL), and 1 μL T4 polynucleotide kinase. Mix well and incubate at $37\text{ }^{\circ}\text{C}$ for 2 h.
2. Add 230 μL RNase-free water to each of the above reaction mixture, respectively. Then recover markers by repeating the **steps 3–7** described in Subheading 3.3.1.
3. Resuspend the pellet with 20 μL FA buffer, respectively.
4. Assemble cast with electrophoresis tank. Use clips to attach a piece of aluminum pad on the surface of the thin glass plate to spread heat when running electrophoresis. Fill in the top and bottom tank with 1X TBE buffer made with RNase-free water.
5. Denature samples at $95\text{ }^{\circ}\text{C}$ for 3 min, then place them on ice for 2 min, followed by a quick spin. Load samples to 15% urea-polyacrylamide gel (Table 2) with the size of $16\text{ cm} \times 16\text{ cm} \times 0.8\text{ mm}$ (the gel volume is about 25 mL/gel).

6. Use a needle and a syringe to clean wells to make sure no urea remains. Use a curved needle and a syringe to remove bubbles at the bottom.
7. Load samples and run electrophoresis at 500 V or 200 mA or 35 W per gel for 1 h.
8. Dismantle the gel cast and leave the gel on the top of thick glass plate matching the labels. Implant three tiny radioactive gel pieces asymmetrically at three of four corners of the gel so one can align the gel to the phosphorimager paper printout at the next step. Radioactive gel pieces can be collected from the gel that was used to purify the size markers after ^{32}P -labeling. Wrap the gel in plastic film and cover it with a phosphor screen and keep them in dark for 30 min.
9. Expose the phosphor screen and print out the image as the original size (Fig. 3a). Align the gel on glass to the paper printout and then use a clean blade to cut the gel framed by ^{32}P -labeled 19 to 24-nt markers. Chop that gel slices into small pieces and carefully transfer them to a 1.5 mL tube with 350 μL 0.4 M NaCl (DEPC treated). Use clean water to rinse and kim wipes to clean the blades between samples.
10. Tumble in the Eppendorf thermomixer at a speed of 1200 rpm/min at 4 $^{\circ}\text{C}$ overnight.
11. Centrifuge at 12,000 g for 15 min at 4 $^{\circ}\text{C}$. Take supernatant out to a new tube. Add 100 μL 0.4 M NaCl (DEPC treated) to each old tube and tumble 30 min at 4 $^{\circ}\text{C}$.
12. Centrifuge at 12,000 g for 5 min at 4 $^{\circ}\text{C}$. Take the supernatants out to the new tubes and add 1 mL ethanol and 1 μL glycogen, mix well and incubate at -80°C for at least 1 h or -20°C overnight.
13. Centrifuge at maximum speed for 15 min at 4 $^{\circ}\text{C}$. Discard supernatant.
14. Wash the pellet with 500 μL 75% ethanol. Then centrifuge at 13800 g for 10 min. Discard ethanol. Spin down quickly and carefully remove the residual ethanol.
15. Keep the tubes on a neat bench and air-dry for at most 5 min. Then resuspend the pellet with RNA dissolve buffer (100 mM KCl, 30 mM Tris-HCl, pH 7.5). Adjust the final cpm to 2000 cpm/ μL .

3.4 RNA Extraction

In order to precisely identify and quantify the *bona fide* sRNAs, high-quality and a reasonable amount of total RNA is required. Our methods are suitable for most plant materials. Based on our experience, 0.1 g well-grinded powder of Arabidopsis seedlings can yield at least 50–75 μg total RNA.

1. Harvest plant tissues at the time according to experiment requirements, wrap in aluminum foil, and immediately freeze them in liquid nitrogen. The samples can be stored at -80°C .

