April 18, 2016
Tetsuya Kurata, Tohoku University

Preparation of 5'-DGE library from plant sample

Summary

Next-generation sequencing (NGS) has become powerful tool for various types of biological studies. Applications for NGS can be mainly categorized into two approaches. One is "sequence-based" analysis including re-sequencing on reference sequence(Sabot et al. 2011), single nucleotide polymorphism (SNP) (Austin et al. 2011, Schneeberger et al. 2009, Uchida et al. 2011), and genome-wide bisulfite-sequencing to define the methylated cytosines (Lister et al. 2009, Lister et al. 2008). The other is "read count-based" quantitative analysis; gene expression (RNA-seq, TSS-seq and 5'-DGE) (Filichkin et al. 2010, Hashimoto et al. 2009, Nishiyama et al. 2012, Tsuchihara et al. 2009), distribution of transcription factor (Kaufmann et al. 2010), histone modifications (Brusslan et al. 2012), and nucleosome occupancy (Stroud et al. 2012) , some of which employ chromatin-immunoprecipitation is employed (ChIP-seq).

In this protocol, I will introduce a procedure for the preparation of libraries to investigate 5'-end sequences of mRNA (5'-DGE) with plant samples. This protocol has been developed using moss (*Physcomitrella patens*) samples at first (Nishiyama, Miyawaki, Ohshima, Thompson, Nagashima, Hasebe and Kurata 2012), but it can be suitable for 5'-DGE analysis in *Arabidopsis* and also other plant species. Though this method was applied to SOLiD systems (Life Technologies), it can be applicable to illumina sequencing platform by using illumina-compatible oligos (Metzker 2010). Furthermore, the

5’-DGE method has an advantage that it needs lower number of sequence reads lather than RNA-seq for quantitative gene expression analysis on annotated genes since the latter uses multiple reads spanning a full length of each mRNA but the former uses the one 5’-end reads from each mRNA transcript. Therefore 5’-DGE could be a more economical method for quantitative gene expression analysis.

Materials and Methods

EQUIPMENTS

Mortar and pestle
1.5 ml LoBind tube (Eppendorf)
Microcentrifuge
Thermal cycler
0.2 ml PCR tube
Razor
Magna-Sep Magnetic Particle Separator, 6 holes (Life Technologies)
Real time quantitative PCR system (qPCR)
Nanosep Centrifugal Devices with Omega Membrane, 100K (Pall)

REAGENTS

Arabidopsis tissue of choice
RNeasy Plant Mini Kit (Qiagen)
FastTrack MAG Micro mRNA Isolation Kit (Life Technologies)
PrimeScript II RT Kit (Takara Bio)
Oligos (see Table 1) [1]*
DTT
25 mM NaOH
KOD-Plus kit (Toyobo)

* [1] These custom-designed oligos were purchased from Integrated DNA Technologies, Inc.

MinElute Reaction Cleanup Kit (Qiagen)

EcoP15I (New England Biolabs)

MAGNOTEX-SA kit (Westburg)

T4 DNA ligase (New England Biolabs)

QuantiTect SYBR Green PCR Master Mix (Qiagen)

Phusion HF Taq system (Finnzymes)

6% polyacrylamide slab gel plate

1x TBE buffer

QIAEX II Gel Extraction Kit (Qiagen)

SYBR Gold (Life Technologies)

SYBR Safe (Life Technologies)

25 bp ladder (Life Technologies)

Safe Imager 2.0 Blue Light Transilluminator (Life Technologies)

H_20 (Deionized and autoclaved)

SOLiD Library TaqMan Quantification Kit (Life Technologies)

Table 1. Oligonucleotides used for preparation of 5'-DGE library

Oligo name	Sequence
BioTEG-P2EcoP15I-rGx3	Biotin-TEG-5'-CTGCCCCGGGTTCCTCATTCTCTCAGCArGrGrG-3'
BioTEG-dT20EcoP15I	Biotin-TEG-5'-CTATCAGCAGTTTTTTTTTTTTTTTTTTT-3'
BioTEG-P2EcoP15I-GG	Biotin-TEG-5'-CTGCCCCGGGTTCCTCATTCTCTCAGCAGGG-3'
P1-A	5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3'
P1-B-NN	5'-N*N*ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG-3'
P1AMP	5-CCACTACGCCTCCGCTTTCCTCTCTAT-3'
P2AMP	5'-CTGCCCCGGGTTCCTCATTCT-3'

