RESEARCH PAPER

Transcriptome-wide analysis of pseudouridylation of mRNA and non-coding RNAs in Arabidopsis

Lirong Sun1†, Yuxing Xu2†, Shenglong Bai1, Xue Bai1, Huijie Zhu1, Huan Dong1, Wei Wang1, Xiaohong Zhu1, Fushun Hao1, and Chun-Peng Song1,*

1 Key Laboratory of Plant Stress Biology, State Key Laboratory of Cotton Biology, School of Life Sciences, Henan University, 85 Minglun Street, Kaifeng 475001, China
2 Department of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

† These authors contributed equally to this work.
* Correspondence: songcp@henu.edu.cn

Received 25 August 2018; Editorial decision 28 May 2019; Accepted 29 May 2019

Editor: Robert Sharwood, Australian National University, Australia

Abstract

Pseudouridine (Ψ) is widely distributed in mRNA and various non-coding RNAs in yeast and mammals, and the specificity of its distribution has been determined. However, knowledge about Ψs in the RNAs of plants, particularly in mRNA, is lacking. In this study, we performed genome-wide pseudouridine-sequencing in Arabidopsis and for the first time identified hundreds of Ψ sites in mRNA and multiple Ψ sites in non-coding RNAs. Many predicted and novel Ψ sites in rRNA and tRNA were detected. mRNA was extensively pseudouridylated, but with Ψs being under-represented in 3′-untranslated regions and enriched at position 1 of triple codons. The phenylalanine codon UUC was the most frequently pseudouridylated site. Some Ψs present in chloroplast 23S, 16S, and 4.5S rRNAs in wild-type Col-0 were absent in plants with a mutation of Svr1 (Suppressor of variegation 1), a chloroplast pseudouridine synthase gene. Many plastid ribosomal proteins and photosynthesis-related proteins were significantly reduced in svr1 relative to the wild-type, indicating the roles of Svr1 in chloroplast protein biosynthesis in Arabidopsis. Our results provide new insights into the occurrence of pseudouridine in Arabidopsis RNAs and the biological functions of Svr1, and will pave the way for further exploiting the mechanisms underlying Ψ modifications in controlling gene expression and protein biosynthesis in plants.

Keywords: Arabidopsis thaliana, proteome, Ψ-sequence, pseudouridine, pseudouridine synthases (PUSs), suppressor of variegation 1 (Svr1), transcriptome.

Introduction

Pseudouridylation, the isomerization of uridine (U) into pseudouridine (Ψ), is the most universal post-transcriptional modification of RNA nucleosides for controlling gene expression in various cellular processes of living organisms (Hsu et al., 2017; Roundtree et al., 2017; Zhao et al., 2017). Accordingly, Ψ is considered as being the fifth ribonucleoside, and has long been known to widely exist in various ribosomal RNAs (25S, 18S, 5.8S, and 5S rRNAs), transfer RNA (tRNA), and small nuclear RNA (snRNA) with high conservation at many positions (Spenkuch et al., 2014; Li et al., 2016; De Zoysa and Yu, 2017; Adachi et al., 2019). In recent years it has been found to be ubiquitously present in mRNA in yeast, mammals...
method, and identified hundreds of pseudouridylation sites of Arabidopsis RNAs by applying a pseudouridine-sequencing method. Pseudouridylation (Ψ) in mRNA, and a number of predicted and novel Ψ sites in rRNAs, tRNAs, and other ncRNAs. We also explored the possible targets and roles of SVR1 in Arabidopsis. Our findings will pave the way for further investigations on the mechanisms and functions of Ψ modifications in plants.

Materials and methods

Plant material

Seeds of Arabidopsis thaliana wild-type (WT, Col-0) and svr1 mutants (SALK_013085, obtained from the Arabidopsis Biological Resource Center) were surface-sterilized using 0.1% HgCl₂ for 5 min, washed with sterile ddH₂O five times, and sown on Murashige–Skoog (MS) agar plate containing 3% (w/v) sucrose. After stratification at 4°C for 2 d, the seeds were germinated and the seedlings were grown in a growth chamber for 2 weeks (21/18°C, 14/10 h day/night, 80–100 μmol m⁻² s⁻¹ light intensity, ~70% relative humidity). The seedlings were then planted in nutritional soil (humus soil:vermiculite 1:1, v/v) for a further 12 d. Fully expanded young leaves were collected, immediately frozen in liquid nitrogen, and stored at −80°C.

Pseudouridine-sequencing library construction

Pseudouridine-sequencing (Pseud-seq, Ψ-sequencing) libraries were prepared according to the method of Carlile et al. (2014, 2015) with some modifications. In brief, ~6 mg total RNA was extracted from 6 g of young leaf samples according to the method of Suzuki et al. (2004). The concentrations of RNA were measured by NanoDrop 2000 (ThermoFisher Scientific) and the quality of RNA was assessed using agarose gel electrophoresis and RNA 6000 Nano Chips on an Agilent 2100 Bioanalyzer (Agilent Technologies). About 10–20 μg of high-quality mRNA was isolated from 2 mg total RNA using a Dynabeads mRNA DIRECT™ Purification Kit and RiboMinus™ Plant Kit for RNA-Seq (Invitrogen). The purity of mRNA was checked using RNA 6000 Nano Chips on an Agilent 2100 Bioanalyzer.

