PLANT SIGNALING & BEHAVIOR
2017, VOL. 12, NO. 5, e1312242 (9 pages)
https://doi.org/10.1080/15592324.2017.1312242

Taylor & Francis
Taylor & Francis Group

ARTICLE ADDENDUM

Whole-transcriptome RNA-seq, gene set enrichment pathway analysis, and exon coverage analysis of two plastid RNA editing mutants

Justin B. Hackett* and Yan Lu

Department of Biological Sciences, Western Michigan University, Kalamazoo, MI, USA

ABSTRACT

In land plants, plastid and mitochondrial RNAs are subject to post-transcriptional C-to-U RNA editing. T-DNA insertions in the *ORGANELLE RNA RECOGNITION MOTIF PROTEIN6* gene resulted in reduced photosystem II (PSII) activity and smaller plant and leaf sizes. Exon coverage analysis of the *ORRM6* gene showed that *orrm6–1* and *orrm6–2* are loss-of-function mutants. Compared to other ORRM proteins, ORRM6 affects a relative small number of RNA editing sites. Sanger sequencing of reverse transcription-PCR products of plastid transcripts revealed 2 plastid RNA editing sites that are substantially affected in the *orrm6* mutants: *psbF*-C77 and *accD*-C794. The *psbF* gene encodes the β subunit of cytochrome b_{559}, an essential component of PSII. The *accD* gene encodes the β subunit of acetyl-CoA carboxylase, a protein required in plastid fatty acid biosynthesis. Whole-transcriptome RNA-seq demonstrated that editing at *psbF*-C77 is nearly absent and the editing extent at *accD*-C794 was significantly reduced. Gene set enrichment pathway analysis showed that expression of multiple gene sets involved in photosynthesis, especially photosynthetic electron transport, is significantly upregulated in both *orrm6* mutants. The upregulation could be a mechanism to compensate for the reduced PSII electron transport rate in the *orrm6* mutants. These results further demonstrated that Organelle RNA Recognition Motif protein ORRM6 is required in editing of specific RNAs in the Arabidopsis (*Arabidopsis thaliana*) plastid.

ARTICLE HISTORY
Received 13 March 2017
Revised 18 March 2017
Accepted 23 March 2017

KEYWORDS
Arabidopsis thaliana; exon coverage analysis; generally applicable gene-set enrichment for pathway analysis; organelle RNA recognition motif; plastid RNA editing; Sanger sequencing; wholetranscriptome RNA-seq

Introduction

RNA editing is modification of specific nucleotides/nucleosides within RNA molecules. Types of RNA editing include nucleotide insertion/deletion, cytidine-to-uridine (C-to-U) and adenosine-to-inosine (A-to-I) deamination, and uridine-to-cytidine (U-to-C) reverse editing.[1,2] RNA editing has been found in viruses, primitive eukaryotes, vertebrates, fungi, and plants, and it occurs in the nucleus, the cytosol, mitochondria, and plastids. RNA editing in plants has been observed in different types of RNAs, but is most abundant in the coding regions of mRNAs.[3] RNA editing in the coding regions of mRNAs restores the function of the encoded protein; therefore, it is considered as a mechanism of correction to compensate for defects in the genomes. Organellar (plastid and mitochondrial) RNA editing occurs in almost all land plants.[3-6] Two types of RNA editing have been found in plastids and/or mitochondria of flowering plants: C-to-U editing in mRNAs and C-to-U and A-to-I editing in tRNAs.[7-10] C-to-U editing in plastid and mitochondrial mRNAs appear to be ubiquitous in land plants.