2. Prechill mortar and pestle with liquid nitrogen. Transfer the frozen materials into a prechilled mortar and grind fully with a prechilled pestle. Add proper liquid nitrogen before samples are thawed. Grind samples until the powder feels fine and smooth.
3. Weigh 100 mg powder in a prechilled 1.5 mL Eppendorf tube. Add 1 mL Trizol and vortex for 30 s. Incubate the sample in room temperature for 2 min. Centrifuge at 21,130 *g* for 10 min at 4 °C.
4. Transfer 1 mL supernatant into a new 1.5 mL tube. Add 0.2 mL chloroform and vortex for 30 s. Centrifuge at 12,000 *g* for 10 min at 4 °C.
5. Carefully transfer aqueous phase without middle layer into a new 1.5 mL tube (*see* Note 7). Add 500 μ L volume of phenol: chloroform:isoamyl alcohol (25:24:1, pH 4.3) into this tube and vortex for 30 s. Centrifuge at 12,000 *g* for 10 min at 4 °C.
6. Carefully transfer aqueous phase without middle layer into a new 1.5 mL tube (*see* Note 7). Add same volume of isopropanol. Mix them well gently by turn tubes upside down for several times. Precipitate 10 min in room temperature. Centrifuge at 12,000 *g* for 10 min at 4 °C.
7. Take supernatant out without touching RNA pellet, use 75% ethanol to wash the pellet once. Centrifuge at 12,000 *g* for 2 min at 4 °C.
8. Take supernatant out fully without touching RNA pellet. Do another round of quick spinning and remove the residual supernatant by pipetting without disturbing of the pellet. Dry RNA pellet 5 min in room temperature in a clean hood (*see* Note 8).
9. Completely dissolve RNA with 30–50 μ L RNase-free water. Store at –80 °C.
10. Use a nanodrop to measure concentration and quality. Take 1 μ g total RNA and mix with 5 μ L FA buffer. Boil the mixture at 95 °C for 3 min and put it on ice for 2 min. Run 1% Agarose gel to check quality. Normalize RNA from different samples to the same concentration (*See* Note 9).

3.5 Preparation of RNA with Radiolabeled Marker Cocktail and Size Selection

1. Prepare a 15% urea-polyacrylamide gel (Table 2) with the size of 16 cm×16 cm×1.5 mm (The gel volume is about 40 mL/gel).
2. Take 1–20 μ g RNA of each sample, dilute it to 20 μ L, add 21 μ L FA buffer and 1 μ L labeled 19/24 nt marker. Mix well.
3. Denature at 95 °C for 2 min and cold down on ice for 2 min.

⁷.To avoid obtaining lower phase, you are suggested to use a 200 μ L pipette to decant upper phase several times.

⁸.Do not over-dry RNA pellet, which might cause difficulties in RNA dissolvment and RNA degradation.

⁹.The absorbance ratio of RNA with good quality at 260 and 280 nm (A₂₆₀/A₂₈₀) should be between 1.8 and 2.0. Gel electrophoresis result should display a 28S/18S ratio as 2/1.

4. Assemble cast with electrophoresis tank. Use clips to fix a piece of aluminum pad with thin glass plate of the cast to share heat when running electrophoresis. Fill in the top and bottom tank with 1× TBE buffer made with RNase-free water.
5. Use a needle and a syringe to clean wells to make sure no urea remains. Use a curved needle and a syringe to remove bubbles at the bottom.
6. Load samples and run electrophoresis at 500 V or 200 mA or 30 W per gel for 1 h.
7. Dismantle the gel cast and leave the gel on the top of thick glass plate matching the labels. Wrap the gel in plastic film. Mark the gel using small radioisotope chunks as described above. Cover it with a phosphor screen and keep them in dark for 30 min to 1 h.
8. Expose the phosphor screen and print out the image as the original size (Fig. 3b). Align the gel on glass to the figure. Draw the position along the outline of the bands. Then use a clean blade to cut the gel along the sketch. Chop that region into small pieces and carefully transfer them to a 1.5 mL tube with 350 μ L 0.4 M NaCl. Use clean water to rinse and kim wipe to clean the blades between samples.
9. Tumble in the Eppendorf thermomixer at a speed of 1200 rpm /min at 4 °C overnight.
10. Centrifuge at 12,000 *g* for 5 min at 4 °C. Take supernatant out to a new tube. Add 100 μ L 0.4 M NaCl to each old tube and tumble 30 min at 4 °C.
11. Centrifuge at 12,000 *g* for 5 min at 4 °C. Mix supernatant to previous one of each sample. Add 1 mL ethanol and 1 μ L glycogen, mix well, and incubate at –20 °C overnight.