About 10–25 μg total RNA or mRNA were fragmented using a NEBNext Magnesium RNA Fragmentation Kit at 94°C for 6–9 min for mRNA or 10–15 min for total RNA. The reactions were quenched and the fragmented RNA was collected. Three-fifths of the RNA fragments were treated with 0.4 M N-cyclohexyl-N′-(2-morpholinoethyl)-carbodiimide methyl-p-toluene sulphonate (CMC) (Sigma) in BEU buffer (7 M urea, 4 mM EDTA, 50 mM bicine, pH 8.5) (+CMC, treatment) and two-fifths of the RNA fragments were treated with BEU buffer alone (−CMC, control). Removal of CMC from Us and Gs and dephosphorylation of RNA fragments were then carried out following the method of Carlile et al. (2015). RNA fragments of 60–100 nt were selected, eluted, and precipitated.

An adapter (5′ Phos/TGGAATTCCTCGGAGGCAAAG/3′ dDC) was synthesized by Sangon Biotech Co, Ltd (Shanghai, China). Adenylation of the adapter was performed using a 5′ DNA Adenylate Kit following the manufacturer’s instructions (New England Biolabs). Ligation of the 3′-end of the 60–100 nt RNA fragments with the pre-adenylated adapter, RNA reverse-transcription (RT) primer: 5′ Phos/GATCGTCGGACTGTAGAACTCTGAACCTGGCCTGAGTTCGAGTGGAGGACACCTGTCCGTGGTGCGGCCGTGCCATATTTC/G/T/CATATTTC3′, and selection of cDNA were then conducted according to the method of Carlile et al. (2015). About 110–140 nt cDNA fragments were selected and circularized using CircLigase™ ssDNA Ligase. PCR amplification of the cDNA was carried out using the circle DNA as templates and specific primers (forward: AATGATACGGCGACACCACGCA; reverse: CAACGACACACTGCAGAACACATACGACATATGAAAAACTGAGGTAGTCCATTTGGAGGGCGCAATCGATCTCCA; where XXXXXX is the barcode) (Carlile et al., 2015). The PCR products were then gel-purified, precipitated, and sequenced on an Illumina Hiseq 2500 with 50-bp single-end reads by Biomarker Technologies Co, Ltd (Beijing, China).
Sequencing data analysis

For analysis of the sequenced data for RNA pseudouridylation, adaptor sequences were removed using the programme cutadapt (v1.14) (Martin, 2011), and low-quality reads were trimmed using the Btrim (v0.2.0) software (Kong, 2011). The remaining sequences were then mapped to the Arabidopsis genome (https://www.arabidopsis.org) using the software TopHat2 (v 2.1.1) (Trapnell et al., 2009).

For determination of Ψ callings, the Arabidopsis gff annotation files for coding genes of proteins, rRNAs, tRNA, and other RNAs, and those for alternative splicing of mRNA precursors were downloaded from the Arabidopsis website (TAIR10). Special python scripts were made for each of the +/+,-CMC library pairs, and peak values for each position 1 nt 3′ of a U (peak position) in the RNA transcriptional fragments were calculated. The length of 5′-untranslated regions (UTRs) and 3′-UTRs of the RNA fragments was regards as 200 bp if they are not annotated. The threshold of peak values was determined according to the method of Carlile et al. (2014):

\[
\text{peak}^+ = \frac{r^+-r^-}{wr^+ + wr^-} \times ws
\]

where \( r^+ \) and \( r^- \) are the number of reads whose 5′-ends map to the position detected in the +CMC and –CMC libraries, respectively, and \( wr^+ \) and \( wr^- \) are the numbers of reads whose 5′-ends map to a window centred at the position assayed (not including reads at that position); \( ws \) is the window size (except that position). Sites with peak positions greater than a specified cut-off were defined as potential Ψs.

For all analyses, the window size was set to 150 and the peak cut-off was set to 1.0. The coverage depth for average reads in windows was at least 5, and the numbers of reads whose 5′-ends mapped to a window was at least 5. A Ψ site that occurred in all three replicates was considered as a mapped site.

Reproductivity analysis of results between independent samples

Pseudo-seq of WT total RNA was repeated three times. The peak values for each putative Ψ site for and other sites in 25S, 18S, and 5.8S rRNAs in each replicate experiment were obtained according to the methods described above. For comparison of the reproducibility of the detected Ψ sites between two replicated samples, all the peak values obtained in the two replicated experiments were visualized in scatterplots.

Validation of mRNA and rRNA pseudouridylation by qPCR

For confirmation of the Pseudo-seq results, changes in melting temperature of quantitative real-time PCR (qPCR) products from 10 mRNA and 18 rRNA fragments containing Ψ sites, and two mRNA and two rRNA harboring no Ψ sites were examined after CMC treatment following Lei and Yi (2017). Total RNA was extracted from leaves of 24-d-old Arabidopsis WT seedlings according to the methods described above. mRNA was isolated using a Dynabeads® mRNA DIRECT™ Purification Kit. Half of the fragmented total RNA and mRNA were treated with 0.4 M CMC as described above for the library preparation. The other half of the RNA was treated with the buffer without CMC. cDNA was synthesized from the RNA using Random Hexamer primer (TaKaRa) and SuperScript II reverse transcriptase (Invitrogen) following Lei and Yi (2017). qPCR experiments were carried out using the cDNA, TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus) (TaKaRa), specific primers (Supplementary Table S1 at PubMed online), and a Lightcycler 480 II real-time PCR system (Roche). High-resolution melting analysis was conducted using the LightCycler® 480 II software according to the method of Lei and Yi (2017).