* phosphorothioate bonds

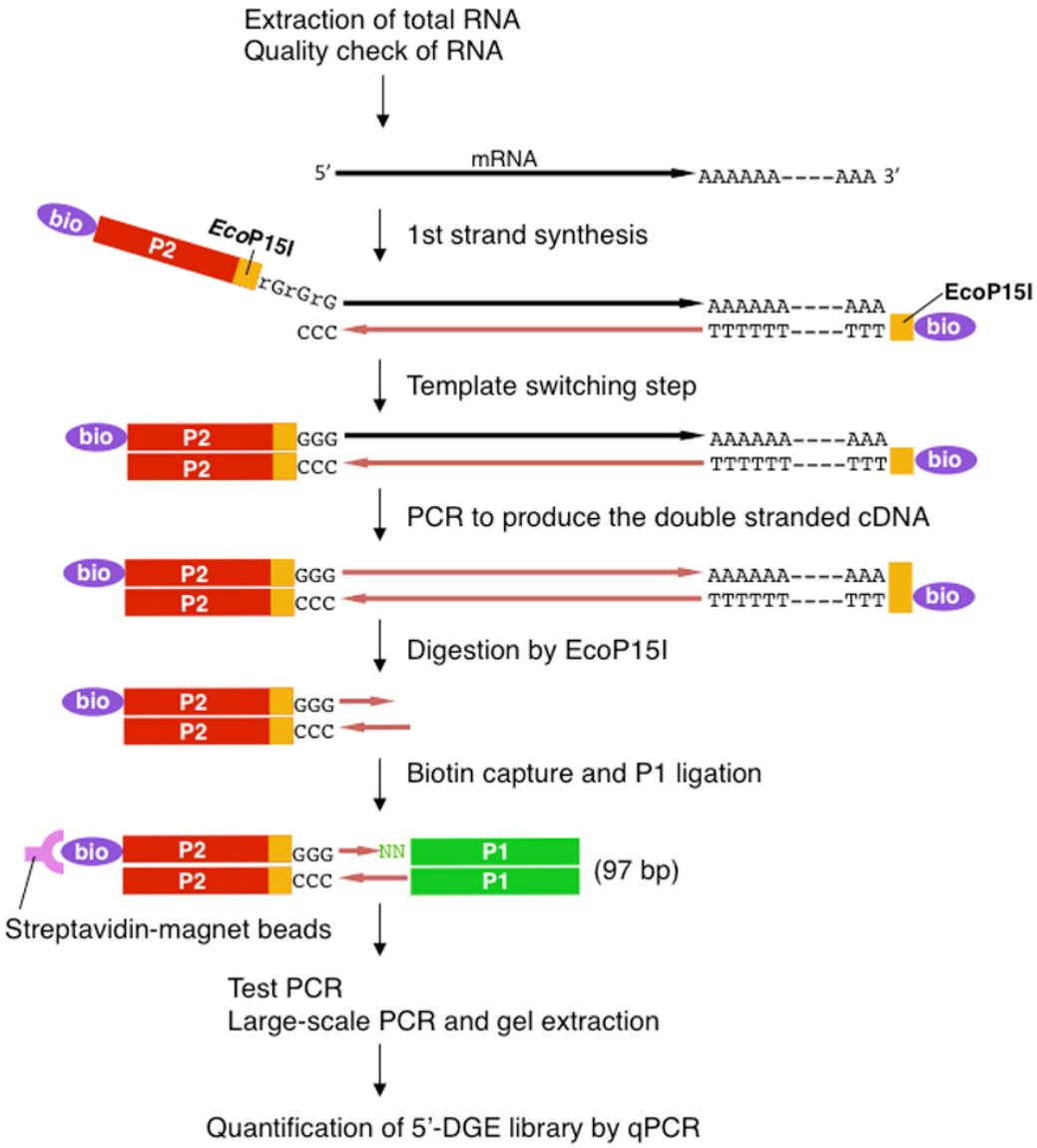

Figure 1. Schematic flow for 5'-DGE library preparation.

