

Article

Genome-Wide Identification of U-To-C RNA Editing Events for Nuclear Genes in *Arabidopsis thaliana*

Ruchika ¹, Chisato Okudaira ¹, Matomo Sakari ¹ and Toshifumi Tsukahara ^{1,2,*} 

¹ Area of Bioscience and Biotechnology, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi City, Ishikawa 923-1292, Japan; mirandianruchika@gmail.com (R.); s1910053@jaist.ac.jp (C.O.); m-sakari@jaist.ac.jp (M.S.)

² Division of Transdisciplinary Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi City, Ishikawa 923-1292, Japan

* Correspondence: tsukahara@jaist.ac.jp; Tel.: +81-761-51-1640

Abstract: Cytosine-to-Uridine (C-to-U) RNA editing involves the deamination phenomenon, which is observed in animal nucleus and plant organelles; however, it has been considered the U-to-C is confined to the organelles of limited non-angiosperm plant species. Although previous RNA-seq-based analysis implied U-to-C RNA editing events in plant nuclear genes, it has not been broadly accepted due to inadequate confirmatory analyses. Here we examined the U-to-C RNA editing in *Arabidopsis* tissues at different developmental stages of growth. In this study, the high-throughput RNA sequencing (RNA-seq) of 12-day-old and 20-day-old *Arabidopsis* seedlings was performed, which enabled transcriptome-wide identification of RNA editing sites to analyze differentially expressed genes (DEGs) and nucleotide base conversions. The results showed that DEGs were expressed to higher levels in 12-day-old seedlings than in 20-day-old seedlings. Additionally, pentatricopeptide repeat (PPR) genes were also expressed at higher levels, as indicated by the log₂FC values. RNA-seq analysis of 12-day- and 20-day-old *Arabidopsis* seedlings revealed candidates of U-to-C RNA editing events. Sanger sequencing of both DNA and cDNA for all candidate nucleotide conversions confirmed the seven U-to-C RNA editing sites. This work clearly demonstrated presence of U-to-C RNA editing for nuclear genes in *Arabidopsis*, which provides the basis to study the mechanism as well as the functions of the unique post-transcriptional modification.

Keywords: uridine-to-cytidine RNA editing; RNA-seq; *Arabidopsis thaliana*; differentially expressed genes (DEGs)



Citation: Ruchika, Okudaira, C.; Sakari, M.; Tsukahara, T. Genome-Wide Identification of U-To-C RNA Editing Events for Nuclear Genes in *Arabidopsis thaliana*. *Cells* **2021**, *10*, 635. <https://doi.org/10.3390/cells10030635>

Academic Editor: Emanuela Felley-Bosco

Received: 1 February 2021

Accepted: 11 March 2021

Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

RNA editing, one of the most promising means of post-transcriptional gene regulation, has been widely investigated in various protozoa [1], mammalian apolipoprotein-B [2], animals [3], fungi [4], bacteria [5,6], and viruses [7,8] as well as in plants [9–11]. A-to-I (Inosine) RNA editing is observed in animal nuclear genes, while C-to-U RNA editing is not limited to animals but is also spreading in plant organelles. The mechanism of cytidine-to-uridine (C-to-U) RNA editing in plant organelles is completely different from that in animal nucleus but also reasonably well understood, mainly owing to the characterization of many RNA editing factors in model systems such as *Arabidopsis thaliana* and *Physcomitrella patens* [12]. In flowering plants, the RNA editing machinery, collectively described as the editosome, consists of at least four proteins including pentatricopeptide repeat (PPR) protein, Multiple Organellar RNA editing factor (MORF)/RNA editing factor interacting protein (RIP), Organelle RNA Recognition Motif (ORRM) proteins, and organelle zinc-finger protein (OZ1).

PPR proteins constitute a large family of proteins, with more than 400 members [13] and are either directly or indirectly responsible for RNA editing [14–16]. Direct selection of target sites is governed by PLS subclass PPR proteins with additional E1 and E2

domains only or further C-terminal DYW domain, which is most likely to catalyze C to U deamination.

In addition to PPR proteins, MORF/RIP, ORRM, and OZ proteins are also required for successful RNA editing and play an important role in the editosome [17]. In the *morfl* loss-of-function mutant, a single amino acid substitution in the conserved MORF domain abrogates the interaction of MORF1 with many PLS-class PPR proteins, implying that direct interaction with PPR proteins is required for the RNA editing function of MORF1 [18]. In *P. patens*, the upstream PPR stretch for RNA recognition linked in *cis* to the downstream E1, E2, and DYW domains is evident in all editing factors. Because of the simplicity of this model system, all organelle editing sites in the moss have been assigned to their corresponding DYW-type editing factors [19]. Reconstitution of target site-specific C-to-U RNA editing in *E. coli* cell as well as *in vitro* with a single DYW-type RNA editing factor from *Physcomitrella patens* suggests the DYW domain catalyzes the cytidine deamination.

While C-to-U RNA editing occurs in chloroplasts and mitochondria of the majority of terrestrial plants, U-to-C RNA editing is rare in land plants, except in hornworts, lycophytes, and ferns, and is, therefore, referred to as an occasional phenomenon [20]. Because of its rare occurrence only in non-model plants, negligible research has been done to explore the mechanism involved in U-to-C RNA editing. Recent finding of novel types of DYW domain, which are limited to species having U-to-C editing, implies that the domains are somehow involved in amination of uridines in plant organelles [21,22]. Therefore, this study was more centered toward the expressed PPR genes. PPR proteins are involved in RNA editing of organellar transcripts. However, their expression and functional role as the editing factors at the nuclear level further need to be uncovered.

In contrast to organellar RNA editing, RNA editing in nuclear genes of plants has not been widely accepted, though it has been suggested by few analyses based on the RNA-seq data. Recently, we also reported U-to-C and A-to-guanosine or inosine (G or I) nucleotide conversions in 12-d-old whole seedlings and leaves of 21-d-old seedlings, respectively [23,24]. However, direct comparison of DNA and cDNA sequencing from the same plant material, which is indispensable to eliminate the possibility of DNA mutations or sequencing errors, was not conducted.

Here, we examined the U-to-C RNA editing in 12-day- and 20-day-old seedlings of *Arabidopsis thaliana*, which showed distinct RNA editing status at least at a single site in previous analysis. RNA-seq data can be used for sequence differences relative to the reference genome to identify both genomic SNPs and RNA editing events. The major challenge in identifying U-to-C RNA editing sites using RNA-seq data is the discrimination of RNA editing sites from genome-encoded, single-nucleotide polymorphisms and technical artifacts caused by sequencing or read-mapping errors. We comprehensively analyzed all candidates for U-to-C RNA editing by Sanger sequencing and confirmed the presence of genuine U-to-C RNA editing events in *Arabidopsis* nuclear genes.

2. Materials and Methods

2.1. Plant Growth Conditions and Sample Collection

Seeds of *Arabidopsis thaliana* col-0 were soaked in water and incubated in the dark at 4 °C for 2–3 days. Seeds were sown in paper pots containing a 1:2:1 mixture of horticultural perlite, peat moss, and vermiculite, and covered with a plastic wrap to maintain the moisture content. The pots were placed in a U-ING Green Farm hydroponic grow box (Japan Trend shop, Osaka, Japan) in a growth room at 22 °C temperature, 45% relative humidity, and a 16-h light/8-h dark cycle. After germination, water and fertilizers were supplied twice a week. Seedlings were harvested at different days and intervals.

2.2. RNA Extraction and cDNA Synthesis

Total RNA was extracted from seedlings using the Qiagen Plant Mini Kit (Hilden, Germany; catalog no. 74904), according to the manufacturer's instructions, and treated with DNase (RQ1 RNase free DNase; Promega, Madison, WI, USA) to remove traces

of contaminating genomic DNA. After DNase treatment, the samples were purified by phenol-chloroform extraction, followed by ethanol precipitation. The purified RNA was quantified using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, cDNA was synthesized using reverse transcriptase (Superscript III; Invitrogen, Carlsbad, CA, USA) and a random hexamer (oligo dT) primer. The sequences of forward and reverse primers are given in Table S3.