Analysis of snoRNA-guided Ψ modifications

To examine the potential pseudouridylation sites guided by Box H/ACA snoRNA, Arabidopsis snoRNA data were obtained from the Plant snoRNA Database (http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home/) and from Chen and Wu (2009) (Supplementary Table 3). The secondary structures of the snoRNAs were obtained using the software Mfold with default parameters (http://unafold.rna.albany.edu/?q=mfold; application run by Biomarker Technologies Co, Ltd), and the sequences containing Ψ sites that matched the known target sequences of Arabidopsis snoRNAs were searched in the Ψ-seq results.

Phylogenetic analysis of PUSs

The PUS genome sequence and gene annotation databases were downloaded from the yeast (https://www.yeastgenome.org/) and Arabidopsis (http://www.arabidopsis.org) databases. The full-length amino acid sequences of PUS proteins were aligned using the ClustalW software with default parameters (Larkin et al., 2007) and a phylogenetic tree was then constructed following the alignment results using the neighbor-joining method and 1000 bootstrap trials with the MEGA 5.0 software (http://www.megasoftware.net).

Proteome analysis

Proteins were extracted from 1 g leaf material of Arabidopsis seedlings and re-dissolved in buffer (8 M urea, 100 mM triethylammonium bicarbonate, pH 8.0). A 2-D Quant kit (GE Healthcare) was used to measure protein concentrations. The proteins were then digested with trypsin, and peptides were labeled with tandem mass tag (TMT). The dynamic changes of the whole proteome were quantified by applying an integrated approach involving TMT labeling and LC-MS/MS. The resulting MS/MS data were processed by the MaxQuant with an integrated Andromeda search engine (v1.5.2.8). Tandem mass spectra were searched against the Uniprot Arabidopsis thaliana database (https://www.uniprot.org/) concatenated with a reverse decay database. Trypsin/P was specified as the cleavage enzyme allowing up to two missing cleavages. The mass error was set to 10 ppm for precursor ions and 0.02 Da for fragment ions. Carbamidomethylation on Cys was specified as the fixed modification, and oxidation on Met and acetylation on the protein N-terminal were specified as variable modifications. The false discovery rate thresholds for proteins, peptides, and modification sites were specified at 1%. The minimum peptide length was set at 7. For the quantification method, TMT-6plex was selected. All the other parameters in the MaxQuant were set to default values (PTM-Biolabs Co., Ltd).

Gene ontology (GO) analysis was conducted for categorization of the proteins encoded by the pseudouridylated mRNA by applying a GO annotation module (http://www.arabidopsis.org) and the agriGO program with default settings (FDR<0.05) (Tian et al., 2017). KEGG pathway analysis (http://www.genome.jp/kegg/) was performed to determine functional categories of the annotations of SVR1-regulated proteins (Kanehisa et al., 2016).

Results

Pseudouridine profiling analysis of total RNA and mRNA in Arabidopsis

To identify Ψ modifications in Arabidopsis ncRNAs and mRNA at transcriptome-wide level, we prepared high-quality cDNA libraries from total RNA and mRNA from leaves of plants grown under normal conditions and then performed Illumina RNA-sequencing using the single-nucleotide-resolution Ψ-seq method (Carlile et al., 2014). The Ψ-seq procedures were developed on the basis of the premature termination of RNA reverse transcription at one nucleotide to the 3′-side of the pseudouridylated site that was generated by CMC treatment prior to cDNA synthesis (Carlile et al., 2014). Twelve cDNA libraries for total RNA (CMC-treated/+CMC, CMC-untreated–/-CMC) and mRNA (CMC-treated/+CMC, CMC-untreated–/-CMC) were constructed and sequenced. Approximately 43.9, 47.3, 36.3, and 37 million reads were obtained from the samples of CMC-treated total RNA, CMC-untreated total RNA, CMC-treated mRNA and
CMC-untreated mRNA, respectively (Supplementary Table S2), and aligned to the Arabidopsis reference genome (TAIR10). All the data were submitted to NCBI (No. SRP156413).

By comparing the reads from a +CMC sample with those from its corresponding −CMC one (the mock control), Ψ modifications in vivo were identified in the RNAs. Based on the similar stringent criteria as described by Carlile et al. (2014), a total of 467 and 451 Ψ sites were identified in ncRNAs and mRNA, respectively. In total, 187 Ψ sites were detected in rRNA, 232 in tRNA, 13 in snRNA, 22 in snoRNA, and 13 in other RNA (Fig. 1, Supplementary Tables S3, S4).