Prior to the widespread use of next-generation sequencing technologies, detection and quantitation of RNA editing events was laborious and limited by the resolution and strand length of Sanger sequencing as well as *a priori* knowledge of primer sets. High-throughput methods using PCR technology[11] brought the number of discovered plastid editing events up to 34. Of the 34 known events, 2 are within non-coding regions. All editing events within the coding regions result in single amino acid changes, except for one which results in a new translation start site. With next-generation sequencing technologies, we are able to simultaneously detect and quantify editing without *a priori* knowledge of the sites. The use of RNA-seq technology led to the discovery of a total of 40 plastid editing sites in Columbia (Col-0) wild-type Arabidopsis.[12,13] Of the 6 newly discovered sites, only 2 are in the coding region of a gene. The 2 coding region editing events occur within the *ndhB* gene, change the third base in the corresponding codons, and result in 2 silent mutations. The other 4 sites are within intronic, intergenic, or 3' UTR regions and are of low-editing-extent ($\leq$12%). The 6 new sites raise questions about the origin of new editing events and the fidelity of editosomes (*i.e.*, editing complexes). Editing events with such low extent and non-obvious biologic implications suggest that some infidelity can exist without detrimental effects and negative selection pressure. It is also possible that the existence of these editing events could be the mechanism that allows for the selection of novel editing events.

C-to-U RNA editing is performed by editosomes of 200 – 400 kDa in size.[14,15] Four types of proteins have been identified as C-to-U RNA editing factors in the plastid: PLS-E-DYW subfamily pentatricopeptide repeat proteins (PPR-DYWs), RNA editing interacting proteins / multiple organellar RNA editing factors (RIPs/MORFs), RNA recognition motif-containing

CONTACT Yan Lu yan.1.lu@wmich.edu Department of Biological Sciences, Western Michigan University, Kalamazoo, MI 49008, USA.
*Present address: MD/PHD Program, School of Medicine, Wayne State University, Detroit, MI 48201, USA.

proteins (ORRMs), and organelle zinc-finger proteins (OZs).[15-20] As described below, each of the 4 types of proteins is considered as a component of the C-to-U RNA editosomes.

PPR-DYW proteins contain multiple PLS-type PRR repeats, an extension (E) domain, and a C-terminal DYW domain.[18] *In silico* search of the Arabidopsis genome identified 87 PPR-DYW proteins, 28% of which are targeted to the plastid, 36% of which are targeted to the mitochondrion.[21] PLS-type PRR repeats allow PPR-DYW proteins to bind single-stranded RNAs in a sequence-dependent manner.[18] The E domain is essential for recruiting a protein with cytidine deaminase activity.[22] The DYW domain contains 3 conserved amino acids—DYW, as well as signature amino acids of classic cytidine deaminases, including the HXE motif for pyrimidine protonation and CXXC for zinc coordination.[22-24] Although 7 classic cytidine deaminases are encoded by the Arabidopsis genome, none of them are active in organellar RNA editing.[3,25] Therefore, PPR-DYW proteins are currently the prime candidate for deaminase activity in plant organelle RNA editing.[3,18] For example, LOW PHOTOSYSTEM II ACCUMULATION 66 (LPA66), a plastid-targeted PPR-DYW protein, is specifically required for RNA editing at the *psbF*-C77 site.[26] C77 is the nucleotide number of the cytidine [C] target relative to the nucleotide A of the translation initiation codon ATG in the *psbF* transcript, which encodes the β subunit of cytochrome b_{559}. A second example is REQUIRED FOR *ACCD* RNA EDITING 1 (RARE1), another plastid-targeted PPR-DYW protein. RARE1 is essential for RNA editing at cytidine 794 (C794) of the *accD* transcript,[27] which encodes the β subunit of acetyl-coenzyme A carboxylase carboxyl transferase. Proteins with PPR and DYW domains do not necessarily specify sequence and perform deamination simultaneously. The DYW domain of PPR-DYW proteins may be recruited by PPR proteins without a DYW domain to carry out deamination activity.[24]

RIP/MORF proteins contain conserved RIP/MORF boxes. [16] The Arabidopsis genome encodes 9 functional RIP/MORF proteins: RIP1/MORF8 is dual-targeted to plastids and mitochondria, RIP2/MORF2 and RIP9/MORF9 are targeted to the plastid, and the rest are targeted to the mitochondrion.[14,16] Unlike PPR-DYW proteins, which are RNA-sequence specific, RIP/MORF proteins are broadly involved in plastid and/or mitochondrial RNA editing. Fourteen plastid sites and 266 mitochondrial sites are affected in the *rip1*/*morf8* mutant and nearly all plastid sites affected in *rip2*/*morf2* and *rip9*/*morf9* mutants.[14,16] RIP/MORF proteins have been found to interact with PPR-DYW proteins via the RIP/MORF box. For example, RIP1/MORF8 interacted with plastid-targeted RARE1 and mitochondrion-targeted MITOCHONDRIAL RNA EDITING FACTOR 10 (MEF10)[14,28] RIP/MORF proteins were also found to interact with themselves and other RIP/MORF proteins, suggesting that these proteins may form homo- and hetero-oligomers.[16]