2.3. Library Preparation for Transcriptome Sequencing

The mRNA from 12-d- and 20-day-old samples were enriched using oligo (dT) beads. A total amount of 3 µg RNA per sample was used as input material for the RNA sample preparations. Then, total RNA was extracted and was sent to the company, Novogene Co., Ltd., for Next Generation Sequencing analysis. Sequencing libraries were generated using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (NEB, Ipswich, MA, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer 5X. First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H⁻). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, MA, USA). Then 3 µL USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The workflow for library preparation and transcriptome sequencing is shown in supporting Figure S1.

2.3.1. Clustering and Sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using HiSeq PE Cluster Kit cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and 125-bp/150-bp paired-end reads were generated.

2.3.2. Data Analysis

Quality control. Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly- N, and low-quality reads from raw data. At the same time, Q20, Q30, and GC content from the clean data were calculated, as given in supporting Table S1. All the downstream analyses were based on the clean data with high quality.

Reads mapping to the reference genome. Reference genome (TAIR 10) and gene model annotation files were downloaded from genome website directly. Index of the reference genome was built using Bowtie v2.2.3 and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12. We selected TopHat as the mapping tool, as it can generate a database of splice junctions based on the gene model annotation file and, thus, a better mapping result than other non-splice mapping tools.

Quantification of gene expression level. High-throughput sequencing (HTSeq v0.6.1, University of Heidelberg, Heidelberg, Germany) was used to count the reads' numbers mapped to each gene. Then the FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene. FPKM, expected number of fragments per

kilobase of transcript sequence per millions base pairs sequenced, considers the effect of sequencing depth and gene length for the reads' count at the same time and is currently the most commonly used method for estimating gene expression levels [25]. HTSeq software was used to analyze FPKM, indicating the gene expression levels in this experiment, using the union mode. The resulting files presented the number of genes with different expression levels and the expression level of single genes.

Differential expression analysis (For DESeq with biological replicates). Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq R package (1.18.0). DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting p -values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted p -value < 0.05 found by DESeq were assigned as differentially expressed. (For DESeq without biological replicates.) Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two conditions was performed using the DESeq R package (1.20.0). The p values were adjusted using the Benjamini and Hochberg method. Corrected p -value of 0.005 and $\log_2(\text{Fold change})$ of 1 were set as the threshold for significantly differential expression.

SNP analysis. Picard-tools v1.96 and samtools v0.1.18 were used to sort, mark duplicated reads, and reorder the bam alignment results of each sample. Genome Analysis Toolkit, GATK2 (v3.2) software was used to perform SNP calling. The mapping status of reads was provided in BAM files, which were visualized using the Integrative Genomics Viewer (IGV) software.

2.4. Sanger Sequencing

After doing PCR with equal amounts of cDNA (100 ng) in each reaction of 20 μL volume, the PCR products were purified by 1% agarose gels and the bands were cut out and frozen. DNA was purified using the QIAquick Gel Extraction kit, and concentration was measured by Nano-Drop. Sequencing of the purified DNA was performed using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Technologies, Waltham, MA, USA) using the forward and reverse primers (Table S3). The raw sequencing data were analyzed using the Sequence Scanner software version 2 (Applied Biosystems) and DNADynamo software.

3. Results

3.1. Identification and Analysis of DEGs by RNA-Seq

The level of gene expression was measured by determining transcript abundance; the greater the transcript abundance, the higher the gene expression level [26]. In RNA-seq analysis, gene expression level is estimated by counting the number of reads mapped onto genes or exons. The lists of descriptions for all expressed genes are given in Supporting data S1. Read count was proportional not only to the actual gene expression level but also to gene length and sequencing depth. Transcriptome data indicated that a total of 33,641 genes were expressed, of which 2140 were specifically expressed in 12-day-old seedlings' genes and 1485 in 20-day-old seedlings' (Figure 1A). The correlation coefficient is an important indicator of the reliability of the experiment: the closer the value of the correlation coefficient to 1, the greater the similarity between samples. The square of the Pearson's correlation coefficient (R^2) should be greater than 0.92 under ideal experimental conditions. In this study, R^2 was greater than 0.8, indicating a slight difference in gene expression between 12- and 20-d-old seedlings (Figure 1B). Volcano plots were used to infer the overall distribution of differentially expressed genes (DEGs). In experiments without biological replicates, the threshold is normally set as $|\log_2(\text{Fold Change})| > 1$ and $q\text{-value} < 0.005$. By contrast, in experiments with biological replicates, DESeq eliminates

biological variation; therefore, we set our threshold as adjusted p -value (padj) < 0.05. (Figure 1C).

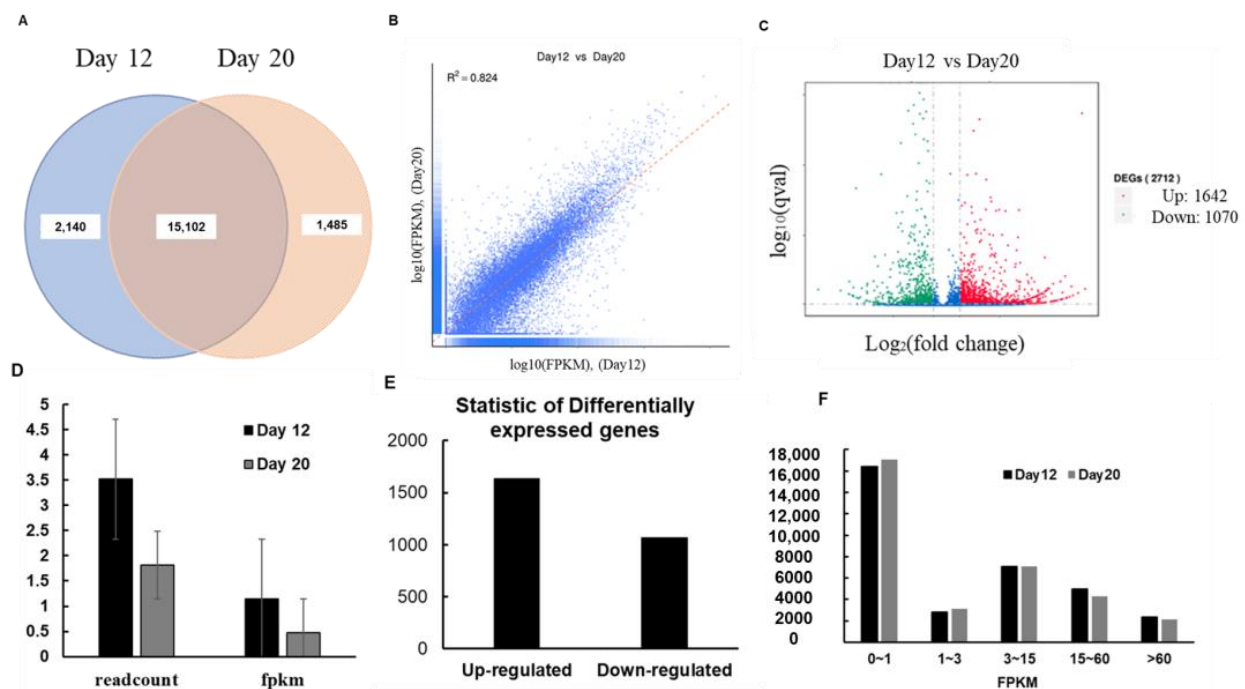


Figure 1. Analysis of genes differentially expressed between 12- and 20-d-old Arabidopsis seedlings. (A) Venn diagram of differentially expressed genes (DEGs). The sum of the numbers in each circle represents the total number of genes expressed within a sample, and the overlap represents genes expressed in both samples. (B) Correlation analysis of gene expression between samples. R^2 indicates the square of the Pearson's correlation coefficient. (C) Volcano plot of DEGs. The x -axis shows the fold change in gene expression between different samples and the y -axis shows the statistical significance of the differences in gene expression. Significantly up- and downregulated genes are highlighted in red and green, respectively. Genes showing no differential expression between 12- and 20-d-old seedlings are shown in blue. Comparison of the expression levels of DEGs (D–F). Comparison of, read count, and FPKM values of DEGs (D) between 12- and 20-d-old seedlings. (E) Summary of DEGs. (F) FPKM statistic.

The FPKM is the most well-known method of gene expression estimation in RNA-seq, as it takes into account the effects of both sequencing depth and gene length on read counts. Figure 1D shows that read counts and FPKM values were higher in 12-d-old seedlings than in the control sample (20-d-old seedlings), indicating higher expression of genes in 12-d-old seedlings. A total of 2712 genes were differentially expressed, of which 1642 were upregulated and 1070 were downregulated (Figure 1E), further indicating higher expression in 12-d-old seedlings. All DEGs are listed in Supporting data S2.