Pseudouridines are widely distributed in mRNA

To map the pseudouridylation profiling of Arabidopsis mRNA in detail, we calculated the ratio of reads ending one nucleotide upstream in the library from CMC-treated mRNA to their corresponding reads in the library from the same RNA without CMC treatment (Carlile et al., 2014). A total of 451 Ψs were identified to be dispersely distributed in 332 gene transcripts (Supplementary Table S4), which accounted for 1.21% of all detected mRNAs (there are 27,416 genes in Arabidopsis according to TAIR10). In total, 53 Ψs were present in the 5′-UTRs, 374 were present in coding sequences (CDS), and 24 were present in the 3′-UTRs (Fig. 2A), individually accounting for 11.75%, 82.93%, and 5.32% of the total Ψ number in mRNA, respectively. Intriguingly, two Ψ sites were characterized within the translation initiation codon AUG while no Ψs were detected in the translation termination codons (Supplementary Table S5).

To explore whether Ψs were unbiasedly distributed along mRNA sequences, we compared their distribution with that of Us in 5′-UTRs, CDS, and 3′-UTRs using χ² tests. Significant differences were found between the enrichments of Ψs and Us in the three regions of the transcripts. The content of Ψs was clearly lower than that of Us in the 5′-UTR while the opposite was found in the CDS and 5′-UTR (P<10⁻⁵) (Fig. 2B), indicating that Ψs preferentially occur in the coding sequences and 5′-UTR of mRNAs in Arabidopsis. Next, we calculated the frequencies of Ψ occurrence in triplet codons of mRNA, and found that UUC, CUU, UUU, and UCU were frequently pseudouridylated (Fig. 2C). Moreover, the frequency of Ψ occurrence was higher in the first U than in the second one when two U bases appeared sequentially in a codon, except for

![Fig. 1. Pseudouridines detected in ncRNAs and mRNA of Arabidopsis. Pseudo-seq reads are shown for (A) 25S rRNA (3000-3200), (B) 18S rRNA (AT3G41768.1, 283-483), (C) tRNA (AT3G50505.1, 1-72), (D) snRNA (AT3G57645.1, 470-670), (E) snoRNA (AT5G66564.1, 1-84) and (F) mRNA (AT1G76180.1, 121-321). The units of the y-axes are reads per min. CMC-dependent peaks of reads are indicated by a dashed red line. The peak values are means (±SE), n=3. (G) Distribution of detected Ψs in various ncRNAs and mRNA.](image-url)
Mapping of pseudouridine profiles in Arabidopsis thaliana RNAs

UUU (Fig. 2C). These data implied that positional bias of Ψs in codons exist in Arabidopsis. To refine the outcome of this analysis, the probability of Ψs occurring in each position of the triplet codon within the coding region was also compared with that of Us using χ² tests. This revealed that significantly more Ψs relative to Us were distributed at position 1 than at positions 2 and 3 (P<0.05) (Fig. 2D). These findings suggested that the base U in position 1 of the codon is preferentially pseudouridylated in Arabidopsis.

To gain insights into the putative roles of Ψs in mRNA, we examined the functions of proteins encoded by the pseudouridylated mRNA. The results showed that the enriched proteins were mainly involved in responses to stimuli or stress, metabolic processes, biosynthetic processes, energy generation, and photosynthesis (Fig. 3; Supplementary Table S6), suggesting that pseudouridylation of mRNA may play important roles in these processes in Arabidopsis.

Methodological validation of the pseudouridylated sites detected by Pseudo-seq

To check the reliability of the Pseudo-seq analysis for the pseudouridylated sites that were determined, three sets of experiments were conducted to test the putative sites of pseudouridylation. In the first set of experiments, we examined the total RNA-sequencing data to define the conserved known Ψs. When we constructed a library from CMC-treated total RNA, and the ncRNAs remaining in the samples allowed us to use the predicted Ψs in rRNA or tRNA as a set of internal positive controls. As previously reported, there are 32 predicted Ψ sites in 25S rRNA, 25 in 18S, and one in 5.8S in Arabidopsis (Brown et al., 2003; Chen and Wu, 2009) and 23 Ψ sites in eukaryotic tRNA (Björk et al., 1987). Of these, we identified 21/32, 14/25, 1/1, and 13/23 in 25S rRNA, 18S rRNA, 5.8S rRNA, and tRNA, respectively (Supplementary Fig S1, S2, Supplementary Table S7). These results were in agreement with the predicted data, although some Ψ sites were not found under our experimental conditions (Björk et al., 1987; Brown et al., 2003; Chen and Wu, 2009). Notably, we discovered 26 novel Ψ sites in 25S rRNA, 15 in 18S rRNA, one in 5.8S rRNA, and 12 in tRNA.

We next examined the reproducibility of the results between two independent samples by comparing the peak values for Ψ sites with those for other sites in rRNAs across pairs of conditions. Scatterplots of the Ψ peak values from the two samples were highly correlated, and clear differences in distribution existed between the values for Ψ sites and those for other sites (Supplementary Fig. S3). The high experimental repeatability provided a good validation of the putative Ψ sites.
In the third set of experiments, we used mutants of a chloroplast pseudouridine synthase SVR1 as a negative control for validation of our deep-sequencing data (Yu et al., 2008). Theoretically, disruptions in PUSs should lead to marked decreases of pseudouridylation of rRNA and tRNA, which result in growth inhibition in yeast and T. gondii, and serious diseases in humans (Zebarjadian et al., 1999; Hoareau-Aveilla et al., 2008; Anderson et al., 2009; Fujiwara and Harigae, 2013). Parallel Pseudo-seq runs of both WT and svr1 mutants were performed in order to compare the reads from CMC-treated samples with those from the corresponding CMC-untreated ones. In total, 15 and four Ψ sites were identified in chloroplast rRNA in the WT and svr1, respectively (Supplementary Fig. S4, Supplementary Table S8). Among these, four Ψ modifications that were detected in the WT chloroplast 23S rRNAs, six in 16S rRNAs, and one in 4.5S rRNAs were absent in the mutant (Supplementary Table S8). By contrast, no significant differences in Ψ sites were found in 25S rRNA and 18S rRNA between the WT and the svr1 plants (Supplementary Fig. S5).