3.6 3' Ligation

1. Centrifuge at 21,130 *g* for 30 min at 4 °C. Decant supernatant and spin down again. Discard residual and air-dry pellet in a clear area for 5 min.
2. Add 7 μ L RNase-free water to each sample to dissolve RNA. Prepare a mix for 3' ligation system: 2.5 μ L 50 μ M 3' adapter, 2 μ L 10× T4 3' ligase buffer (no ATP), 1 μ L 10 × BSA, and 6 μ L 50% DMSO. The total volume is 11.5 μ L. Multiply the volume for one reaction by 1.1-fold of sample numbers due to pipetting variations.
3. Mix well. Incubate the reaction mixture at 95 °C for 1 min to denature the RNA and immediately place on ice for 2 min.
4. Add 1.5 μ L Rnl2 K227Q to each sample. Tap tubes to mix well. Centrifuge at 5000 *g* for 1 min at 4 °C. Keep samples on ice in 4 °C overnight.
5. Add 20 μ L FA buffer to each sample. Run gel electrophoresis and recover RNA as mentioned in Subheading 3.5. The image is shown in Fig. 3c.

3.7 5' Ligation

1. Centrifuge at 21,130 *g* for 30 min at 4 °C. Discard supernatant and spin down again. Discard residual and air-dry pellet in a clear area for 5 min.
2. Add 6.1 µL RNase-free water to each sample to dissolve RNA. Prepare a mix for 5' ligation system: 1 µL 100 µM 5' adapter (Table 2), 2 µL 10 × RNA ligation buffer, 0.4 µL 10 mM ATP, 2 µL 10× BSA, 6 µL 50% DMSO. The total volume is 11.4 µL. Multiply the volume for one reaction by 1.1-fold of sample numbers due to pipetting variations.
3. Mix well. Incubate the reaction mixture at 95 °C for 1 min to denature the RNA and immediately place on ice for 2 min.
4. Add 1 µL RNase Inhibitor, 1.5 µL Rn11. Tap to mix well. Incubate samples at 37 °C for 3 h.
5. Add 20 µL FA buffer to each sample. Run gel electrophoresis and recover RNA as described in Sect. 3.5 but add 1 µL 50 µM SR-RT primer to each sample together with 350 µL NaCl (DEPC treated). The image is shown in Fig. 3d.

3.8 Reverse Transcription

1. Centrifuge at 21,130 *g* for 30 min at 4 °C. Discard supernatant and spin down again. Discard residual and air-dry pellet in a clear area for 5 min.
2. Add 10.8 µL RNase-free water and 1.2 µL dNTP mix (10 mM each) to each sample to dissolve RNA. Prepare a mix containing 2 µL 0.1 M DTT and 5× first strand buffer 4 µL for each sample. The total volume is 6 µL. Multiply the volume for one reaction by 1.1-fold of sample numbers due to pipetting variations.
3. Mix well. Incubate the reaction mixture at 90 °C for 1 min and immediately place samples on ice for 2 min.
4. Add 1 µL RNase inhibitor and 1 µL SSII to each sample. Tap to mix well. Centrifuge at 5000 *g* for 1 min. Incubate samples at 42 °C for 1.5 h.
5. Add 40 µL 150 mM KOH/20 mM Tris-Base to each sample. Mix well. Incubate samples at 90 °C for 10 min.
6. Neutralize pH value of each cDNA sample with a certain volume of 150 mM HCl to about 7–8 using Hydrion S/R Dispenser pH papers.

3.9 PCR

1. Prepare Mastermix for a small-scale PCR for one sample (8 µL 10 × KOD Buffer, 3.2 µL 25 mM MgSO₄, 3.2 µL dNTPs (2 mM each), 0.8 µL 5' primer (SR primer containing P5 sequence), 0.8 µL 3' primer (different index attached with P7 sequence), 8 µL cDNA, 1.6 µL KOD, and 54.4 µL water). The total volume is 80 µL. Aliquot into 4 reactions evenly. Repeat for other samples with the replacement of cDNA and corresponding index primers.