To compare gene expression levels under different conditions, an FPKM distribution diagram was used. The final FPKM value represents the mean of biological replicates. In general, an FPKM value of 0.1 or 1 was used as a threshold to determine whether a gene is expressed or not. The number of genes with different expression levels is shown in Figure 1F.

3.2. Comparison of Nucleotide Differences between Genomic DNA in Database and RNA-Seq of 12- or 20-D-Old Seedlings

Comparison of RNA-seq data of 12-day- or 20-day-old Arabidopsis Col-0 plants to genomic DNA sequence in the database identified 12 types of possible nucleotide conversion patterns in transcripts: G-to-A, C-to-U, U-to-C, U-to-A, A-to-G, C-to-A, A-to-T, G-to-T, C-to-G, A-to-C, G-to-C, and U-to-G. Among these patterns, U-to-C was the third most predominant after G-to-A and C-to-U. Single-nucleotide base differences are listed in Supporting data S3. RNA-seq analysis revealed 590 different sites, of which 79 sites (13%)

represented possible U-to-C conversion. Out of 253 genes showing nucleotide differences, 50 contained possible U-to-C conversion (Figure 2A,B). A list of candidate U-to-C RNA editing sites detected in Arabidopsis seedlings is given in Table 1.

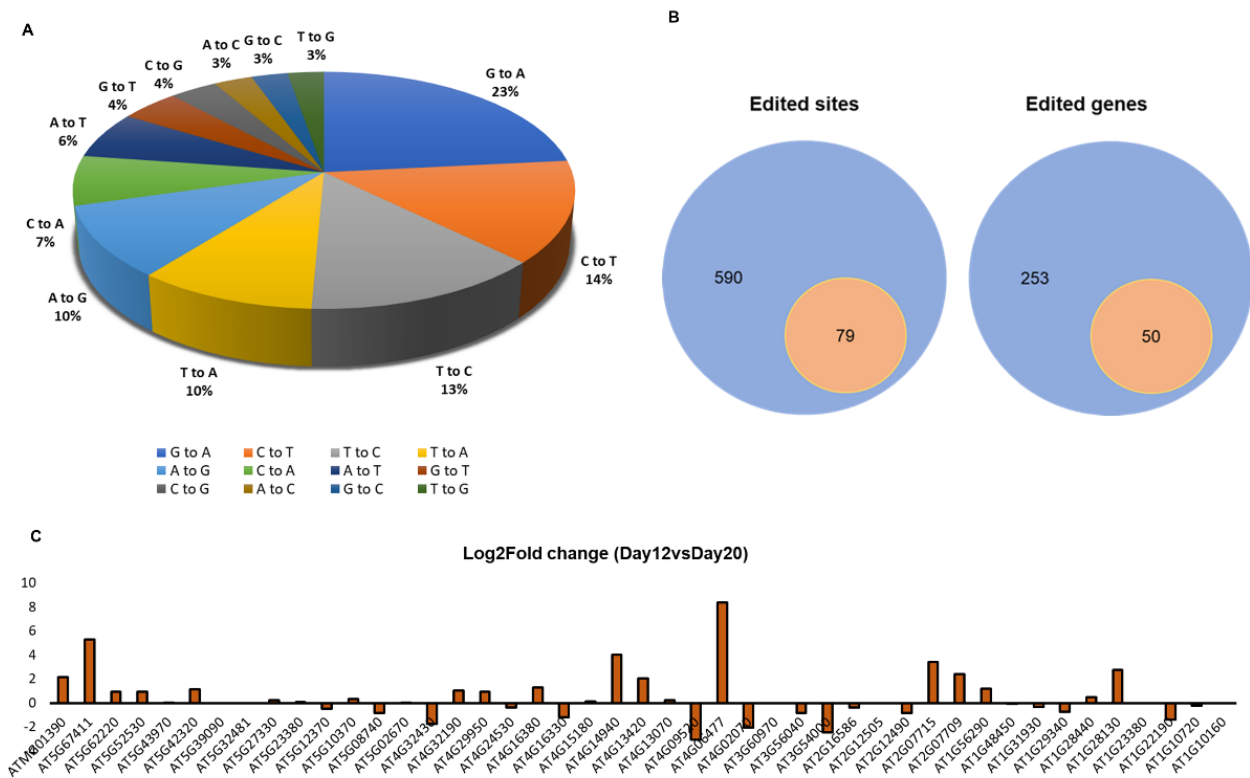


Figure 2. Analysis of single-nucleotide base conversions identified in 12-d-old Arabidopsis seedlings by RNA-seq. (A) Pie chart showing the percentage for genes identified with single-nucleotide base conversions. (B) Number of total edited sites and edited genes (blue), and number of sites and genes with U-to-C mutations (orange). (C) Log₂FC values for the genes identified with U-to-C nucleotide conversion. Genes were expressed to higher levels in 12-d-old seedlings than in 20-d-old seedlings (C).

Next, we analyzed the next-generation sequencing (NGS) data of Arabidopsis for expressed PPR genes using the Bioinformatics & Evolutionary Genomics website. The expressed PPR genes are listed in descending order of expression in Supporting data S4. The right column contains the genes containing all nucleotide base conversions. Out of 465 expressed PPR genes, 10 genes including AT3G62470, AT1G50270, AT1G16830, AT1G63080, AT1G06580, AT3G56550, AT1G09820, AT3G53360, AT2G22410, and AT4G32430 showed nucleotide conversion (Figure 3A). Out of 54 U-to-C variant genes, one gene, AT4G32430, was found as PPR gene (Figure 3B). The list of expressed genes, PPR genes that differed in base nucleotide conversions, the genes that differed in U-to-C base conversion, and the PPR gene that differed in U-to-C base conversion are shown in Figure 3C.

Table 1. List of candidate U-to-C RNA editing sites detected in Arabidopsis seedlings at different developmental stages.

S.No.	Position	Reads		Gene ID	Description
		12 Days	20 Days		
1	3412532	56	0	AT2G07715	Ribosomal Proteins L2, RNA binding domain
2	8544440	34	0	AT4G14940	Amine oxidase
3	26898977	2	0	AT5G67411	GRAS family transcription factor
4	8297931	4	0	AT1G23380	KNOTTED1-like homeobox gene 6
5	14657330	14	0	AT4G29950	Ypt/Rab-GAP domain of gyp1p superfamily protein
6	3392826	107	16	AT2G07709	-
7	362386	175	44	ATMG01390	-
8	7191444	105	197	AT2G16586	Unknown
9	3061212	498	2	AT4G06477	-

Table 1. Cont.