Validation of mRNA and rRNA pseudouridylation using a qPCR-based method for Ψ site recognition

To further validate the putative Ψ sites in ncRNAs and mRNA, we applied a qPCR-based method for locus-specific detection of pseudouridine, which was based on Ψ-CMC-induced mutation/deletion in cDNA synthesis causing read-through qPCR products with different melting temperatures (Lei and Yi, 2017). We selected 11 Ψ sites as targets in the mRNA (AT1G20620, Ψ-477, Ψ2600, Ψ3215; AT1G76180, Ψ221; AT2G41100, Ψ-157; AT3G01500, Ψ892; AT3G04640, Ψ38; AT3G45140, Ψ3384; AT3G56795, Ψ51; ATCG01020, Ψ113) and two negative control sites (AT1G20620, 473–568 bp; AT2G41100, 800–889 bp), and examined the melting curves of the related qPCR products from samples with and without CMC treatment. We found that the melting curves in each CMC-treated mRNA sample containing a putative Ψ site were significantly different from those in the corresponding CMC-untreated sample, while the melting curves of qPCR products from mRNA without Ψs in the +CMC samples were quite similar to those in the –CMC samples (Fig. 4). We also examined the melting curves of a total of 10 qPCR products from 25S rRNA (Ψ783, Ψ973, Ψ1060, Ψ2339, Ψ2489, Ψ2965) and 18S rRNA (Ψ761, Ψ1000, Ψ1486, Ψ1634), and two negative control fragments (203–287 bp in 25S rRNA and 498–585 bp in 18S rRNA). Clear alterations in melting temperature were observed in all the curves from samples containing Ψs compared with Ψ-free samples (Fig. 5). Collectively, these results suggested that the detected Ψ modifications did indeed exist in these mRNA and rRNA molecules in Arabidopsis.

The melting curves of the qPCR data from 23S rRNA (Ψ1346, Ψ2623) and 16S rRNA (Ψ49, Ψ1159) fragments of the WT and svr1 after CMC treatment were compared with the corresponding curves with no CMC treatment. The melting temperatures for the WT samples were markedly different, whilst those from the svr1 mutant were similar (Fig. 6), indicating that the identified Ψs are catalysed by SVR1 in Arabidopsis.

Possible mechanisms for isomerization of Us to Ψs in Arabidopsis RNAs

Ψs are formed through isomerization of Us by PUSs and RNA-dependent mechanisms. Ten Ψ-catalysing enzymes including nine PUSs (PUS1–9) and Cbf5p are found in yeast (Rintala-Dempsey and Kothe, 2017). Among these, PUS4 and PUS7 have been determined to specifically recognize the core consensus sequences ‘GUΨCNANYCY’ and ‘UGΨAG’, respectively (Lovejoy et al., 2014; Schwartz et al., 2014; Rintala-Dempsey and
We therefore examined all the pseudouridylated RNA harboring these specific sequences in Arabidopsis. A total of 11 and 67 putative PUS4 targets were present in mRNA and tRNA, respectively, and five \( \Psi \)s were putative PUS7 targets in tRNA (Supplementary Table S9), indicating that these \( \Psi \)s are likely to be catalysed by homologs of PUS4 and PUS7 in Arabidopsis (Supplementary Fig. S6).

To determine whether \( \Psi \)s within rRNAs and mRNA were synthesized by snoRNA-guided mechanisms, we obtained 50 Arabidopsis H/ACA snoRNAs and predicted their secondary structure, and analysed the H/ACA box‐targeting sequences. In total, four unique sites were identified in 25S rRNA, five in 18S rRNA, and six in mRNA (Supplementary Fig. S7). These sites perfectly matched to the canonical H/ACA snoRNA targets, suggesting that snoRNA‐dependent Cbf5p orthologs in Arabidopsis are responsible for the formation of these \( \Psi \)s (Supplementary Fig. S6).

Possible biological roles of \( \Psi \) modifications of RNAs in Arabidopsis

To further understand the putative roles of \( \Psi \)s in Arabidopsis rRNA and tRNA, we examined the locations of the \( \Psi \)s in these RNAs. We found that \( \Psi \)1000, \( \Psi \)1104, and \( \Psi \)1118 were located in the decoding site of the 40S ribosome subunit and that \( \Psi \)2844, \( \Psi \)2855, \( \Psi \)2870, \( \Psi \)2884, \( \Psi \)2945, and \( \Psi \)2965 were in the peptidyltransferase center and A‐site of 60S ribosome subunit (figure 3 in Sloan et al., 2017). The decoding sites,
center, and A-site are functionally important regions and essential for protein synthesis (Sloan et al., 2017). Hence, these Ψ sites may play pivotal roles in protein translation in Arabidopsis.