ORRM proteins contain a RRM; 5 ORRM proteins have been found to be involved in plastid or mitochondrial RNA editing.[15,17,19,29] The plastid-targeted ORRM1 is the founding member of this protein family.[17] Unlike other ORRMs, ORRM1 contains 2 RIP/MORF boxes, which are required for its interaction with PPR-DYW protein RARE1. The *orrm1* mutant showed near complete loss of editing at 12 plastid sites. ORRM2, ORRM3, and ORRM4 are targeted to the mitochondrion and none of them have RIP/MORF boxes. [15,19] The *orrm2, orrm3*, and *orrm4* mutants displayed decreased editing extents at 35, 32, and 262 mitochondrial RNA editing sites, respectively. ORMM6 is targeted to the plastid and it does not have RIP/MORF boxes. [29] The *orrm6* mutants demonstrated reduced editing extents at the *psbF*-C77 and *accD*-C794 RNA editing sites. ORRM3 was found to interact with RIP1/MORF8, ORRM2, and itself; ORRM4 was found to interact with ORRM3 and itself; and ORRM6 was found to interact with RIP1/MORF8, RIP2/MORF2, RIP9/MORF9, and itself.[29] The interactions among ORRMs suggest that ORRMs may form homo- and/or hetero-oligomers.

OZ proteins contain multiple Ran-binding-protein-2 (RanBP2, $CXXCX_{10}CXXC$) type zinc-finger domains.[20] The Arabidopsis genome encodes 4 OZ proteins: OZ1 is targeted to the plastid; OZ2, OZ3, and OZ4 are targeted to the mitochondrion. A loss-of-function mutation in the *OZ1* gene resulted in major loss of editing at 14 plastid sites and significant changes in the editing extent at 16 other plastid sites.[20] Despite of the large number of editing sites altered in the *oz1* mutant, C targets on the same transcripts are differentially affected, suggesting that OZ1 action is site-specific. Consistent with this hypothesis, some RanBP2-type zinc fingers have been shown to bind single-stranded RNAs on a sequence-specific manner.[30] OZ1 interacted with itself, ORRM1, RIP1/MORF8, as well as plastid-targeted PPR-DYW proteins CHLORORESPIRATORY REDUCTION 28 (CRR28) and ORGANELLE TRANSCRIPT PROCESSING 82 (OTP82).[20] Like RIP/MORF and ORRM proteins, OZ1 interacted with itself in yeast 2-hybrid assays, suggesting that OZ1 may form homooligomers.

In this work, we used a combination of Sanger sequencing and whole-transcriptome RNA-seq to explore the editing site specificity of ORRM6, a unique plastid-targeted RNA editing factor. In addition, we conducted gene set enrichment pathway analysis and identified several Gene Ontology gene sets that were significantly and consistently upregulated in the loss-of-function *orrm6* mutants.

Exon coverage analysis confirmed that orrm6–1 and orrm6–2 are loss-of-function mutants with truncated transcripts

Quantitative RT-PCR showed that the functional *ORRM6* transcript (*i.e.*, the full length *ORRM6* transcript without the T-DNA insertion) cannot be detected in either *orrm6* mutant (Fig. 1 in ref.[29]). This suggests that *orrm6–1* and *orrm6–2* are loss-of-function mutants. To confirm this, BEDtools and Integrative Genomics Viewer were used to examine the exon coverage of the *ORRM6* transcript (Fig. 1) with RNA-seq data from the wild type and the *orrm6* mutants. The relative abundance of the first exon in the *orrm6* mutants is approximately 60% lower than that in the wild type (Fig. 1B). In the *orrm6* mutants, exons 2-4 are almost completely absent (Fig. 1). These data confirmed that T-DNA insertions in the first intron of the *ORRM6* gene result in a truncated *ORRM6* transcript lacking