Modified from Nishiyama et al. (2012)

PROCEDURE

The workflow for library preparation is shown in Figure 1.

Step 1. Harvest plant samples and freeze them in liquid N_2 *[2].

1) Grind 50~100 mg Arabidopsis tissues by motor and pestle in liquid N_2 [3]*

Step 2. Extract total RNA.

1) Extract total RNA using the RNeasy Plant Mini Kit (Qiagen).

Step3. Quality check of extracted total RNA.

1) Check the quality of total RNA with the 2100 Bioanalyzer (Agilent) according to the manufacture protocol [4]*.

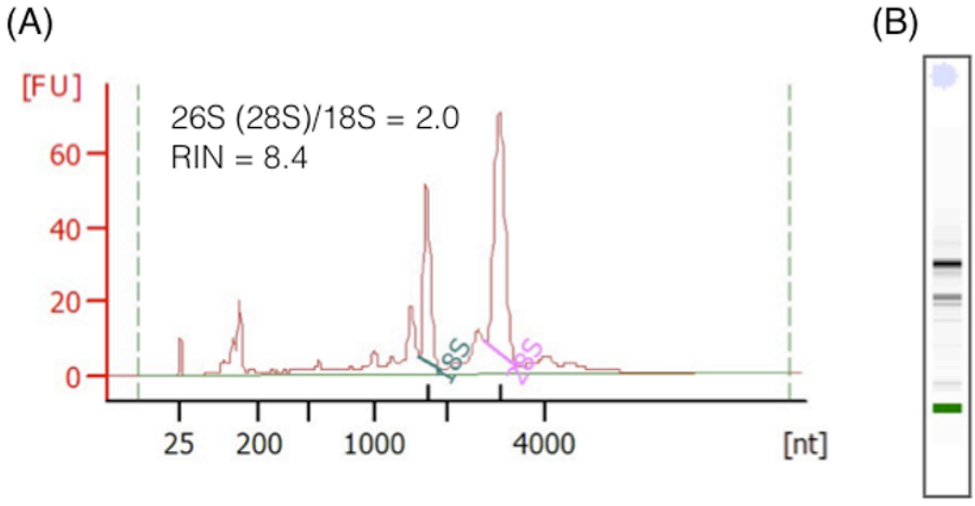

Figure 2. Quality check of total RNA from Arabidopsis.

Agilent RNA 6000 Pico Kit was used in this sample.
(A) Electropherogram, (B) Gel image. (Provided by Kenji Saeki and Dr. Miyo T. Morita).

* [2] To reduce the stochastic error on gene expression among samples, it is recommended to conduct the experiments by biological replicates; independent samples or sample-pools should be prepared from n=2, 3

* [3] If not immediately use, store it in deep freezer for several months. For other plant, enough amount to prepare 5~10 μg total RNA should be used.

* [4] Agilent RNA6000 Nano or Pico Kit is employed. In some case, it is better to dilute the RNA sample before run. Samples with good quality show a intensity ratio of 26S rRNA/18S rRNA around 2, and RIN number is other parameter for RNA. Alternatively, conventianal agarose gel electrophoresis for RNA is applicable when enough amount RNA is obtained.

Step 4. Enrichment of polyA RNA fraction.

1) Use 5 to 10 μg of total RNA was used [5]*.
2) Carry out polyA enrichment according to the manufacture protocol.
3) Elute the polyA fraction by adding 20 μl RNase-free water supplied in kit and elute again with 5 μl RNase-free water (total 25 μl)
4) Transfer a required amount of the polyA-enriched fraction into a 0.2 ml PCR tube [6]* [7]*.

* [5] To maximize the nucleic acid fragment recovery, 1.5 ml LoBind tubes should be used.

* [6] If an equalization step is omitted, quantitative bias among libraries will appear.

* [7] At least, 100 ng polyA RNA is needed for the next step.

