Journal of Bioscience and Bioengineering
VOL. xx No. xx, 1–6, 2018

www.elsevier.com/locate/jbiosc

Comprehensive analysis of mRNA internal cleavage sites in *Arabidopsis thaliana*

Daishin Ueno, Shotaro Yamasaki, Taku Demura, and Ko Kato*

Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

Received 2 November 2017; accepted 25 December 2017
Available online xxx

The major obstacle of efficient transgene expression seems to be gene silencing, and one of the important factors in gene silencing is mRNA stability. Regulation of mRNA stability is an important aspect of the control of gene expression. mRNAs are degraded by both exonucleolytic digestion and endonucleolytic cleavage. However, with the exception of small RNA–guided cleavage, the mechanisms underlying endonucleolytic cleavage-dependent RNA degradation remain to be elucidated. High-throughput approaches for genome-wide profiling of RNA cleavage sites, collectively termed degradome sequencing, have been developed by several groups. These analyses have contributed to the identification of mRNA cleavage sites in plants, but due to selection of poly (A) mRNA in library preparation, these approaches cannot identify cleavage sites in a fully accurate manner. To address this issue, we developed a new experimental method, truncated RNA end sequencing (TREseq), which enabled us to accurately identify many cleavage sites. TREseq can also be used to estimate the efficiency of mRNA cleavage, revealing differences in base frequencies near cleavage sites that reflect differences in cleavage efficiency. These results will contribute to gain important knowledge about the stability of the transgene mRNA in the future.

[**Key words:** mRNA degradation; mRNA cleavage sites; Genome-wide analysis; Plant; Transgene expression]

Post-transcriptional gene silencing (PTGS) is one of the important factors for transgene expression, and it seems to be dependent on stability of transgene mRNA. Gene expression in eukaryotes is modulated at multiple levels, from initiation of transcription through mRNA degradation. mRNA degradation, a critical step in control of gene expression (1), can proceed in two ways: deadenylation followed by removal of the cap and 5′–3′ or 3′–5′ exonucleolytic digestion (2), or endonucleolytic cleavage followed by exonucleolytic digestion (3). Deadenylation-dependent RNA degradation has been well studied in yeast (4). By contrast, little is known about endonucleolytic cleavage-dependent RNA degradation, except in the context of small RNA–guided cleavage. High-throughput approaches have been developed for genome-wide profiling of mRNA cleavage sites, collectively termed degradome sequencing: these methods include parallel analysis of RNA ends (PARE) (5) and genome-wide mapping of uncapped transcripts (GMUCT) (6). These analyses have contributed to the identification of cleavage sites in plants (5,7–10), and have also revealed short motifs adjacent to cleavage sites (11). However, precise identification of multiple cleavage sites in one mRNA has yet to be achieved. Moreover, because poly (A) selection was performed for library preparation in these analyses, many of the cleavage sites identified with these methods are distributed near the 3′ ends of genes. Consequently, it is difficult to accurately estimate the efficiency of mRNA cleavage at each site.

To address these technical challenges, we developed a new experimental procedure, truncated RNA end sequencing (TREseq). Due to improvements in experimental procedures, this method can detect multiple cleavage sites in a single mRNA with minimal 3′ bias, and can also estimate cleavage efficiency. Using this technique, we identified a G-rich sequence enriched near cleavage sites that was positively associated with cleavage efficiency. In addition, we discovered a novel short motif using Dreme, which is a motif discovery tool. This knowledge could facilitate the removal of easily cleavable sequences from transgene mRNAs, as well as provide new insight into techniques for efficient transgene expression.

MATERIALS AND METHODS

Plant material *Arabidopsis thaliana* T87 cell suspension was cultured in modified Murashige–Skoog medium, as described previously (12).

Truncated RNA end sequencing For RNA isolation, T87 cells were harvested 3 days after inoculation in two independent biological replicates. Total RNA was isolated using the TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA), followed by purification using the RNeasy kit (Qiagen, Hilden, Germany) with on-column DNase I treatment according to the manufacturer's instructions.