2. Program PCR reactions as below: 95 °C 2 min first, then program cycles containing 95 °C 20 s, 55 °C 30 s, and 68 °C 30 s. The cycle numbers are 12, 15, 18, and 21, matching to each reaction above.
3. Run a 2.5% agarose gel. Put 4 products from one sample together. Select a cycle in which the amplification is at an exponential stage. The expected band size should be between 100 bp and 200 bp (Fig. 4a).
4. After defining the cycle numbers, prepare large-scale PCR reaction system: 30 µL 10× KOD Buffer, 12 µL 25 mM MgSO₄, 12 µL dNTPs (2 mM each), 3 µL 3' primer, 3 µL 5' primer, 30 µL cDNA, 6 µL KOD, and 204 µL water. The total volume is 300 µL. Aliquot each reaction into 8 PCR tubes. Run PCR program as described in **step 2** but with the corresponding cycle for each sample.
5. After amplification, combine PCR products from the same sample to a 1.5 mL Eppendorf tube. Add 300 µL phenol: chloroform:isoamyl alcohol (25:24:1) pH 8.0 to each tube. Vortex for 20 s. Centrifuge at full speed at 4 °C for 10 min.
6. Take the upper, aqueous phase to a new 1.5 mL tube. Add 300 µL chloroform to the supernatant and vortex for 20 s. Centrifuge at full speed at 4 °C for 10 min.
7. Decant 250 µL upper, aqueous phase to a new 1.5 mL tube without touching the wall of the tube. Add 25 µL 3 M NaAc and 750 µL ethanol. Mix well by rotating tubes upside down. Then, incubate tubes at -20 °C overnight.

3.10 Pme I Digestion

This step is to specifically remove the PCR products amplified from the spiked and radiolabeled markers because the marker sequence contains a Pme I digestion site. The digestion must be complete to avoid sequencing of RNA markers.

1. Centrifuge samples above in full speed at 4 °C for 10 min. Remove the supernatant. Air-dry DNA pellet for 3 min.
2. Resuspend DNA in Pme I reaction mix containing 5 µL 10 × Cutsmart, 1 µL Pme I and 44 µL water for each reaction. The volume for each reaction is 50 µL in total. Multiply the volume for one reaction by 1.1-fold of sample numbers due to pipetting variations. Mix well and aliquot to each sample.
3. After DNA is fully dissolved, incubate reaction at 37 °C at least 4 h.
4. Add 5 µL DNA loading buffer to each sample and mix well.
5. Prepare 3% low-melting gel. Gently transfer gel to a big tank. Load samples with empty well between different ones. Run gel electrophoresis with 0.5× TBE as running buffer.
6. Shot gel pictures under UV light (Fig. 4b).
7. Use a clean blade to cut gel and transfer gel slice to a 1.5 mL Eppendorf tube. Use water to rinse blade and use kim wipe to wipe it before processing next sample. After all the cutting finished, take gel pictures to make sure that target bands are completely covered (Fig. 4c).

8. Do gel extraction with the QiaQuick gel extraction kit or a comparable kit according to the manufacturer's instructions. Recover the DNA in 30 μ L of 10 mM Tris-HCl pH 8.0 buffer.
9. Use nanodrop to measure cDNA concentration. Normalize the samples to 100 ng/ μ L. Run gel electrophoresis to normalize since nanodrop analysis may be not accurate enough (Fig. 4D). Pool cDNA libraries and the mixture is now ready for Illumina sequencing.

3.11 Computing Analysis Workflow

Since sRNA seq has been widely conducted in numerous laboratories, correspondingly, lots of computational tools have been developed [21–23]. Based on the published tools, open sources, and our own experience, we perform our computing data analysis according to the pipeline shown in Fig. 5. The software used here are listed in Table 3 [18].

1. Uncompress downloaded files to get fastq format file.
Usually the sequencing results are transferred as compressed files.

```
$ tar -xvf example.tar
```



```
$ gunzip example.gz
```
2. Perform Quality control (QC) to filter the low-quality reads and adapter trimming.

Given that the quality of NGS data can be affected by contamination of adapters or primers, false duplicates from PCR, read errors among other artifacts, QC is a critical step before further data processing [24, 25]. We use fastQC (Table 3) to check the sequence quality. Clean fastq file will be generated. Reads generated through sRNA seq may cover adapters partially. Untrimmed adapters will compromise the mapping accuracy [23]. Several tools can be used for this step including fastx [26], cutadapt [27], skewer [28], trimmomatic [29], etc. Fastx is developed earlier and still widely used tool to trim adapters, especially for 3' adapters. This method has a fair balance between sensitivity and specificity with a medium speed [28]. Herein we remove 3' adapter by using fastx clipper to get "example-rmadapt.fastq" (See Note 10).

```
$ fastx_clipper -a NNNNTGGAATTCTCGGGTGCCAAGG -Q 33 -c -l 18 -o /
example-rmadapt.fastq -i /example.fastq -v;
```

3. tRNA and rRNA removal.