Step 5. Reverse transcription and template-switching reaction using PrimeScript II reverse transcriptase with following oligos.

1) Prepare 5 μl of each reaction consisting of;
   poly A-enriched RNA sample (>100 ng)
   1.2 μM BioTEG-dT20EcoP15I oligo
   1.2 μM BioTEG-P2EcoP15I-rGx3 oligo
   fill up to 5 μl with H_20, resolve the RNA pellet by tapping
2) Incubate it at 72°C for 2 min in a thermal cycler for denaturing.
3) Prepare 10 μl of each reaction consisting of
   5 μl Denatured polyA and oligo solution
   2 mM DTT (dithiothreitol)
   0.5 mM dNTP
   1x PrimeScript II buffer
   1 μl PrimeScript II
   fill up to 10 μl with H_20 in 0.2 ml PCR tube per one library [8]*
4) Incubate it at 42°C for 60 min in a thermal cycler.

* [8] We routinely use 8-stripe 0.2 ml PCR tubes for multiple samples.

Step 6. Alkaline lysis of mRNAs.

1) Add 1 μl of 25 mM NaOH.
2) incubate it at 68°C for 30 min in a thermal cycler.

Step 7. Conduct PCR by three cycles to obtain double strand cDNAs.

1) Prepare 100 μl of PCR mix containing of;
   11 μl first-strand cDNA
   0.2 mM dNTP
   1.2 mM $MgSO_4$
   0.3 μM BioTEG-P2EcoP15I-GG oligo
   0.3 μM BioTEG-dT20EcoP15I oligo
   1x KOD-Plus buffer
   2 μl KOD-Plus DNA polymerase
2) PCR condition
   95°C for 3 min
   72°C for 10 min
   95°C for 20 sec

   95°C for 5 sec
   68°C for 8 min] X3

Step 8. Cleanup of cDNAs.

1) Purify cDNA with the MinElute Reaction Cleanup Kit (Qiagen).
2) After the binding step, wash the MinElute column with 750 μL of 35% guanidine hydrochloride solution [9]*.
3) Elute cDNA with 13 μL EB in 1.5 ml LoBind Tube.

* [9] Treatment by guanidine hydrochloride is effective to remove residual oligos.

Step 9. Digest the purified cDNA fragments with EcoP15I.

1) 50 μl reaction consists
   8 μl purified cDNA
   5 μl 10X NEB buffer 3
   0.5 μl 100X BSA
   5 μl 10X ATP
   1 μl EcoP15I (10 units)
   fill up to 50 μl H_20
2) Incubate it at 37°C for o/n.

Step 10. Capture the biotin-labeled 5'-end cDNA fragments.

1) Add MAGNOTEX-SA beads into 1.5 ml LoBind Tube [5]* [10]*.
2) Insert the tube into the magnetic separator, and place for a minute.
3) Discard the supernatant.
4) Mix the following components, then add to the tube.
   50 μl EcoP15I-digested cDNA
   50 μl 2X Binding Buffer
5) Mix well by pipetting, and rotate at room temperature for 10 minutes.
6) Insert the tube into the magnetic separator, and place for a minute.
7) Discard the supernatant.
8) Wash the beads with 200 μL of 1x Binding buffer
9) Insert the tube into the magnetic separator, and place for a minute.
10) Discard the supernatant.
11) Repeat the washing step twice.

* [10] For multiple samples, 8-stripe 0.2 ml PCR tubes are useful but a compatible magnetic separator (e.g. DynaMag-PCR; Life Technologies) is required.

Step 11. Ligate P1 adaptor to captured biotin-labeled 5'-end cDNA fragments.