In addition, Ψs were observed to occur in positions 13 (D stem), 27 (anticodon loop), 38 (anticodon loop), 39 (anticodon stem), and 55 (TΨC stem-loop) in tRNA (Supplementary Fig. S2B). These Ψs have been determined to play significant roles in stabilization of the tRNA secondary and tertiary structure, and to contribute to the accurate and efficient decoding ability of tRNAs in yeast (Charette and Gray, 2000; Lorenz et al., 2017). We calculated the frequency of Ψ sites detected in each nucleotide of tRNA and found that the most frequently pseudouridylated sites occurred at positions 27, 38–40, and 54–56 (Fig. 7). All of these sites except for 56 were previously predicted and are highly conserved in eukaryotic tRNA (Björk et al., 1987).

To examine whether the mutation in SVR1 altered protein synthesis, differences in leaf proteomes between the WT and svr1 were analysed by TMT labeling and LC-MS/MS methods. The results revealed that the expression of 155 proteins was clearly reduced in svr1 relative to the WT (ratio<0.667, P<0.05) (Fig. 8). It was notable that the abundances of 53 chloroplast ribosomal proteins and 63 photosynthesis-associated proteins were clearly decreased in svr1. The ribosomal proteins included RPL35, rpl36, RPL28, RPL27, rpl2-A, and RPL9, and the photosynthesis proteins included eight psa proteins (psaA-H),
Mapping of pseudouridine profiles in Arabidopsis thaliana RNAs

six psb proteins (psbA-F), psbL, ribulose bisphosphate carboxylase large chain (rbCL), ribulose bisphosphate carboxylase small chain 1A (RBCS-1A), RBCS-1B, RBCS-3B, ATP6, and petC (Rieske Fe-S). The two classes of proteins accounted for nearly three-quarters of all reduced proteins in the mutant (Fig. 8; Supplementary Table S10) (all data have been submitted to EMBL-EBI, No. PXD011814). Among them, psaF, rbCL, ATP6, and petC have been reported to be reduced in svr1 compared with the WT (Yu et al., 2008). These findings indicated that loss of function of SVR1 inhibited the biosynthesis of many ribosome and photosynthesis-related proteins, thus regulating protein translation in Arabidopsis.

Discussion

In this study, we identified for the first time hundreds of Ψ sites within mRNA in Arabidopsis using the Ψ-seq method (Carlile et al., 2014). In addition, dozens of Ψ sites in rRNAs, tRNAs, and other ncRNAs were also detected (Supplementary Tables S3, S4). These results imply that RNA pseudouridylation may be ubiquitous not only in heterotrophic organisms but also in autotrophic organisms.

The Ψ-seq method described by Carlile et al. (2014, 2015) has been shown to be successful for finding Ψs in RNA in planta, and several aspects of our results confirmed that it is appropriate for use in Arabidopsis. Firstly, we detected 21 out of 32 predicted Ψ sites in 25S rRNA, 14/25 sites in 18S rRNA, 1/1 in 5.8S rRNA, and 13/23 in tRNA (Supplementary Figs S1, S2, Supplementary Table S7; Björk et al., 1987; Brown et al., 2003; Chen and Wu, 2009). Our identification of the evolutionarily conserved sites of pseudouridylation from ncRNAs suggested that the CMC treatment of rRNA and tRNA was effective, and this could be used as a criterion for identification of Ψs with high confidence. Secondly, the good correlation of the peak values for the characterized Ψ sites in rRNAs between two independent sequencing runs verified the repeatability of the method. The distributions of the Ψ sites on scatterplots...
were significantly different from those of other sites in rRNAs (Supplementary Fig. S3). A certain minimum number of reads for the residues of interest in the sequencing data had been set to minimize the false positive discovery signal (Nakamoto et al., 2017). Thirdly, validation was provided by the qPCR-based method for locus-specific pseudouridine detection, which is dependent on mutations/deletions in cDNA synthesis caused byΨ-CMC leading to qPCR products with altered melting temperatures (Lei and Yi, 2017). Our results showed that the melting curves of the qPCR products from 10 mRNA and 10 rRNA fragments containing detected Ψs with CMC treatment clearly differed from those without CMC treatment, while the curves from four negative RNA fragments were very similar between the +CMC and –CMC samples (Figs 4, 5), implying that the identified pseudouridylation modifications genuinely occurred in rRNAs and mRNA. Finally, disruption of the chloroplast pseudouridine synthase gene SVR1 caused marked decreases in the levels of mature chloroplast 23S, 16S, 5S, and 4.5S rRNAs (Yu et al., 2008). We consistently found that four, six, and one Ψ sites individually identified in 23S, 16S, and 4.5S rRNAs, respectively, in the WT were not present in the svr1 mutant (Supplementary Table S8). However, no significant differences in Ψ modifications were observed in 25S rRNA and 18 rRNA (Supplementary Fig. S5). Collectively, these data suggest that Ψ-seq is a powerful tool for the accurate detection of Ψ modifications in rRNAs of Arabidopsis.