S.No.	Position	Reads		Gene ID	Description
		12 Days	20 Days		
10	9226791	28	69	AT4G16330	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
11	5816271	12	6	AT3G17050	-
12	9255546	268	99	AT4G16380	Heavy metal transport/detoxification superfamily protein
13	14198871	647	240	AT3G41768	-
14	16918673	55	46	AT5G42320	Zn-dependent exopeptidases superfamily protein
15	17708862	0	21	AT3G47965	Unknown
16	24989428	27	0	AT5G62220	glycosyltransferase 18
17	21320395	0	12	AT5G52530	dentin sialophosphoprotein-related
18	2848835	146	86	AT5G08740	NAD(P)H dehydrogenase C1
19	15546833	13	0	AT4G32190	Myosin heavy chain-related protein
20	3392918	144	14	AT2G07709	-
21	21319578	0	5	AT5G52530	dentin sialophosphoprotein-related
22	21077241	0	2	AT1G56290	Cwfl-like family protein
23	7622202	0	2	AT4G13070	RNA-binding CRS1/YhbY (CRM) domain protein
24	17692876	29	0	AT5G43970	translocase of outer membrane 22-V
25	10266697	46	6	AT1G29340	plant U-box 17
26	7869982	17	0	AT5G23380	Protein of unknown function (DUF789)
27	7836325	19	0	AT1G22190	Integrase-type DNA-binding superfamily protein
28	19998466	36	0	AT3G54000	Unknown
29	603074	13	5	AT5G02670	Unknown
30	22561577	0	2	AT3G60970	multidrug resistance-associated protein 15
31	6025041	0	27	AT4G09520	Cofactor-independent phosphoglycerate mutase
32	909133	0	2	AT4G02070	MUTS homolog 6
33	7797368	0	4	AT4G13420	high affinity K ⁺ transporter 5
34	8662474	0	3	AT4G15180	SET domain protein 2
35	12669828	0	2	AT4G24530	O-fucosyltransferase family protein
36	15653919	0	2	AT4G32430	Pentatricopeptide repeat (PPR) superfamily protein
37	5075516	0	2	AT2G12490	-
38	17587422	0	2	AT2G42200	squamosa promoter binding protein-like 9
39	17958701	0	2	AT2G43200	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
40	526197	0	5	AT3G02515	-
41	20795012	69	64	AT3G56040	UDP-glucose pyrophosphorylase 3
42	3264804	0	2	AT5G10370	helicase domain-containing protein/IBR domain-containing protein/zinc finger protein-related
43	9633752	0	2	AT5G27330	Prefoldin chaperone subunit family protein
44	12108844	0	9	AT5G32481	-
45	15644809	0	4	AT5G39090	HXXXD-type acyl-transferase family protein
46	3332097	0	2	AT1G10160	-
47	3564739	0	2	AT1G10720	BSD domain-containing protein
48	9825469	0	6	AT1G28130	Auxin-responsive GH3 family protein
49	9997031	0	2	AT1G28440	HAESA-like 1
50	4006628	0	13	AT5G12370	exocyst complex component sec10
51	5097198	0	5	AT2G12505	-
52	11465954	0	11	AT1G31930	extra-large GTP-binding protein 3
53	7014676	0	2	AT3G20087	N/A
54	15766171	0	2	AT2G37585	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
55	7191297	249	171	AT2G16586	Unknown
56	17908527	0	2	AT1G48450	Protein of unknown function (DUF760)

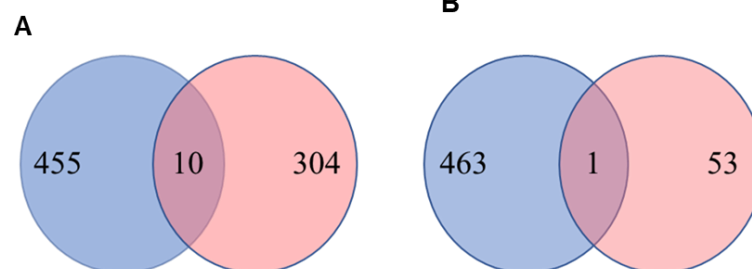


Figure 3. Cont.

C

Names	total	elements
PPR genes that differ in base nucleotide conversion	10	AT3G62470 AT1G50270 AT1G16830 AT1G63080 AT1G06580 AT3G56550 AT1G09820 AT3G53360 AT2G22410 AT4G32430
Expressed PPR genes	465	AT1G19290 AT2G37320 AT5G59200 AT1G15510 AT3G18970 AT3G05240 AT1G69290 AT3G29230 AT1G80270 AT4G17616 AT3G60960 AT2G41080 AT1G79540 AT1G77340 AT1G03540 AT2G33760 AT1G62350 AT1G11290 AT3G49240 AT3G49710 AT1G74750 AT1G62910 AT4G26680 AT1G62260 AT5G50990 AT5G39680 AT5G44230 AT2G16880 AT4G33170 AT2G20720 AT2G36240 AT1G76280 AT5G01110 AT2G13600 AT1G02060 AT5G40405 AT3G46790 AT5G14080 AT1G53330 AT5G48910 AT5G02830 AT5G61400 AT3G15200 AT4G20770 AT5G15340 AT3G06430 AT1G11900 AT4G08210 AT5G66500 AT5G39710 AT5G08305 AT3G47840 AT1G73710 AT3G46870 AT1G80880 AT5G48730 AT3G59040 AT4G02820 AT1G51965 AT4G37170 AT4G21170 AT3G02650 AT3G04130 AT5G67570 AT5G57250 AT3G58590 AT1G77170 AT4G14050 AT3G29290 AT5G25630 AT3G26630 AT4G35130 AT4G14190 AT1G60770 AT1G28020 AT5G28370 AT2G17525 AT5G24830 AT1G11710 AT1G43010 AT2G01740 AT1G13410 AT3G49730 AT4G18840 AT1G09900 AT3G56030 AT2G17033 AT4G01570 AT2G32230 AT2G21090 AT2G37230 AT1G22830 AT1G32415 AT1G17630 AT2G34370 AT2G44880 AT2G27800 AT1G55890 AT2G29760 AT4G14820 AT1G77150 AT5G55740 AT1G80550 AT1G64430 AT5G11310 AT2G03380 AT4G26800 AT2G37310 AT5G27110 AT3G46610 AT3G61170 AT1G63330 AT2G18940 AT1G73400 AT1G52620 AT4G30700 AT2G19280 AT1G26460 AT5G62370 AT5G08490 AT3G13880 AT2G03880 AT1G71420 AT5G14820 AT1G08610 AT5G61990 AT5G65570 AT5G02860 AT3G22470 AT3G17370 AT4G31070 AT3G13770 AT2G06000 AT5G40400 AT3G02010 AT1G28690 AT1G74580 AT2G20540 AT3G09060 AT2G17140 AT5G41170 AT5G19020 AT2G39230 AT2G34400 AT2G48000 AT1G05600 AT5G06540 AT4G18520 AT3G22150 AT4G02750 AT5G50280 AT3G01580 AT3G49740 AT5G14770 AT1G31790 AT5G36300 AT4G30825 AT1G12250 AT3G51320 AT1G20300 AT4G32450 AT1G16480 AT5G60960 AT2G26790 AT1G30610 AT3G25060 AT4G39530 AT2G32630 AT2G25580 AT3G50420 AT2G04860 AT1G62680 AT2G17670 AT2G36980 AT5G28380 AT5G18475 AT1G68930 AT1G34160 AT2G02150 AT2G20710 AT1G12775 AT1G22960 AT1G09220 AT2G36730 AT5G61370 AT3G62890 AT4G14850 AT5G52850 AT1G31920 AT1G18485 AT1G62720 AT3G04750 AT2G41720 AT4G21190 AT5G52630 AT1G62930 AT5G18950 AT2G35130 AT4G01400 AT1G31430 AT5G14350 AT1G61870 AT1G02370 AT2G15630 AT3G03580 AT5G13230 AT1G05670 AT4G16470 AT3G49142 AT1G55630 AT3G15130 AT5G37570 AT2G39620 AT1G80150 AT2G02750 AT4G21065 AT4G21705 AT3G14580 AT1G15480 AT1G19520 AT5G43820 AT3G47530 AT3G07290 AT5G66631 AT1G06710 AT3G15590 AT4G16835 AT1G31840 AT3G25970 AT5G15010 AT5G28340 AT3G22670 AT3G11380 AT2G44920 AT1G09190 AT4G04790 AT4G20740 AT3G23020 AT5G12100 AT3G02330 AT5G47360 AT1G13630 AT1G33350 AT3G61520 AT5G10690 AT3G16710 AT4G19220 AT1G18900 AT4G20090 AT5G39350 AT5G43790 AT5G28460 AT5G59900 AT3G16010 AT1G63150 AT1G74400 AT1G64310 AT5G56310 AT4G22760 AT3G04760 AT5G42310 AT4G18975 AT1G71210 AT5G64320 AT3G53170 AT4G01030 AT3G24000 AT4G31850 AT1G63130 AT5G53490 AT1G06140 AT5G46460 AT2G33680 AT5G38730 AT3G06920 AT2G38420 AT5G15280 AT1G64583 AT3G25210 AT2G01860 AT5G59600 AT1G62670 AT3G09040 AT1G26500 AT1G71060 AT4G34830 AT1G66345 AT1G13800 AT1G12700 AT2G42920 AT5G55840 AT1G74800 AT4G13650 AT1G10330 AT5G46680 AT4G04370 AT4G33990 AT1G62914 AT3G48810 AT1G02150 AT1G79490 AT1G08070 AT2G30100 AT3G08820 AT2G46050 AT5G04780 AT3G42630 AT5G18390 AT3G13150 AT1G26900 AT1G71460 AT1G52640 AT1G43980 AT1G63400 AT1G63320 AT1G20230 AT5G06400 AT2G40720 AT3G18020 AT1G03510 AT2G01360 AT3G61360 AT4G01990 AT3G60980 AT5G13770 AT3G29840 AT3G18990 AT5G42450 AT1G10270 AT4G39620 AT4G18750 AT4G38150 AT1G74900 AT3G49170 AT2G18520 AT2G17210 AT1G13040 AT4G19440 AT1G11630 AT1G19720 AT5G09450 AT3G29660 AT5G15300 AT3G15930 AT3G14330 AT3G23330 AT3G13160 AT1G12300 AT1G77360 AT1G07740 AT2G35030 AT5G39980 AT3G21470 AT2G01510 AT4G25270 AT1G69350 AT4G38010 AT4G16390 AT5G27460 AT1G03100 AT1G30290 AT3G57430 AT2G45350 AT1G663070 AT2G31400 AT3G26782 AT1G74850 AT3G28540 AT4G21300 AT1G03560 AT1G09410 AT2G28050 AT1G10910 AT4G36680 AT3G20730 AT2G15980 AT5G68520 AT4G11690 AT3G60050 AT1G64580 AT5G46590 AT5G03800 AT1G68980 AT5G65560 AT1G59720 AT2G30780 AT5G16860 AT1G58570 AT1G01970 AT5G40410 AT5G09950 AT5G27270 AT3G02490 AT1G53600 AT2G15690 AT1G09680 AT1G14470 AT5G65820 AT5G46100 AT1G62590 AT3G18840 AT1G74630 AT4G15720 AT3G18110 AT5G15980 AT5G47460
PPR genes that differ in U-to-C base nucleotide conversion	1	AT4G32430
Genes that differ in U-to-C base nucleotide conversion	53	AT4G29950 AT4G13070 AT1G31930 AT5G32481 AT1G10720 AT1G48450 AT5G42320 AT1G28130 AT5G08740 AT5G23380 AT3G56040 AT4G16380 AT1G29340 AT5G12370 AT3G17050 AT4G09520 AT2G18586 AT3G60970 AT4G24530 AT3G41768 AT2G12500 AT5G67411 AT3G20087 AT4G13420 AT2G12490 AT4G32190 AT2G07709 AT2G43200 AT1G56290 AT5G62220 AT1G10160 AT5G52530 AT4G06477 AT1G23380 AT2G07715 AT3G47965 AT4G16330 AT5G27330 AT2G37585 AT5G39090 AT4G14940 AT3G02515 AT3G54000 ATMG01390 AT1G28440 AT4G15180 AT1G22190 AT5G02670 AT2G42200 AT5G10370 AT5G43970 AT4G02070