1) Suspend beads in 100 μl of 1x T4 DNA ligase buffer
2) Insert the tube into the magnetic separator, and place for a minute.
3) Discard the supernatant.
4) Repeat the step twice. At the second time, transfer the beads to a fresh tube.
5) Add the following solutions into the tube as listed order,

42 μl H_20
2 μl 50 μM P1 adaptor [11]*
5 μl 10X T4 DNA ligation buffer
1 μl T4 DNA ligase
total 50 μl

6) Rotate the tube at room temperature (20-25℃) for 2 hours.
7) Insert the tube into the magnetic separator, and place for a minute.
8) Wash the beads with 100 µl of 1x Binding buffer.
9) Insert the tube into the magnetic separator, and place for a minute.
10) Discard the supernatant.
11) Repeat the above wash step twice.
12) Wash the beads with 100 µl of 1x Phusion HF buffer.
13) Insert the tube into the magnetic separator, and place for a minute.
14) Discard the supernatant.
15) Suspend the beads with 50 µl of 1x Phusion HF buffer.
16) Transfer bead solution into 1.5 ml LoBind tube.

* [11] To prepare P1 adaptor,
1) Mix 100 µM P1-A with equal volume of 100 µM P1-B-NN.
2) Denature it at 95℃ for 10 min.
3) Leave it at room temperature.

Step 12. Quantify of captured template for PCR.

1) Prepare fifty-fold diluted beads solution.
2) Prepare standard samples [12]*: standard were 2.5, 0.25, 0.025, 0.0025, 0.00025 pg/μl.
3) Prepare the following 10 μl of PCR mix for each reaction,
    2 μl diluted template
    5 μl 2x QuantiTect SYBR Green PCR Master Mix
    0.03 μl 100 μM P1 Fw primer
    0.03 μl 100 μM P1 Rv primer
    fill up to 10 μl by H_20
4) Conduct absolute quantification at triplicate for each sample.

* [12] Standard was prepared by cloning P1-P2 sandwich 5'-DGE fragment amplified in any library PCR. It should be confirmed that cloned standard has P1-P2 sandwich structure.

Step 13. Test amplification of 5'-DGE library.

1) Examine several amounts of templates to optimize a condition for the large-scale PCR condition [13]* [14]*.
2) Use 5 to 25 pg library-templates.
3) Prepare 50 μl of PCR mix consisting of;
    Prepared library-templates
    0.2 mM dNTP
    0.5 μM P1AMP
    0.5 μM P2AMP
    1x Phusion HF buffer
    0.5 μl Phusion HF DNA polymerase
    fill up to 50 μl by H_2O
4) PCR condition
    98°C for 30 sec

    98°C for 10 sec
    72°C for 30 sec] X12 [15]*

* [13] Over-amplification produces extra bands (see Figure 2, 3).

* [14] This step can be omitted if PCR condition for similar samples is optimized.

* [15] PCR in which total 15 cycle including PCR at step7 is done. We confirmed linearity between non-amplification sample and amplified one by qPCR (data not shown). It is possible to reduce PCR cycles in this step.

4°C ∞

5) Perform ethanol precipitation with normal (not LoBind) 1.5 ml since pellets come off when LoBind tubes are used [5]*.

6) Dissolve the pellet in 8 μl of H_20.

7) Prepare 6% polyacrylamide slab gel plate using 1X TBE [16]*.

8) Add loading dye to dissolve amplified-samples.

9) Use 25-bp ladder as a marker.

10) Run at 180 constant voltage for 25 min.

11) Stain gel in 1X TEBE containing 5,000 fold diluted SYBR-Gold-TBE solution for 15 min.

12) Decide the amount of templates for the large-scale PCR reaction [17]*.

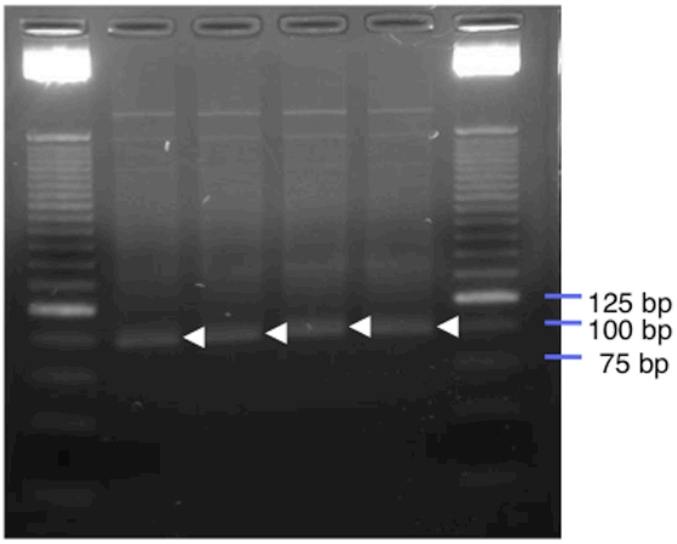

Figure 3. Gel image after 3% AGE

Arrow head indicates amplified 5'-DGE library bands. 3% MetaPhore agarose in 1X TAE was used. (Photograph by Masumi Ohshima).