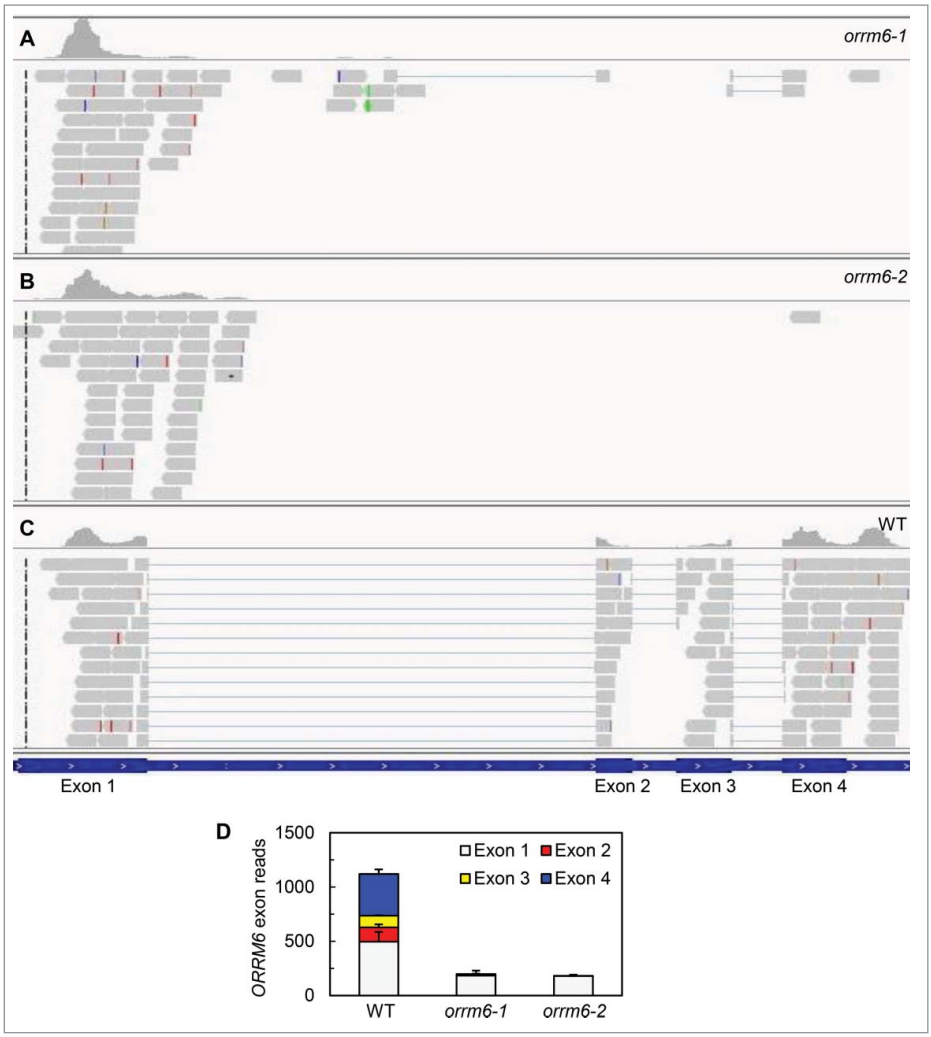

Figure 1. Analysis of *ORRM6* transcript levels in the wild-type and *orrm6* mutant plants. (A-C) Visual representation of exon reads in RNA samples prepared from *orrm6–1, orrm6–2,* and wild-type (WT) plants. Integrated Genomic Viewer was used to visually display exon coverage of *ORRM6*. Each rectangle represents a single 50 bp read and thin lines show reads spanning introns. (D) Numbers of *ORRM6* exon reads in RNA samples prepared from the wild-type and *orrm6* mutant plants. Total RNA was extracted from mature leaves and analyzed with whole-transcriptome RNA-seq. The values (mean + SE, n = 4) have been normalized by the library size.