In total RNA, sRNAs only account for about 1% or less, while rRNA and tRNA consist 90%. Even if 28S and 18S rRNA can be largely excluded by size selection, we may still have contaminations of 5S rRNA (~120 nt) and tRNA

¹⁰. After this step, one can use fastQC (Table 3) to check the distribution of reads in different lengths. Usually reads amount of 24 nt should be higher or even about twice of that of 21 nt except some tissues with actively dividing and developing cells [39]. If the samples are from materials including vegetative organs or inflorescence buds, 24 nt reads amount is similar or even lower than 21 nt reads amount, and the RNA might have a certain level of degradation.

(~70 nt). Remove the potentially contaminated tRNA and rRNA is an important step. Here we use bowtie, a fast and efficient program to align short reads to large genomes, to remove tRNA and rRNA by using bowtie to get “example_rmstructure-rma-dapt.fastq”.

```
$ bowtie-build structure_RNA.fa structure_RNA
```

```
$ bowtie -v 0 -a structure_RNA -p 10 -q /aim_directory/example-rmadapt.fastq --un /file_localized_directory/ example_rmstructure-rmadapt.fastq;
```

4. Extract data of reads number between 18 to 30 nt.

Most sRNAs range from 18 to 30 nt in length, after removal of contamination of structural RNAs, further cleaning of sRNA reads is required to keep the 18–30 nt sRNAs. Extract data of reads number between 18 and 30 nt by awk command and obtain “example_18-30 nt.fastq”.

```
$ awk 'BEGIN {OFS = "\n"} {header = $0 ; getline seq ; getline qheader; getline qseq ; if (length(seq) >= 18 && length (seq) <= 30) {print header, seq, qheader, qseq}}' < example_rmstructure-rmadapt.fastq > example_18-30nt.fastq;
```

5. Data alignment.

Alignment is to map reads back to reference genomes. More than 60 alignment tools have been developed up to date [30]. Among them, bowtie/bowtie2 shows an ultrafast speed and a high-quality score [30]. In our pipeline, we use bowtie to align reads to TAIR10 to generate “example.sam” [31]. Sequence alignment map (sam) file is the output of aligners.

```
$ bowtie -v 0 -a TAIR10 -p 10 -q /aim_directory/ example_18–30 nt.fastq -S /aim_directory/ example.sam;
```

6. Reads sorting.

A sam file contains all the details of the mapping information of reads. The size of a sam file is larger than the corresponding fastq file. However, the reads information is unsorted. Usually, binary alignment map (bam) file, the compressed binary version of sam format, is more frequently used in data analysis since bam file occupies less memory space and supports random retrieval in a certain region [32]. Samtools is a powerful program to convert bam to sam and sam to bam [32]. We use samtools to compress sam file into bam file and meanwhile sort reads by chromosome at the same time. “example.sorted.n.bam” file is the output.

```
$ samtools view -bS example.sam | samtools sort -n - example_sorted;
```

7. Reads counting.

After alignment, we next will count the reads number mapped to loci or regions according to certain features (e.g., pri-miRNAs, exons, transposons, or isoforms). Several programs have been developed to count reads such as cufflinks [33], featureCount [34], and HTSeq [35]. Here we use HTSeq package to count how

many reads map to each feature. HTSeq is a Python library to facilitate the rapid development of scripts for processing and analyzing sequencing data, which include parsers for common file formats for a variety of types of input data. Here we use the annotation file (gff3) downloaded from miRbase (<http://www.mirbase.org/>) (see Note 11).

```
$ htseq-count --format=bam --stranded=no --type=miRNA --idattr=Name
example.sorted.n.bam file_downloaded_from_miRBase.gff3 > example.count;
```

The output file is a count matrix with rows of sRNA names and columns of samples.

8. Normalization.

After counting reads, we get the reads number of each sRNA. Given that the read counts in one region is also decided by sequencing depth, we need to normalize counts number if we want to compare expression levels of one gene between different samples. There are different methods to perform normalization including Total count (TC), Upper Quartile (UQ), Median (Med), Trimmed Mean of M-values (TMM), Quantile (Q), etc. [36]. Several R packages are commonly used for detection of differentially expressed genes from RNA-Seq or sRNA-seq data. We often use DeSeq2 or EdgeR to analyze the raw count data [37]. These two methods intend to minimize the effect from genes of extreme expression and use medium expressed genes to normalization. The analysis process includes three main steps: (1) normalization, (2) dispersion estimation, and (3) test for differential expression. We use EdgeR normalized by TMM methods as an example to illustrate the quantification process. Other packages or normalize parameters could also be used based on different scenarios.