We examined naturally occurring mRNA pseudouridylation sites in Arabidopsis and found that Ψs were widely distributed in 5′-UTRs, CDS, and 3′-UTRs of mRNA (Fig. 2A; Supplementary Table S4). Consistent with previous results obtained from yeast, mammals (including humans), and T. gondii (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014; Li et al., 2015; Nakamoto et al., 2017), the level of Ψ modifications relative to U in the 3′-UTR was markedly lower than that in the 5′-UTR and CDS region, i.e. Ψ is under-represented in 3′-UTRs of Arabidopsis mRNA (Fig. 2B). Furthermore, we observed that the probability of pseudouridylation at position 1 of the triple codon was clearly higher than that at positions 2 or 3 within the coding region (Fig. 2D), in accordance with previous data from humans and T. gondii (Li et al., 2015; Nakamoto et al., 2017). Mapping of Ψ sequences has shown that the valine codon (GUA) and phenylalanine codons (UUU and UUC) are the most frequently pseudouridylated in yeast and humans, respectively (Carlile et al., 2014; Li et al., 2015). We also found that UUU and UUC were the most frequently modified in Arabidopsis (Fig. 2C). Taken together, these findings suggest that biased distributions of Ψs in general regions and in specific codon positions of mRNA may be a common feature not only in animals but also in plants, implying that a sequence-specific mechanism for Ψ modifications within mRNA may be conservative between both groups.

In total, we found 451 Ψ sites in Arabidopsis mRNA (Supplementary Table S4), a number comparable to that in yeast (50–100) and human cells (100–400). No pseudouridylation was detected in the translation termination codons of Arabidopsis mRNA, although artificial modification of Ψ on yeast mRNA allows conversion of nonsense stop codons into sense codons (Karijolich and Yu, 2011). Our GO analysis results showed that the Ψ-containing transcripts that were enriched mainly played roles in responses to stimuli and stress, and in metabolic processes, biosynthetic processes, energy generation, and photosynthesis (Fig. 3; Supplementary Table S6),
indicating that mRNA pseudouridylation may be essential for these processes in Arabidopsis.

In summary, our analysis of Ψ-seq data in Arabidopsis showed that Ψ sites were highly biased towards the 3′-end of tRNAs. Our selected size of RNA fragments was at least 60 nt, and mature tRNAs are only about 70–100 nucleotides long (Torres et al., 2014). Thus, many RT stop products that resulted from Ψs at 3′-ends of tRNAs would have been too short for read-mapping, and some Ψs that occurred at the 3′-end of tRNAs might not have been detected. In addition, we found that the expression levels of 53 chloroplast ribosomal proteins and 63 photosynthesis-associated proteins were significantly reduced in the mutant compared with the WT (Fig. 8; Supplementary Table S10). Among the photosynthesis-related proteins, four (psaF, rbcL, ATPα, and petC) have been previously reported to be reduced in svr1 (Yu et al., 2008). Our results add significantly to the findings by Yu et al. (2008) and suggest that SVR1 plays a pivotal role in modulating the translation of many chloroplast proteins, including ribosomal proteins and those associated with photosynthesis (Fig. 8). Our discovery of pseudouridylation of mRNA in Arabidopsis raises the question of the significance of modified nucleosides in mRNA in relation to coordination of responses to environmental cues, and this should be the focus of further research.

Supplementary data

Supplementary data are available at JXB online.

Fig. S1. Positions of Ψs in 25S rRNA in Arabidopsis.

Fig. S2. Positions of Ψs in 18S rRNA and tRNA in Arabidopsis.

Fig. S3. Correlations of detected Ψ sites and other sites in rRNAs between two replicated experiments.

Fig. S4. Peak plots and coverage plots for chloroplast rRNA in the wild-type and the svr1 mutant.

Fig. S5. Peak plots and coverage plots for 25S rRNA and 18S rRNA in the wild-type and the svr1 mutant.

Fig. S6. Phylogenetic relationships of PUSs between yeast and Arabidopsis.

Fig. S7. Putative Ψ sites in rRNAs and mRNA guided by H/ACA snoRNA.

Table S1. Specific primers used in qPCR.
Table S2. Pseudo-seq profiles of total RNA and mRNA.
Table S3. Detected Ψ sites in various ncRNAs.
Table S4. Ψ sites in mRNA.
Table S5. Ψ sites present in initiation codons.
Table S6. GO enrichment of Ψ-containing mRNAs in the wild-type.
Table S7. Comparison of predicted Ψ sites with detected Ψ sites in rRNAs and tRNA.
Table S8. Comparison of Ψs in chloroplast rRNA between the wild-type and the svr1 mutant.
Table S9. Putative Ψ targets of homologs of PUS4 and PUS7 in Arabidopsis.
Table S10. Reduced proteins in svr1 relative to the wild-type.

Acknowledgements

We thank Professor Wendy V. Gilbert and Dr Thomas M. Carlile (Department of Biology, Massachusetts Institute of Technology, USA) as well as Professor Zhen Su and Wen Ying Xu (China Agricultural University) for kindly providing us with experimental procedures and for giving us many informative suggestions. We also thank Dr Aimin Chen of the School of Mathematics and Dr Yuge Li of the School of Life Sciences in Henan University for their help. This work was supported by the Ministry of Agriculture of China (2016ZX08009-003) and National Program on Key Basic Research Project (973 Program, 2012CB1143001).

Author contributions

LS and C-PS designed the experiments; LS, XB, HZ, and HD performed the experiments; LS, XY, SB, and WW analysed the sequencing data; C-PS, FH, and XZ wrote the manuscript.