exons 2–4. The first exon encodes the first 82 amino acids, which correspond to the chloroplast transit peptide and the region between the chloroplast transit peptide and the RRM. Therefore, it is likely that the correspondingly truncated ORRM6 protein does not contain the RRM. T-DNA insertions could also result in early termination codons and the Integrative Genome Viewer detects improper splicing of the *orrm6–1* transcript. Both of these could lead to nonsense-mediated decay of the *ORRM6* transcript.[31]

Sanger sequencing and whole-transcriptome RNA-seq revealed that ORRM6 is required at 2 plastid RNA editing sites: psbF-C77 and accD-C794

Sanger sequencing of reverse transcription-PCR products from the wild-type and *orrm6* mutant plants suggested that ORRM6 is required in editing at 2 plastid RNA editing sites: *psbF*-C77 and *accD*-C794. To investigate whether ORRM6 plays additional roles, we performed Sanger sequencing of reverse transcription-PCR products of other plastid transcripts (Fig. 2). The results from Sanger sequencing revealed that other plastid RNA editing sites are not substantially affected by the loss-of-function mutations in the *ORRM6* gene (Fig. 2).

To quantitatively determine the editing extents at *accD*-C794, *psbF*-C77, and other plastid RNA editing sites, we performed whole-transcriptome RNA-seq of Illumina TruSeq libraries made from total mRNAs, as described previously.[17,32] In brief, total leaf RNAs were extracted from the wild type and the *orrm6* mutants (4 biologic replicates per genotype), rRNAs in the total RNA samples were depleted before mRNA library construction, and the resulting mRNA libraries were sequenced on an Illumina HiSeq 2500. After removal of adaptor sequences and initial checks on sequence quality, sequences (reads) were mapped to a reference genome and assembled into a table with the TopHat software and were visualized with the Integrative Genomics Viewer.[32,33] The proportion of edited and unedited mRNAs was calculated for each of the known plastid RNA editing sites. Whole-transcriptome RNA-Seq data identified 2 plastid RNA editing sites that are substantially and consistently reduced in both *orrm6–1* and *orrm6–2* mutants: *psbF*-C77 and *accD*-C794. The editing extent at *accD*-C794 was 85.8% in the wild type and it was reduced to 30.7% in *orrm6–1* and 29.0% in