Figure 3. The next-generation sequencing (NGS) data of Arabidopsis for expressed PPR genes. Out of 465 expressed PPR genes, 10 genes including AT3G62470, AT1G50270, AT1G16830, AT1G63080, AT1G06580, AT3G56550, AT1G09820, AT3G53360, AT2G22410, and AT4G32430 showed nucleotide conversion (A). Out of 54 U-to-C variant genes, one gene, AT4G32430, was found as PPR gene (B). The list of expressed genes, PPR genes that differed in base nucleotide conversions, the genes that differed in U-to-C base conversion, and the PPR gene that differed in U-to-C base conversion are shown in (C).

3.3. Identification of Genes Harboring U-To-C RNA Editing Site

We selected the genes of both samples that had a minimum number of reads to be able to infer an editing event. This minimum number should be reasonably high to minimize the impact of sequencing artifacts. For example, the T-to-C change at position 14,198,871 in AT3G41768 was supported by 647(29%) and 240 (19%) in 12-d-old and 20-d-old seedlings, respectively. In addition, there were some variants that were supported by 100% of the reads in both samples (12- and 20-d-old). Therefore, these are several editing events that

seem to be polymorphisms. For the same gene, we found many reads in the same U-to-C conversions. Genes with higher read coverage were further examined for the confirmation of U-to-C RNA editing sites. Genes, such as AT2G16586, AT5G02670, AT5G42320, AT4G16380, and AT5G08740, showed 249, 13, 55, 268, and 146 reads at the converted sites, respectively. Genes showing extremely low reads (0, 2) were also analyzed by RT-PCR. However, very few sites were confirmed as editing events. Because many reads mapped to each U-to-C conversion site, we considered that these nucleotide conversions were caused by RNA editing [27]. The flowchart for methodology for identification of U-to-C RNA editing sites is shown in Figure 4A. Furthermore, we validated the RNA-seq-based candidates experimentally by Sanger sequencing of both genomic, gDNA, and cDNA for all candidate genes. We extracted DNA and mRNA from the same aliquots of seedling samples. By sequencing the paired DNA and cDNA samples and analyzing each chromatogram by two individuals independently we confirmed the U-to-C RNA edited sites. The cDNA showed a double peak, representing T and edited C nucleotides, while no double peak was observed in gDNA sequencing. The sequencing was performed using sense primer targeting at the editing sites. Validation using PCR and Sanger sequencing verified seven genes, including AT2G16586, AT5G42320, AT5G02670, AT3G41768, AT4G32430, AT3G47965, and AT5G52530, containing U-to-C RNA editing sites. The Sanger sequence chromatograms for all seven edited genes are showed in Figure 5. The raw sequencing data were analyzed using the software, DNADyanamo and Sequence Scanner version 2 (Applied Biosystems). When the edited and unedited products were presented together, a dual peak (T (unedited) and C (edited)) was observed at the target site. The editing efficiency was calculated from peak area and a list of genes showing U-to-C RNA editing in 12-d- and 20-day old Arabidopsis seedlings is given in Table 2. Furthermore, we also investigated the editing efficiency at different developmental stages of Arabidopsis, such as four days and eight days. It was found that no editing occurred at early stages of development, like in four days, while a few editing could be identified in 8-day-old seedlings (Table S3). The U-to-C RNA editing sites were majorly located within the UTRs of mature mRNAs.

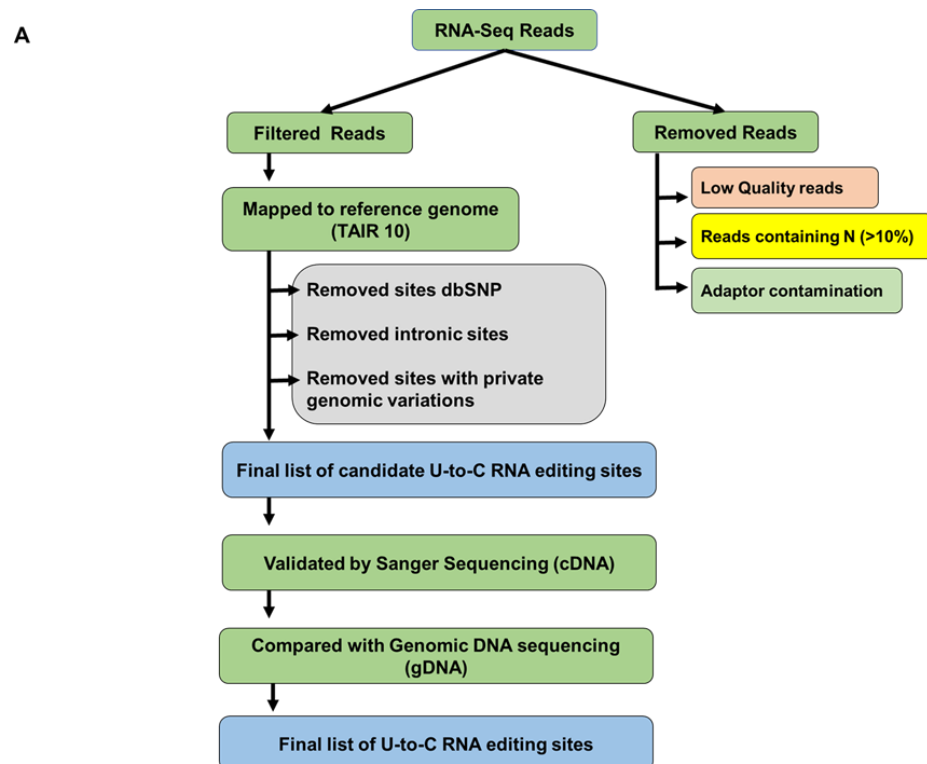


Figure 4. Cont.