- a. Load the raw reads count to R and set the group of 2 samples with 3 biological replicates.


```
example_all <- read.table(file = " example.count ")
exampleSet <- example_all [,1:6]
group_list <- factor(c(rep("WT",3), rep("mutant",3)))
```
- b. Filter the low count reads in each sample and normalize the raw reads count to total reads.


```
exampleSet <- exampleSet [rowSums(cpm(exampleSet) > 1) >= 2,]
exampleSet <- DGEList(counts = exampleSet, group = group_list,
lib.size=c(total reads of each sample))
exampleSet <- calcNormFactors(exprSet, method="TMM")
```
- c. Use qCML (quantile-adjusted conditional maximum like-lihood) to estimate the dispersion factors

¹¹Make sure that the chromosome names in the index file used for bowtie alignment are consistent with those in the gff3 file.

```
exampleSet <- estimateCommonDisp(exampleSet)
```

```
exampleSet <- estimateTagwiseDisp(exampleSet)
```

- d. Generate differential gene expression list with exactTest; normally, genes with adjusted p value < 0.05 were thought as significantly differentially expressed.

```
example <- exactTest(exampleSet)
```

```
tTag <- topTags(example, n=nrow(exampleSet))
```

```
tTag <- as.data.frame(tTag)
```

```
write.csv (tTag,file = "WT_vs_mutant_edgeR.csv")
```

9. Generating reports, tables, and figures.

In an attempt to provide a user-friendly way to the whole analysis process, many integrated packages have also been developed for users to generate reports, tables, and figs [38]. SARTools proposed a comprehensive, easy-to-use, DESeq2 and edgeR-based R pipeline that covers all the steps of a differential expression analysis, and automatic reports including data quality control, correlation between replicates and plot for test of differential expression [37]. Customized parameters can be defined for each script. For example:

- a. Input files (raw reads counts file generated from **step 7**) and the working directory
- b. Project name and related information
- c. Experimental design (including samples, replicates and different treatments, etc.)
- d. Normalization and statistical test methods (a sizeFactor, i.e., a scaling coefficient that applies to raw counts, and a normalization factor, i.e., a scaling coefficient that applies to library sizes.)
- e. Filtering process (setting p value and lowest count reads)

For detailed usage of SARTools, the user manual can be found online (Table 3). Based on the tables, the expression of specific sRNAs of interest can be further analyzed. Clustering analysis such as heatmap (<http://biit.cs.ut.ee/clustvis/>), Gene Ontology (GO) enrichment analysis (<http://geneontology.org/docs/go-enrichment-analysis/>), and others can be used according to different experimental purposes.

Acknowledgments

The work in the X. Zhang's laboratory was supported by the NIH grant R01GM132401. D.S. was supported by the China Scholar Council Fellowship.