* [16] We used mini gel type slab PAGE system: plate size-106 mm (width) x 100 mm (length) (BE-211, BIO CRAFT). Alternatively, usual 3% agarose gel electrophoresis is useful with low-length suitable agarose (e.g. MetaPhore Agarose, Takara Bio) can be used.

* [17] Appearance of extra bands at 120 bp indicates over-amplification. In that case, template amount will be reduced. 75 pg of template is routinely used for large-scale PCR per a library.

Step 14. Amplify 5'-DGE library in large scale.

1) Set the following 250 μl of PCR mixture,

75 pg library-template

0.2 μM dNTP
0.5 μM P1AMP
0.5 μM P2AMP
1x Phusion HF buffer
2.5 μl Phusion HF DNA polymerase
fill up to 250 μl with H_2O

2) Divide PCR mixture into five PCR tubes.

3) PCR condition,

98°C for 30 sec

98°C for 10 sec
72°C for 30 sec] X12

72°C for 10 min
4°C ∞

4) Combine five PCR solution into one normal (not LoBind) 1.5 ml since pellet came off when LoBind tube was used.

3) Perform ethanol precipitation.

4) Dissolve the pellet in 8 μl H_20.

5) Prepare 6% polyacrylamide slab gel plate usung 1X TBE.

6) Add loading dye to the samples.

7) Use 25 bp ladder as a marker.

8) Run at 180 V constant voltage for 25 min.

9) Stain gel in 1X TBE containing 5,000 fold diluted SYBR-Safe-TBE solution for 15 min.

10) Excise 97 bp band with a razor on a blue light transilluminator [18]* [19]*.

11) Chop the excised gel.

12) Transfer chopped gels into 1.5 ml tube.

13) Add 50 μl of diffusion buffer from QIAEX II Gel Extraction Kit and incubate at 50 °C for 30 min.

14) Centrifuge the sample for 1min.

15) Transfer the supernatant to NANOSEP 100K OMEGA

* [18] SYBR Safe is better than ethidium bromide regarding sensitivity and safeness.

* [18] UV transilluminator with a long wave length mode (near 365 nm) is applicable.

centrifugal device.

16) Add Centrifuge NANOSEP 100K OMEGA to filtrate residual gels.

17) Follow the manual of QIAEX II Gel Extraction Kit before elution step.

18) Elute by 20 μl of H_20 into 1.5 ml LoBind tube and elute again with 5 μl of H_20 (total 25 μl) [5]*.

19) Store the obtained 5'-DGE libraries at 4 °C.

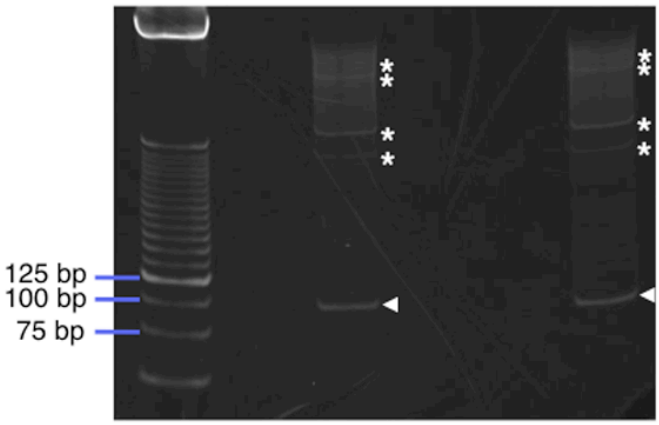

Figure 4. Gel image after large-scale amplification with 6% PAG.