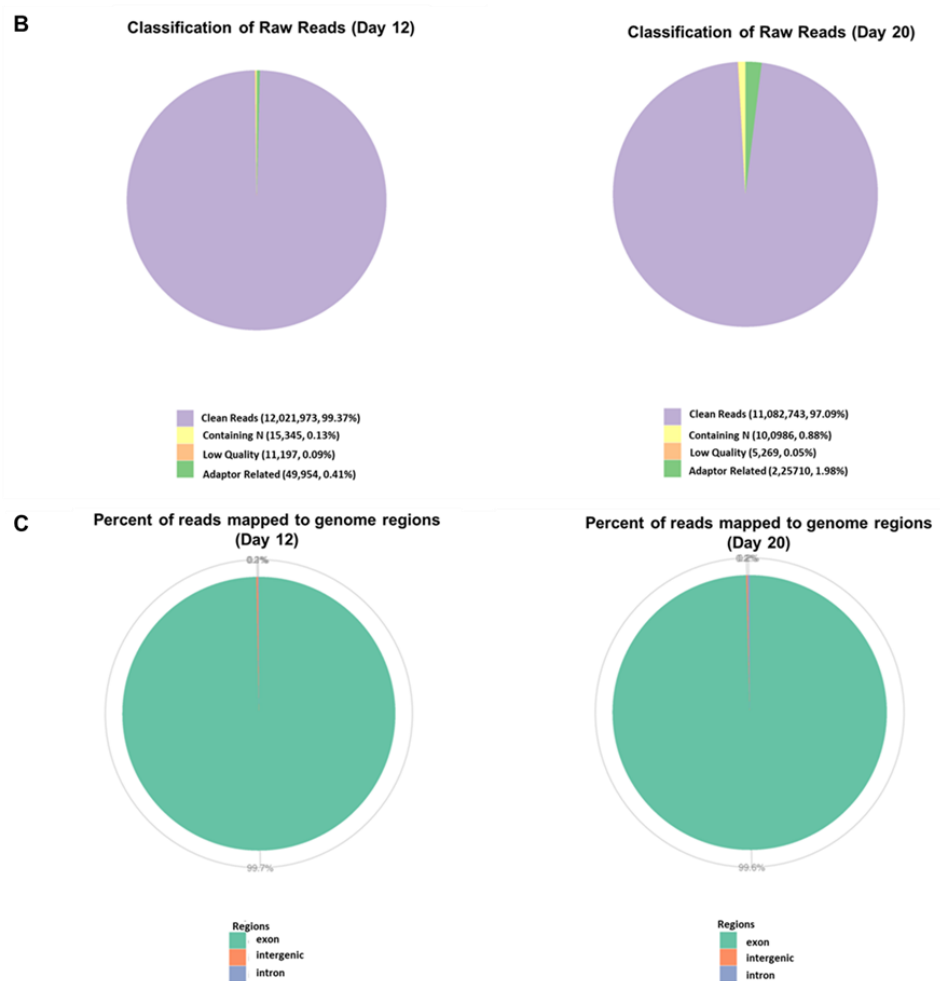


Figure 4. The flowchart for methodology for identification of U-to-C RNA editing site. (A) Raw reads are filtered to remove reads containing adapters or reads of low quality, so that downstream analyses are based on clean reads. The filtering process is as follows. (1) Discard reads with adaptor contamination. (2) Discard reads when uncertain nucleotides constitute more than 10% of either read ($N > 10\%$). (3) Discard reads when low-quality nucleotides (base quality less than 20) constitute more than 50% of the read. For mapping sequences, TopHat2 was chosen for plant genomes. The mismatch parameter was set to 2 and other parameters were set to default. Appropriate parameters were also set, such as the longest intron length. Only filtered reads were used to analyze the mapping status of RNA-seq data to the reference genome. Edited sites were further validated and confirmed by RT-PCR. (B) Clean reads for day 12 and day 20. (C) Percentage of reads mapped to genome regions for day 12 and day 20.

Table 2. List of genes identified with U-to-C RNA editing sites in 12-day- and 20-day-old Arabidopsis seedlings.

S.No.	Position	Edited Site	Gene ID	RNA Editing Efficiency (in %)		Encoded Protein
				12 Days	20 Days	
1.	14198871	5' UTR	AT2G16586	77.30	65.74	Transmembrane protein
2.	16918673	CDS	AT5G42320	24.20	0	Zn-dependent exopeptidase superfamily protein
3.	603074	5' UTR	AT5G02670	0	22.80	Hypothetical protein
4.	7191297	3' UTR	AT3G41768	45.54	49.65	Ribosomal RNA
5.	15653919	3' UTR	AT4G32430	0	20.43	PPR-like superfamily protein
6.	17708862	3' UTR	AT3G47965	24.54	22.48	Hypothetical protein
7.	21320395	CDS	AT5G52530	20.65	0	Dentin sialophosphoprotein-like protein

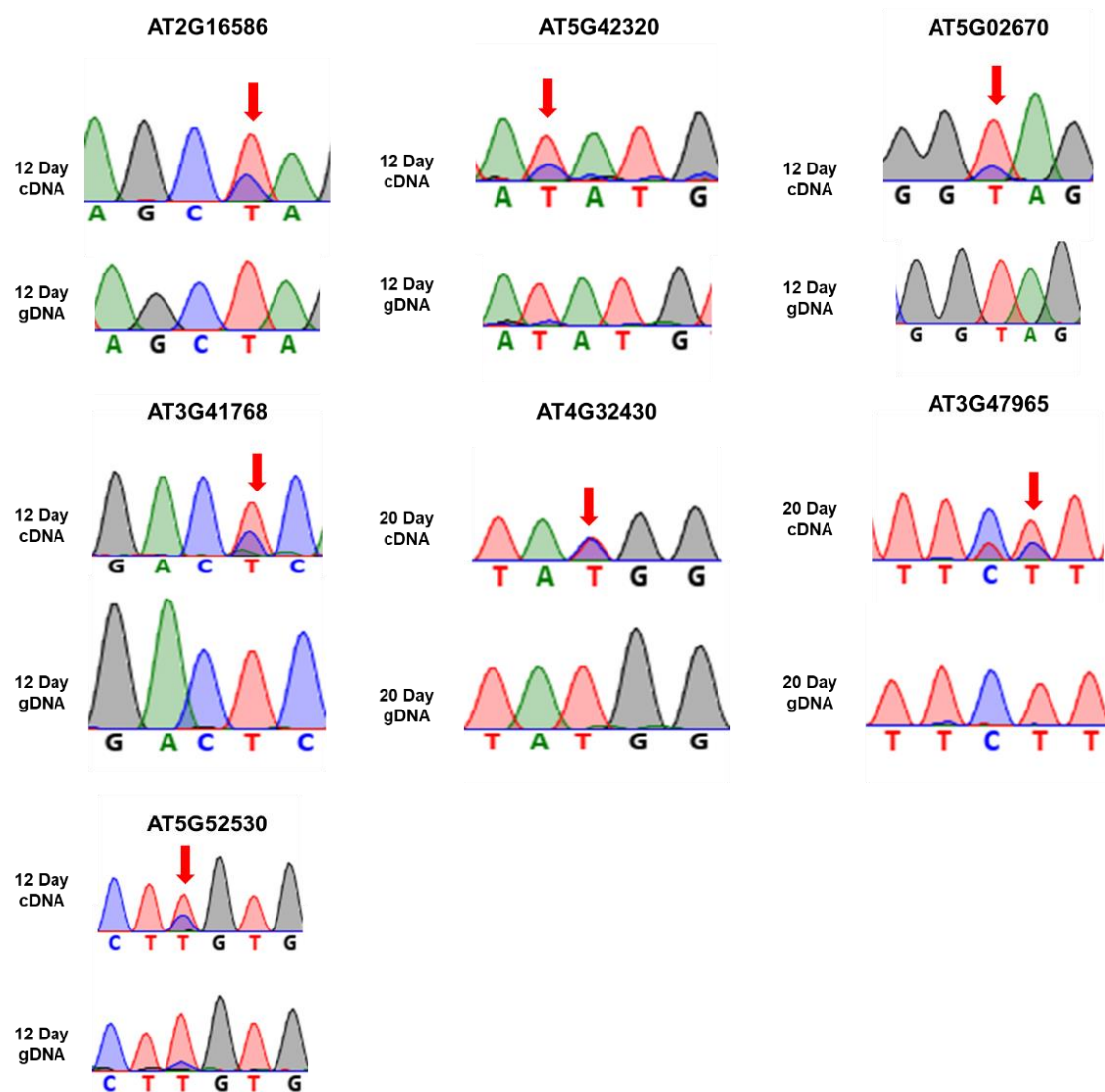


Figure 5. The Sanger sequence chromatogram depicting the U-to-C types of RNA editing events in 12-d- and 20-day-old seedlings from the same tissues of Arabidopsis via cDNA and genomic, gDNA using forward primers. Arrows indicate the position of RNA editing.