For library preparation and sequencing, TREseq is a modification of a previously described method, non-Amplified non-Tagging Illumina Cap Analysis of Gene Expression (nAnT-iCAGE) (13). First, total RNA was depleted of rRNA using the Ribo-Zero rRNA Removal Kit (Plant Seed/Root) (Illumina, San Diego, CA, USA). Next, the depleted total RNA was reverse-transcribed with random or oligo dT primer. Subsequently, mRNA–cDNA hybrids were separated into Cap mRNA–cDNA hybrid (Cap mRNA) and Cap-less mRNA–cDNA hybrid (Cap-less mRNA) fractions using the Cap-trapping method (13) for samples reverse-transcribed with each primer (i.e., random or oligo dT). Cap mRNA (random) and Cap-less mRNA (random) were used to determine transcription start sites (TSS) and cleavage sites, respectively. Cap mRNA (random) was also used for estimating mRNA abundance (more than one tags per million (TPM) in each gene was used). Cap-less mRNA (oligo dT) was compared with Cap-less mRNA (random). Experimental procedures are shown in Fig. 1. A cDNA library was constructed for each sample using the nAnT-iCAGE method. Cap-less mRNA (random or oligo dT) was purified using AMPure XP (Beckman Coulter,

* Corresponding author. Tel.: +81 743 72 5461; fax: +81 743 72 5469.
E-mail address: kou@bs.naist.jp (K. Kato).

1389-1723/$ – see front matter
https://doi.org/10.1016/j.jbiosc.2017.12.022

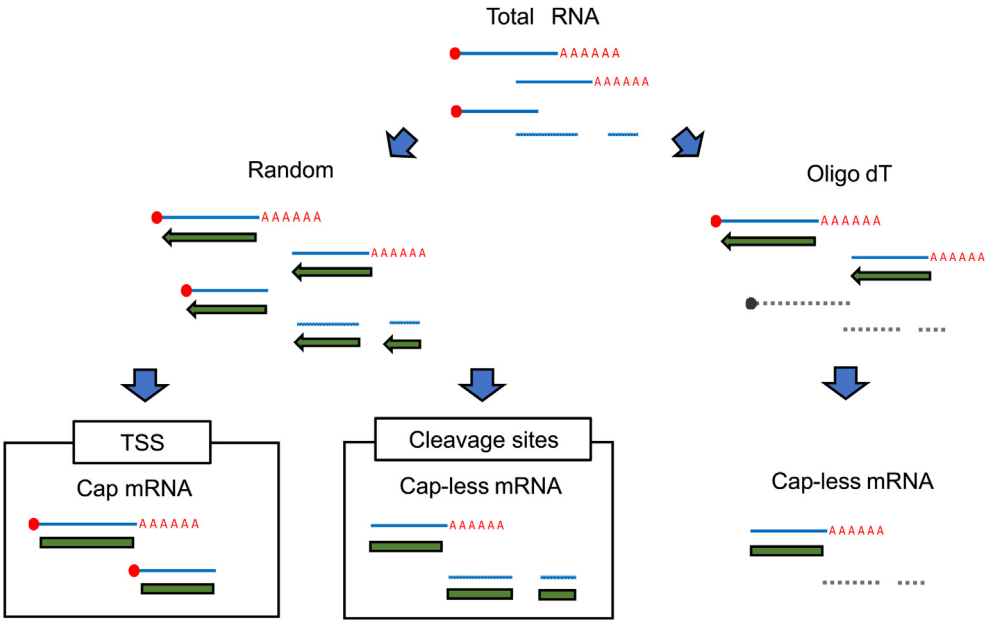

FIG. 1. Experimental procedures of truncated RNA end sequencing (TREseq). Total RNA was reverse-transcribed with random or oligo dT primer. (A) Cap mRNA and Cap-less mRNA (random) were used to determine transcription start sites (TSS) and cleavage sites, respectively. Cap mRNA (random) was also used to estimate mRNA abundance. Cap-less mRNA (oligo dT; poly $(A)^+$ mRNA) was used for comparison with cleavage site data.

Indianapolis, IN, USA) according to the manufacturer's instructions before the cDNA library was constructed. Each library was sequenced on an Illumina NextSeq 500 (Illumina).