References

1. Ma Z, Zhang X (2018) Actions of plant argonautes: predictable or unpredictable? *Curr Opin Plant Biol* 45:59–67 [PubMed: 29857309]
2. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297 [PubMed: 14744438]
3. Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small rna. *Nat Protoc* 3:1077 [PubMed: 18536652]
4. Kroh EM, Parkin RK, Mitchell PS et al. (2010) Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-pcr (qrt-pcr). *Methods* 50:298–301 [PubMed: 20146939]
5. Chen C, Ridzon DA, Broomer AJ et al. (2005) Real-time quantification of microRNAs by stem-loop rt-pcr. *Nucleic Acids Res* 33: e179–e179 [PubMed: 16314309]
6. Git A, Dvinge H, Salmon-Divon M et al. (2010) Systematic comparison of microarray profiling, real-time pcr, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* 16:991–1006 [PubMed: 20360395]
7. Hafner M, Landgraf P, Ludwig J et al. (2008) Identification of microRNAs and other small regulatory rnas using cdna library sequencing. *Methods* 44:3–12 [PubMed: 18158127]
8. Wang H, Zhang X, Liu J et al. (2011) Deep sequencing of small rnas specifically associated with arabidopsis ago1 and ago 1 uncovers new ago functions. *Plant J* 67:292–304 [PubMed: 21457371]
9. Williams Z, Ben-Dov IZ, Elias R et al. (2013) Comprehensive profiling of circulating microRNA via small rna sequencing of cdna libraries reveals biomarker potential and limitations. *Proc Natl Acad Sci U S A* 110:4255–4260 [PubMed: 23440203]
10. Baran-Gale J, Kurtz CL, Erdos MR et al. (2015) Addressing bias in small rna library preparation for sequencing: a new protocol recovers microRNAs that evade capture by current methods. *Front Genet* 6:352 [PubMed: 26734062]
11. Hafner M, Renwick N, Farazi TA et al. (2012) Barcoded cdna library preparation for small rna profiling by next-generation sequencing. *Methods* 58:164–170 [PubMed: 22885844]
12. Hafner M, Renwick N, Brown M et al. (2011) Rna-ligase-dependent biases in mirna representation in deep-sequenced small rna cdna libraries. *RNA (New York, NY)* 17:1697–1712
13. Raabe CA, Tang T-H, Brosius J et al. (2013) Biases in small rna deep sequencing data. *Nucleic Acids Res* 42:1414–1426 [PubMed: 24198247]
14. Zhu H, Hu F, Wang R et al (2011) Arabidopsis argonaute10 specifically sequesters mir166/165 to regulate shoot apical meristem development. *Cell* 145:242–256 [PubMed: 21496644]
15. Zhang Z, Liu X, Guo X et al. (2016) Arabidopsis ago3 predominantly recruits 24-nt small rnas to regulate epigenetic silencing. *Nat Plants* 2:16049 [PubMed: 27243648]
16. Zhang Z, Hu F, Sung MW et al. (2017) Risc-interacting clearing 3'-5' exoribonucleases (rices) degrade uridylylated cleavage fragments to maintain functional risc in *Arabidopsis thaliana*. *elife* 6:e24466 [PubMed: 28463111]
17. Wang Z, Ma Z, Castillo-González C et al. (2018) Swi2/snf2 atpase chr2 remodels pri-mirnas via serrate to impede mirna production. *Nature* 557:516–521 [PubMed: 29769717]
18. Ma Z, Castillo-Gonzalez C, Wang Z et al. (2018) Arabidopsis serrate coordinates histone methyltransferases atxr5/6 and rna processing factor rdr6 to regulate transposon expression. *Dev Cell* 45: 769–784 e766 [PubMed: 29920280]
19. Guo X, Ma Z, Zhang Z et al. (2017) Small rna-sequencing links physiological changes and rddm process to vegetative-to-floral transition in apple. *Front Plant Sci* 8:873 [PubMed: 28611800]
20. Lee Y, Kim VN (2005) Preparation and analysis of droscha Rna silencing. Springer, pp 17–28
21. Kuksa PP, Amlie-Wolf A, Katani Ž et al. (2018) Spar: Small rna-seq portal for analysis of sequencing experiments. *Nucleic Acids Res* 46:W36–W42 [PubMed: 29733404]
22. Hoogstrate Y, Jenster G, Martens-Uzunova ES (2014) Flaimapper: computational annotation of small ncna-derived fragments using rna-seq high-throughput data. *Bioinformatics* 31:665–673 [PubMed: 25338717]

23. Wu X, Kim T-K, Baxter D et al. (2017) Srrnanalyzer—a flexible and customizable small rna sequencing data analysis pipeline. *Nucleic Acids Res* 45:12140–12151 [PubMed: 29069500]
24. Davis MPA, van Dongen S, Abreu-Goodger C et al. (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63:41–49 [PubMed: 23816787]
25. Patel RK, Jain M (2012) Ngs qc toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619 [PubMed: 22312429]
26. Gordon A, Hannon G (2010) Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit 5
27. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12
28. Jiang H, Lei R, Ding S-W et al. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform* 15:182–182
29. Lohse M, Bolger AM, Nagel A et al. (2012) Robina: a user-friendly, integrated software solution for rna-seq-based transcriptomics. *Nucleic Acids Res* 40:W622–W627 [PubMed: 22684630]
30. Fonseca NA, Rung J, Brazma A et al. (2012) Tools for mapping high-throughput sequencing data. *Bioinformatics* 28:3169–3177 [PubMed: 23060614]
31. Langmead B, Trapnell C, Pop M et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25 [PubMed: 19261174]
32. Li H, Handsaker B, Wysoker A et al. (2009) The sequence alignment/map format and samtools. *Bioinformatics* 25:2078–2079 [PubMed: 19505943]
33. Trapnell C, Roberts A, Goff L et al. (2012) Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc* 7:562 [PubMed: 22383036]
34. Liao Y, Smyth GK, Shi W (2013) Feature-counts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930 [PubMed: 24227677]
35. Pyl PT, Anders S, Huber W (2014) Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169 [PubMed: 25260700]
36. Dillies M-A, Rau A, Aubert J et al. (2012) A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis. *Brief Bioinform* 14:671–683 [PubMed: 22988256]
37. Varet H, Brillet-Guéguen L, Coppée J-Y et al. (2016) Sartools: a deseq2-and edger-based r pipeline for comprehensive differential analysis of rna-seq data. *PLoS One* 11:e0157022 [PubMed: 27280887]
38. Russo F, Angelini C (2014) Rnaseqgui: a gui for analysing rna-seq data. *Bioinformatics* 30:2514–2516 [PubMed: 24812338]
39. Tabara M, Ohtani M, Kanekatsu M et al. (2018) Size distribution of small interfering RNAs in various organs at different developmental stages is primarily determined by the dicing activity of dicer-like proteins in plants. *Plant Cell Physiol* 59:2228–2238 [PubMed: 30032266]