4. Discussion

In our knowledge, this is the first report of U-to-C RNA editing for nuclear genes confirmed by both RNA-seq and Sanger sequencing in flowering plants. In this study, total RNA extracted from 12-d- and 20-d-old seedlings was examined by high-throughput RNA-seq.

Total RNA isolated from 12-d-old seedlings was examined by NGS, and DEGs were identified based on FPKM values and read counts. The results showed that DEGs were expressed to higher levels in 12-d-old seedlings than in 20-d-old seedlings. This was confirmed by higher FPKM values and read counts and more upregulated genes in 12-d-old seedlings than in 20-d-old seedlings. The ANOVA test was performed for comparing the gene expression levels. The summary for regression analysis of differentially expressed genes among the replicates of 12-d- and 20-d-old seedlings is given in Table S5. Additionally, PPR genes were also expressed to higher levels in 12-d-old than in 20-d-old seedlings, as indicated by the \log_2FC values. These data suggest that DEGs are more likely to be expressed in young Arabidopsis seedlings than in older seedlings. Therefore, more mutations could occur at this stage of development because RNA editing events are more frequent in seedlings than in any other plant tissue.

While investigating for RNA editing events to create a global map of high-quality candidates, an appropriate balance between sensitivity (identifying a highly inclusive set of possible edits) and specificity (being more confident that a call is, in fact, a true RNA edit) is required. We considered it better to have a fewer number of candidate RNA editing events that are more likely to be true than to have a larger number with an increased percentage of false positives. We undoubtedly did not score a substantial number of true, low-level, U-to-C RNA editing events in the process. Up to 90% of nucleotide variants that are not SNPs (either in dbSNP or private genomic SNVs) are U-to-C calls; this suggests they are likely to be U-to-C editing candidates. Furthermore, more than 85% of these candidates are located in UTRs. Our candidate U-to-C RNA editing sites had a different variant frequency from known SNPs. They tended to cluster predominantly in the untranslated regions.

We investigated single-nucleotide base changes and the percentage of read coverage was calculated (Table 1). We predicted 12 types of nucleotide differences, including possible U-to-C conversions. RT-PCR products of the genes including the candidate U-to-C conversions were subjected to Sanger sequencing. A total of seven genes, AT2G16586, AT5G42320, AT1G05670, AT3G41768, AT4G32430, AT3G47965, and AT5G52530, were identified as targets for U-to-C RNA editing (Table 2). The UTRs of genes encoding proteins involved in RNA metabolism and RNA binding, including PPR proteins, Zn-finger (ZnF)-related proteins, ribosomal protein L2, transmembrane proteins, and two hypothetical proteins, were identified as target of U-to-C editing. Interestingly, the ribosomal RNA, AT3G41768, was identified for 45.65% of U-to-C RNA editing efficiency. Since about 50% of genes are affected with editing, it might have had significant effect on their functions. Similarly, the transmembrane protein, AT2G16586, was identified with 77.3% of U-to-C RNA editing efficiency, which may affect its general physiology. In addition, the PPR gene, AT4G32430, was also identified with 20.43% U-to-C RNA editing.

While RNA editing in introns or UTR regions can affect mRNA stability, translation, or splicing activity because of the modification of its secondary structure, those in coding region can also affect the translated polypeptide sequence [28–30]. In this study, we demonstrated that most U-to-C RNA editing events are located in UTRs, which may affect the secondary structure and, consequently, the stability of mRNA.

In Arabidopsis, C-to-U and U-to-C RNA editing have been reported at the translation borders of nuclear transcripts, AT1G29930.1 and AT1G52400.1 [31]. These deamination (C-to-U) and amination (U-to-C) events accumulated at adjacent sites; therefore, the possibility that the deamination reaction serves as the amino group donor for the amination reaction was proposed, although the frequency of amination was higher than that of deamination [31]. Although this hypothesis is attractive, we could not detect the same RNA editing events in our RNA-seq data. Thus, the amino group donor of the U-to-C amination in plants is unclear. However, in cDNA AT3G47965 there is also a small T superposing with the C just upstream the edited T, showing the possible immediate donor of amino group. Previously, an extensive research on editing sites in nuclear transcripts for mRNA by Parallel Analysis of RNA Ends (PARE) and Massively Parallel Signature Sequencing (MPSS) data was reported. It showed that all 12 RNA editing patterns may exist in the nuclear genes and that perhaps the numbers of editing sites in a specific pattern may vary. The study suggested that RNA editing is an essential RNA-based regulatory layer not only for mitochondrial and chloroplast genes but also for nuclear genes. However, a global vision of RNA editing in plant nuclear protein-coding transcripts has not been realized. Therefore, this work intended to uncover the occurrence of RNA editing events in the nuclear genes of Arabidopsis.

We further compared the gene expression levels for seven identified U-to-C RNA editing target genes among different tissues (Figure S2). The green bar shows the genes expressed in seedling stage of development of Arabidopsis. The day-specific characteristic of the U-to-C RNA editing events implied that these were post-transcriptional modifications, not genomic mutations. These editings were identified as a growth-dependent RNA editing efficiency alteration. Day 4 seedlings did not have RNA editing, at least

(Table S4). It indicates that the enzyme important for this editing events might have been expressed at defined stages of seedling development. Next, to validate whether the identified RNA editing sites were true positive, we searched for evidence of the identified RNA editing sites in *Arabidopsis* RNA-seq data generated by public laboratories, using online software <http://signal.salk.edu/atg1001/3.0/gebrowser.php>, accessed on 5 January 2021. All seven identified U-to-C RNA editing sites AT2G16586, AT5G42320, AT5G02670, AT3G41768, AT4G32430, AT3G47965, and AT5G52530 were aligned against the publicly available RNA-seq databases (Supporting data S6) and confirmed our findings. The target T sites were identified as edited C sites in various databases. The comparative analysis of *Arabidopsis* RNA-seq is shown in Figure S3. The edited sites are indicated within red boxes. Further studies are needed to better understand the processes involved in U-to-C RNA editing, including the identification of *cis* or *trans* regulatory elements, isolation of editing enzymes, and validation of editing sites.

5. Conclusions

Our findings confirm the uridine-to-cytidine RNA editing sites in some nuclear genes in *Arabidopsis thaliana*. A comprehensive analysis of RNA-seq data to detect nucleotide base conversions was performed. In this study, we examined U-to-C RNA editing in *Arabidopsis* seedlings at different developmental stages. Sanger sequencing identified the sites and efficiency of seven U-to-C editing events. Most U-to-C RNA editing here identified occurred in the UTR of mature mRNAs. Thus, we confirmed the presence of U-to-C RNA editing in nuclear genes of plants. We provided the experimental basis to explore the mechanism involved in the amination of U-to-C editing and functions and effects of U-to C RNA editing on mRNA stability, other RNA modifications, and translation.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4409/10/3/635/s1>. Figure S1: The workflow for library preparation and transcriptome sequencing. Figure S2: Comparative analysis of gene expression levels for seven identified U-to-C RNA editing target genes among different tissues. Green bar shows the genes expressed in seedling stage of development of *Arabidopsis*. Figure S3: Validation of target U-to-C RNA editing sites on *Arabidopsis* RNA-seq database. Table S1: Data table for quality control. Table S2: List of candidate U-to-C RNA editing sites detected in *Arabidopsis* seedlings showing the percentage of read coverage. Table S3: List of candidate genes for U-to-C RNA editing sites in *Arabidopsis* seedlings at different developmental stages showing the forward and reverse primer sequences. Table S4: List of genes identified with U-to-C RNA editing in *Arabidopsis* seedlings at different developmental stages. Table S5: Summary for regression analysis of differentially expressed genes among the replicates of 12-day- and 20-day-old seedlings. Supporting data S1: Descriptions for all genes. Supporting data S2: Lists of DEGs. Supporting data S3: Lists of single-nucleotide conversions. Supporting data S4: Expressed PPR gene lists. Supporting data S5: Description for U-to-C conversion. Supporting data S6: RNA-seq database table.