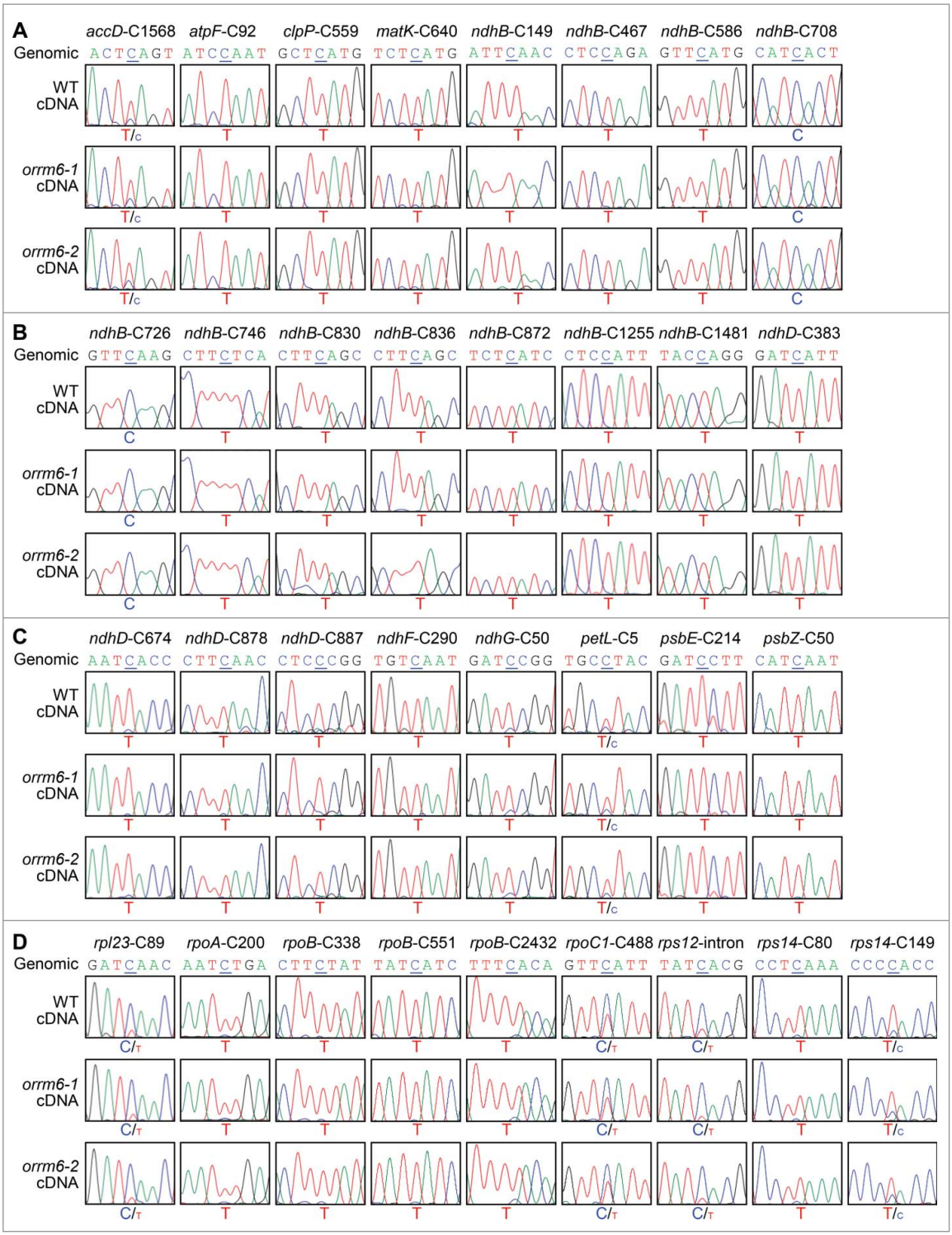

Figure 2. Sanger sequencing of plastid RNA editing sites in the wild-type and *orrm6* mutant plants. (A) Analysis of RNA editing at *accD*-C1568, *atpF*-C92, *clpP*-C559, *matK*-C640, *ndhB*-C149, *ndhB*-C467, *ndhB*-C586, and *ndhB*-C708. (B) Analysis of RNA editing at *ndhB*-C726, *ndhB*-C746, *ndhB*-C830, *ndhB*-C836, *ndhB*-C872, *ndhB*-C1255, *ndhB*-C1481, and *ndhD*-C383. (C) Analysis of RNA editing at *ndhD*-C674, *ndhD*-C878, *ndhD*-C887, *ndhF*-C290, *ndhG*-C50, *petL*-C5, *psbE*-C214, and *psbZ*-C50. (D) Analysis of RNA editing at *rpl23*-C89, *rpoA*-C200, *rpoB*-C338, *rpoB*-C551, *rpoB*-C2432, *rpoC1*-C488, *rps12*-intron, *rps14*-C80, and *rps14*-C149. RT-PCR products surrounding the editing sites were directly sequenced. The 7-nucleotide sequences encompassing the cytidine target (underlined) were shown. The corresponding genomic sequences of these 2 sites were displayed as controls.

orrm6–2; the editing extent at *psbF*-C77 was 97.1% in the wild type and it was reduced to 6.8% in *orrm6–1* and 6.3% in *orrm6–2* (Table 1). In addition to *psbF*-C77 and *accD*-C794, whole-transcriptome RNA-seq showed that editing at the intergenic site between *ndhK* and *ndhJ* is absent in both *orrm6* mutants (Table 1). The biologic significance of this intergenic RNA editing site is not clear. In the wild-type Arabidopsis, only approximately 4.2% of the C target at this RNA editing site was changed to T (Table 1). It is possible that some infidelity exists without detrimental effects on the plant. Taken together, these

Table 1. Editing extents at plastid RNA editing sites in wild-type and *orrm6* mutant plants.