References

Adachi H, De Zoysa MD, Yu YT. 2019. Post-transcriptional pseudouridylation in mRNA as well as in some major types of noncoding RNAs. Biochimica et Biophysica Acta 1862, 230–239.
Anderson MZ, Brewer J, Singh U, Boothroyd JC. 2009. A pseudouridine synthase homologue is critical to cellular differentiation in Toxoplasma gondii. Eukaryotic Cell 8, 398–409.
Björk GR, Ericson JU, Gustafsson CE, Hagervall TG, Jönsson YH, Wikström PM. 1997. Transfer RNA modification. Annual Review of Biochemistry 56, 263–287.
Brown JW, Echeverria M, Qu LH, Lowe TM, Bachelerie JP, Hüttenhofer A, Kastenmayer JP, Green PJ, Shaw P, Marshall DF. 2003. Plant snoRNA database. Nucleic Acids Research 31, 432–435.
Carlile TM, Rojas-Duran MF, Gilbert WV. 2015. Pseudo-Seq: genome-wide detection of pseudouridine modifications in RNA. Methods in Enzymology 560, 219–245.
Carlile TM, Rojas-Duran MF, Zinzhteyn B, Shin H, Bartoli KM, Gilbert WV. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515, 143–146.
Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. IUBMB Life 49, 341–351.
Chen HM, Wu SH. 2009. Mining small RNA sequencing data: a new approach to identify small nuclear RNAs in Arabidopsis. Nucleic Acids Research 37, e69.
De Zoysa MD, Yu YT. 2017. Posttranscriptional RNA pseudouridylation. The Enzymes 41, 151–167.
Fujiwara T, Harigae H. 2013. Pathophysiology and genetic mutations in congenital sideroblastic anemia. Pediatrics International 55, 675–679.
Hoareau-Aveilla C, Henry Y, Leblanc T. 2008. Dyskeratosis congenita, a disease caused by defective telomere maintenance. Médecine/Sciences 24, 390–398.
Hsu PJ, Shi H, He C. 2017. Epitranscriptomic influences on development and disease. Genome Biology 18, 197.
Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44, D457–D462.
Karjiloch J, Yu YT. 2011. Converting nonsense codons into sense codons by targeted pseudouridylation. Nature 474, 395–398.
Kong Y. 2011. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98, 152–163.
Larkin MA, Blackshields G, Brown NP, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lei Z, Yi C. 2017. A radiolabeling-free, qPCR-based method for locus-specific pseudouridine detection. Angewandte Chemie International Edition 56, 14878–14882.
Li X, Ma S, Yi C. 2016. Pseudouridine: the fifth RNA nucleotide with renewed interests. Current Opinion in Chemical Biology 33, 108–116.

Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C. 2015. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nature Chemical Biology 11, 592–597.

Lorenz C, Lüns CE, Mörl M. 2017. tRNA modifications: impact on structure and thermal adaptation. Biomolecules 7, 35.

Lovejoy AF, Riordan DP, Brown PO. 2014. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS ONE 9, e110799.

Lu S, Li C, Zhang Y, Zheng Z, Liu D. 2017. Functional disruption of a chloroplast pseudouridine synthase desensitizes arabidopsis plants to phosphate starvation. Frontiers in Plant Science 8, 1421.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads EMNet.journal 17, 10–12.

Nakamoto MA, Lovejoy AF, Cygan AM, Boothroyd JC. 2017. mRNA pseudouridylation affects RNA metabolism in the parasite Toxoplasma gondii. RNA 23, 1834–1849.

Rintala-Dempsey AC, Kothe U. 2017. Eukaryotic stand-alone pseudouridine synthases − RNA modifying enzymes and emerging regulators of gene expression? RNA Biology 14, 1185–1196.

Roundtree IA, Evans ME, Pan T, He C. 2017. Dynamic RNA modifications in gene expression regulation. Cell 169, 1187–1200.

Schwartz S, Bernstein DA, Mumbach MR, et al. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA, Cell 159, 148–162.

Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT. 2017. Tuning the ribosome: the influence of tRNA modification on eukaryotic ribosome biogenesis and function. RNA Biology 14, 1138–1152.

Spenkuch F, Motorin Y, Helm M. 2014. Pseudouridine: still mysterious, but never a fake (uridine)! RNA Biology 11, 1540–1554.

Suzuki Y, Kawazu T, Koyama H. 2004. RNA isolation from siliques, dry seeds, and other tissues of Arabidopsis thaliana. BioTechniques 37, 542, 544.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community. Nucleic Acids Research 45, W122–W129.

Torres AG, Batlle E, Ribas de Pouplana L. 2014. Role of tRNA modifications in human diseases. Trends in Molecular Medicine 20, 306–314.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Yu F, Liu X, Alsheikh M, Park S, Rodermel S. 2008. Mutations in SUPPRESSOR OF VARIEGATION1, a factor required for normal chloroplast translation, suppress var2-mediated leaf variegation in Arabidopsis. The Plant Cell 20, 1786–1804.

Zebbarjadian Y, King T, Fournier MJ, Clarke L, Carbon J. 1999. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. Molecular and Cellular Biology 19, 7461–7472.

Zhao BS, Roundtree IA, He C. 2017. Post-transcriptional gene regulation by mRNA modifications. Nature Reviews. Molecular Cell Biology 18, 31–42.