Author Contributions: Conceptualization, R. and T.T.; methodology, R. and T.T.; software, R. and M.S.; validation, R., C.O., and M.S.; formal analysis, R. and T.T.; investigation, R. and C.O.; resources, T.T.; data curation, R.; writing—original draft preparation, R.; writing—review and editing, R., M.S., and T.T.; visualization, R.; supervision, T.T.; project administration, T.T.; funding acquisition, T.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (17H02204 and 18K19288).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are highly grateful to Mizuki Takenaka for his valuable discussion.

Conflicts of Interest: Authors declare no conflict of interests.

References

1. Benne, R. RNA-editing in trypanosome mitochondria. *Biochim. Biophys. Acta Gene Struct. Expr.* **1989**, *1007*, 131–139. [[CrossRef](#)]
2. Chen, S.; Habib, G.; Yang, C.; Gu, Z.; Lee, B.; Weng, S.; Silberman, S.R.; Cai, S.; Deslypere, J.; Rosseneu, M. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **1987**, *238*, 363–366. [[CrossRef](#)]
3. Walkley, C.R.; Li, J.B. Rewriting the transcriptome: Adenosine-to-inosine RNA editing by ADARs. *Genome Biol.* **2017**, *18*, 205. [[CrossRef](#)] [[PubMed](#)]
4. Liu, H.; Wang, Q.; He, Y.; Chen, L.; Hao, C.; Jiang, C.; Li, Y.; Dai, Y.; Kang, Z.; Xu, J.R. Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Res.* **2016**, *26*, 499–509. [[CrossRef](#)]
5. Wolf, J.; Gerber, A.P.; Keller, W. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.* **2002**, *21*, 3841–3851. [[CrossRef](#)]
6. Bar-Yaacov, D.; Pilpel, Y.; Dahan, O. RNA editing in bacteria: Occurrence, regulation and significance. *RNA Biol.* **2018**, *15*, 863–867. [[CrossRef](#)]
7. Thomas, S.M.; Lamb, R.A.; Paterson, R.G. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **1988**, *54*, 891–902. [[CrossRef](#)]
8. Cattaneo, R.; Kaelin, K.; Baczkó, K.; Billeter, M.A. Measles virus editing provides an additional cysteine-rich protein. *Cell* **1989**, *56*, 759–764. [[CrossRef](#)]
9. Covello, P.S.; Gray, M.W. RNA editing in plant mitochondria. *Nature* **1989**, *341*, 662–666. [[CrossRef](#)]
10. Gualberto, J.M.; Lamattina, L.; Bonnard, G.; Weil, J.H.; Grienenberger, J.M. RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* **1989**, *341*, 660–662. [[CrossRef](#)]
11. Hiesel, R.; Wissinger, B.; Schuster, W.; Brennicke, A. RNA editing in plant mitochondria. *Science* **1989**, *246*, 1632–1634. [[CrossRef](#)] [[PubMed](#)]
12. Barkan, A.; Small, I. Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 415–442. [[CrossRef](#)]
13. Lurin, C.; Andrés, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyère, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; et al. Small Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **2004**, *16*, 2089–2103. [[CrossRef](#)]
14. Chateigner-Boutin, A.L.; Hanson, M.R. Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res.* **2003**, *31*, 2586–2594. [[CrossRef](#)] [[PubMed](#)]
15. Schallenberg-Rüdinger, M.; Kindgren, P.; Zehrmann, A.; Small, I.; Knoop, V. A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J.* **2013**, *76*, 420–432. [[CrossRef](#)] [[PubMed](#)]
16. Leu, K.C.; Hsieh, M.H.; Wang, H.J.; Hsieh, H.L.; Jauh, G.Y. Distinct role of Arabidopsis mitochondrial P-type pentatricopeptide repeat protein-modulating editing protein, PPME, in nad1 RNA editing. *RNA Biol.* **2016**, *13*, 593–604. [[CrossRef](#)]
17. Sun, T.; Bentolila, S.; Hanson, M.R. The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.* **2016**, *21*, 962–973. [[CrossRef](#)]
18. Bayer-Császár, E.; Haag, S.; Jörg, A.; Glass, F.; Härtel, B.; Obata, T.; Meyer, E.H.; Brennicke, A.; Takenaka, M. The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochim. Biophys. Acta Gene Regul. Mech.* **2017**, *1860*, 813–828. [[CrossRef](#)]
19. Schallenberg-Rüdinger, M.; Knoop, V. Coevolution of organelle RNA editing and nuclear specificity factors in early land plants. In *Genomes and Evolution of Charophytes, Bryophytes and Ferns. Advances in Botanical Research*; Rensing, S.A., Ed.; Elsevier Academic Press: Amsterdam, The Netherlands, 2016; Volume 78.
20. Gerke, P.; Szövényi, P.; Neubauer, A.; Lenz, H.; Gutmann, B.; McDowell, R.; Small, I.; Schallenberg-Rüdinger, M.; Knoop, V. Towards a Plant Model for Enigmatic U-to-C RNA Editing: The Organelle Genomes, Transcriptomes, Editomes and Candidate RNA Editing Factors in the Hornwort *Anthoceros Agrestis*. *New Phytologist* **2020**, *225*, 1974–1992. [[CrossRef](#)]
21. Oldenkott, B.; Yang, Y.; Lesch, E.; Knoop, V.; Schallenberg-Rüdinger, M. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Commun. Biol.* **2019**, *2*, 85. [[CrossRef](#)]
22. Gutmann, B.; Royan, S.; Schallenberg-Rüdinger, M.; Lenz, H.; Castleden, I.R.; McDowell, R.; Vacher, M.A.; Tonti-Filippini, J.; Bond, C.S.; Knoop, V.; et al. Small, the expansion and diversification of pentatricopeptide repeat RNA-editing factors in plants. *Mol. Plant* **2019**. [[CrossRef](#)]
23. Qulsum, U.; Azad, T.A.; Tsukahara, T. Analysis of tissue-specific RNA editing events of genes involved in RNA editing in *Arabidopsis thaliana*. *J. Plant Biol.* **2019**, *62*, 351–358. [[CrossRef](#)]
24. Qulsum, U.; Tsukahara, T. Tissue-specific alternative splicing of pentatricopeptide repeat (PPR) family genes in *Arabidopsis thaliana*. *Biosci. Trends* **2019**, *12*, 569–579. [[CrossRef](#)] [[PubMed](#)]
25. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
26. Ramaswami, G.; Zhang, R.; Piskol, R.; Keegan, L.P.; Deng, P.; O’Connell, M.A.; Li, J.B. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **2013**, *10*, 128–132. [[CrossRef](#)] [[PubMed](#)]
27. Yang, E.; van Nimwegen, E.; Zavolan, M.; Rajewsky, N.; Schroeder, M.; Magnasco, M.; Darnell, J.E., Jr. Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Res.* **2003**, *13*, 1863–1872.

28. Drescher, A.; Hupfer, H.; Nickel, C.; Albertazzi, F.; Hohmann, U.; Herrmann, R. Maier R C-to-U conversion in the intergenic ndhI/ndhG RNA of plastids from monocot plants: Conventional editing in an unconventional small reading frame? *Mol. Genet. Genom.* **2002**, *267*, 262–269. [[CrossRef](#)]
29. Farré, J.; Akinin, C. Alejandro araya BC RNA editing in mitochondrial for splicing Trans -Introns is required for splicing. *PLoS ONE* **2012**, *7*, 1–10. [[CrossRef](#)]
30. Wu, B.; Chen, H.; Shao, J.; Zhang, H.; Wu, K.; Liu, C. Identification of symmetrical RNA editing events in the mitochondria of *Salvia miltiorrhiza* by strand-specific RNA sequencing. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)]
31. Meng, Y.; Chen, D.; Jin, Y.; Mao, C.; Wu, P.; Chen, M. RNA editing of nuclear transcripts in *Arabidopsis thaliana*. *BMC Genom.* **2010**, *11*, 1–7. [[CrossRef](#)]