Arrow heads indicates amplified 5'-DGE library bands. Extra bands are shown by asterisks.
(Photograph by Masumi Ohshima).

Step 15. Quantification of the obtained 5'-DGE libraries.

1) Prepare 1,000-fold diluted library.

2) Quantify 5'-DGE library by commercial quantification kit as according to manufacture protocol [20]*.

* [20] Quantification kit should be selected according to the the available qPCR system

KEY POINTS AND PIT FALLS

1) Design your experiment essentially according to Fisher's three principles: replication, randomization and local cotrol (Fisher 1971) for multiple samples.
2) Prepare good quality total RNA.
3) Adjust the amounts of polyA among samples.
4) Conduct the library preparation in parallel with multiple samples as possible to reduce the technical bias.
5) Before NGS run of 5'-DGE library, it is better to check the sequence by conventional PCR product cloning and sanger sequencing with multiple clones.

Acknowledgements

I would like to thank Drs. Naoyuki Uchida and Keiko Torii for critical reading and kind encouragement. Ms. Masumi Ohshima provides photographs. I also thank Dr. Miyo T. Morita and Kenji Saeki for providing bioanalyzer data. This method was developed during Hasebe Reprogramming Evolution Project (Japan Science and Technology Agency).

REFERENCES

Austin, R.S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N.J., Bonetta, D., Zhang, J., Fung, P., Gong, Y., Wang, P.W., McCourt, P. and Guttman, D.S. (2011) Next-generation mapping of Arabidopsis genes. *Plant J*, **67**, 715-725.

Brusslan, J.A., Rus Alvarez-Canterbury, A.M., Nair, N.U., Rice, J.C., Hitchler, M.J. and Pellegrini, M. (2012) Genome-wide evaluation of histone methylation changes associated with leaf senescence in Arabidopsis. *PLoS*

One, **7**, e33151.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K. and Mockler, T.C. (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res*, **20**, 45-58.

Fisher, R.A. (1971) The Design of Experiments. 9th Edition, Macmillan.

Hashimoto, S., Qu, W., Ahsan, B., Ogoshi, K., Sasaki, A., Nakatani, Y., Lee, Y., Ogawa, M., Ametani, A., Suzuki, Y., Sugano, S., Lee, C.C., Nutter, R.C., Morishita, S. and Matsushima, K. (2009) High-resolution analysis of the 5'-end transcriptome using a next generation DNA sequencer. *PLoS One*, **4**, e4108.

Kaufmann, K., Wellmer, F., Muino, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C. and Riechmann, J.L. (2010) Orchestration of floral initiation by APETALA1. *Science*, **328**, 85-89.

Lister, R., Gregory, B.D. and Ecker, J.R. (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol*, **12**, 107-118.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523-536.

Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat Rev Genet*, **11**, 31-46.

Nishiyama, T., Miyawaki, K., Ohshima, M., Thompson, K., Nagashima, A., Hasebe, M. and Kurata, T. (2012) Digital gene expression profiling by 5'-end sequencing of cDNAs during reprogramming in the moss Physcomitrella patens. *PLoS One*, **7**, e36471.

Sabot, F., Picault, N., El-Baidouri, M., Llauro, C., Chaparro, C., Piegu, B., Roulin, A., Guiderdoni, E., Delabastide, M., McCombie, R. and Panaud, O. (2011) Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J*, **66**, 241-246.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D. and Andersen, S.U. (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods*, **6**, 550-551.

Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S.E. and Gutierrez, C. (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, **109**, 5370-5375.

Tsuchihara, K., Suzuki, Y., Wakaguri, H., Irie, T., Tanimoto, K., Hashimoto, S., Matsushima, K., Mizushima-Sugano, J., Yamashita, R., Nakai, K., Bentley, D., Esumi, H. and Sugano, S. (2009) Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Research*, **37**, 2249-2263.

Uchida, N., Sakamoto, T., Kurata, T. and Tasaka, M. (2011) Identification of EMS-induced causal mutations in a non-reference Arabidopsis thaliana accession by whole genome sequencing. *Plant Cell Physiol*, **52**, 716-722.