Fig. 1.
Work flow for the construction of a sRNA library

Fig. 2.

Synthesis of the 5' adenylated 3' adapter oligonucleotides. (a) Setting up an O_2 -free reaction system using a balloon with an excess of N_2 (in flask B). (b) Adding solution A dropwise to solution B. (c) Detection of synthesized ImpA through TLC. The product pointed by an arrow indicates the synthesized ImpA. (d) Visualization of adenylated adapter through UV shadowing. The adenylated adapter (upper band) and the unadenylated adapter (lower band), and the silica gel plate are marked by black arrows. In addition to recovery of

the pre-adenylated product, the unadenylated adapter should also be recovered for a repeat assay when needed

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 3. Tracking of sRNA construction processing through radio-autographic monitor. **(a)** Radioisotope labeled 19 nt and 24 nt markers before purification. **(b)** Separation of sRNAs from total RNAs using spiked 19 and 24 nt internal markers. **(c)** Imaging of sRNA and internal markers after the 3' ligation of the pre-adenylated adapters. **(d)** Imaging of sRNA and internal markers after the 5' ligation

Fig. 4. PCR to amplify the sRNA library. **(a)** Small-scale PCR with different cycles to optimize the cycle number, where the cDNA application is at the exponential stage. The product size lies between 100 and 200 bp. In the examples here, amplification with 14 or 15 cycles is an ideal selection. Note: amplification cycles might vary for different samples. **(b)** Large-scale PCR with the optimized cycles from the small-scale experiments. **(c)** Leftovers after the recovery of sRNA-cDNA libraries. **(d)** Normalization steps before making the library pool

Fig. 5.
Computational pipeline for sRNA-seq data analysis

Table 1

Markers, adapters, and primers

Product	Index primer sequence
19 nt size marker	5'-rCrGrUrArCrGrCrGrGrUrUrUrArArArCrGrA-3'
24 nt size marker	5'-rCrGrUrArCrGrCrGrGrArArUrArGrUrUrArArArCrUrGrU-3'
SR Primer	5'-AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCG-s-A-3'
SR-RT Primer	5'-GCCTTGGCACCCGAGAATTCCA-3'
3' adapter	5'-NNNNTGGAATTCTCGGGTGCCAAGG-NH ₂ -3'
5' adapter	5'-rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-3'

Table 2

Different concentrated urea-polyacrylamide gel recipe

Components	Concentration		
	10%	15%	20%
UreaGel buffer	8 mL	8 mL	15 mL
UreaGel concentrate	32 mL	48 mL	120 mL
UreaGel diluent	40 mL	24 mL	15 mL
10% APS	640 µL	640 µL	1.2 mL
TEMED	32 µL	32 µL	60 µL
Total volume	80 mL	80 mL	150 mL

Table 3

Software used in bioinformatic analysis

Software	Source	Website
Bowtie v1.1.2	[31]	http://bowtie-bio.sourceforge.net/index.shtml
FASTX-Toolkit v0.0.13	The Hannon Lab	http://hannonlab.cshl.edu/fastx_toolkit/
Samtools	N/A	http://samtools.sourceforge.net/
HTSeq v0.11.1	[35]	https://htseq.readthedocs.io/en/release_0.11.1/
SARTools	[37]	https://github.com/PF2-pasteur-fr/SARTools
FastQC v0.11.8	N/A	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/