For data analysis, each sample yielded more than 20 million single-end reads. We used the modified MOIRAI system to process TREseq data (14). Low-quality reads or reads originating from rRNA were removed. The remaining reads were mapped to the TAIR version 10 reference genome (The Arabidopsis Information Resource, www.arabidopsis.org) using bwa, which is a mapping tool. If a mismatch was found within 3 nt of the 5' end of a mapped read, this read was removed from the analysis. Tags were counted at the 5' end of each read using the MOIRAI pipeline (14). Reads were annotated with reference to TAIR10 representative gene models. For calculation of mRNA abundance, data were processed by CAGE analysis (15). Specifically, reads whose 5′ ends were not G (because the cap is 7-methylguanosine) were removed during the filtering process, but the 5′ G residues in the remaining reads were removed before mapping. Reads between 500 bp upstream from the 5′ end of the gene and the start codon (ATG) were annotated. TREseq reads are available in the DDBJ Sequence Read Archive (DRA) database with accession number DRA005995.

5-Bromouridine immunoprecipitation chase (BRIC) For RNA isolation, two days after inoculation, T87 cells were incubated for 16 h in the presence of 150 μM 5-bromouridine (BrU) in two independent biological replicates. Initiation of incubation times was adjusted to enable harvest 3 days after inoculation. After BrU-containing medium was replaced with BrU-free medium, cells were harvested at 0, 1, 3, and 6 h. Total RNA was isolated using the TRIzol Reagent and further purified by LiCl precipitation (1.5 M final concentration of LiCl). After addition of spike-in controls, BrU-labeled total RNA was isolated using the 5-BrU immunoprecipitation chase (BRIC) kit (MBL, Nagoya, Aichi, Japan).

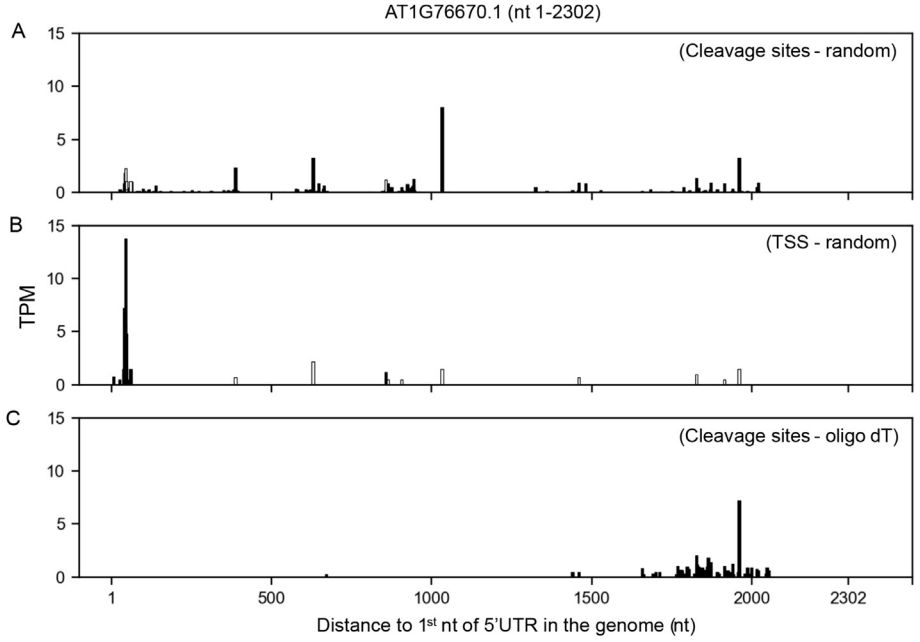

FIG. 2. Comparison of datasets. 5′ ends of the reads were counted after mapping. (A) Cap-less mRNA (random), (B) Cap mRNA (random), and (C) Cap-less mRNA (oligo dT). 1 indicates first nucleotide of the 5′UTR in the genome. Open bars indicate noise in both datasets, which was removed.

Please cite this article in press as: Ueno, D., et al., Comprehensive analysis of mRNA internal cleavage sites in *Arabidopsis thaliana*, J. Biosci. Bioeng., (2018), https://doi.org/10.1016/j.jbiosc.2017.12.022