Gene	Genome position[a]	Edited nucleotide[b]	Δ amino acid	WT	*orrm6–1*	*orrm6–2*	*orrm6* - WT[c]
accD	57868	C794	S265→L	85.8 ± 3.2	30.7 ± 3.0***	29.0 ± 2.0***	−56.0
accD	58642	C1568	3′-UTR	83.5 ± 1.6	78.8 ± 2.2	78.3 ± 1.3	−4.9
atpF	12707	C92	P31→L	94.9 ± 0.2	95.1 ± 0.3	93.8 ± 0.4	−0.4
atpH	13210	C298	3′-UTR	4.4 ± 0.6	4.5 ± 0.3	4.9 ± 0.3	0.3
clpP	69942	C559	H187→Y	81.3 ± 0.5	82.8 ± 0.7	78.0 ± 0.7**	−1.4
matK	2931	C640	H214→Y	71.9 ± 8.6	74.4 ± 4.8	79.9 ± 0.3	5.2
ndhB	97016	C149	S50→L	56.9 ± 19.9	66.6 ± 11.4	67.3 ± 5.6	10.1
ndhB	96698	C467	P156→L	76.6 ± 4.9	84.0 ± 0.8	86.1 ± 0.5	8.5
ndhB	96579	C586	H196→Y	74.6 ± 3.2	78.8 ± 1.3	81.9 ± 1.0	5.8
ndhB	96457	C708	I236→I	3.6 ± 0.5	3.2 ± 0.2	3.2 ± 0.4	−0.4
ndhB	96439	C726	F242→F	3.5 ± 0.4	3.1 ± 0.3	2.7 ± 0.2	−0.6
ndhB	96419	C746	S249→F	84.8 ± 5.8	72.5 ± 3.3	76.6 ± 2.2	−10.3
ndhB	95650	C830	S277→L	74.7 ± 1.1	75.9 ± 2.0	77.5 ± 1.7	2.0
ndhB	95644	C836	S279→L	66.2 ± 2.5	61.2 ± 2.3	61.0 ± 2.5	−5.0
ndhB	95608	C872	S291→L	61.9 ± 4.5	64.9 ± 1.0	59.3 ± 4.5	0.2
ndhB	95225	C1255	H419→Y	87.8 ± 1.1	87.5 ± 1.4	88.0 ± 1.4	0.0
ndhB	94999	C1481	P494→L	88.2 ± 1.0	84.3 ± 0.7*	85.2 ± 0.7	−3.5
ndhD	117166	C2	T1→M	46.4 ± 4.1	36.8 ± 5.7	51.2 ± 2.0	−2.5
ndhD	116785	C383	S128→L	95.4 ± 1.2	96.9 ± 0.3	96.7 ± 0.8	1.4
ndhD	116494	C674	S225→L	73.8 ± 2.0	74.3 ± 4.9	78.3 ± 2.0	2.5
ndhD	116290	C878	S293→L	79.1 ± 2.0	83.7 ± 1.3	86.2 ± 0.5*	5.8
ndhD	116281	C887	P296→L	76.7 ± 5.7	87.9 ± 0.6	87.9 ± 0.9	11.2
ndhF	112349	C290	S97→L	23.4 ± 4.8	21.4 ± 3.6	26.8 ± 2.7	0.7
ndhG	118858	C50	S17→F	68.4 ± 6.3	81.3 ± 4.0	75.9 ± 1.4	10.2
ndhK-ndhJ	49209	C726[d]	Intergene	4.2 ± 0.5	0.0 ± 0.0**	0.0 ± 0.0**	−4.2
petL	65716	C5	P2→L	35.1 ± 12.3	50.4 ± 10.4	55.8 ± 11.4	18.0
psbE	64109	C214	P72→S	99.3 ± 0.1	99.1 ± 0.2	99.2 ± 0.0	−0.2
psbF	63985	C77	S26→F	97.1 ± 0.3	6.8 ± 0.2***	6.3 ± 0.3***	−90.5
psbZ	35800	C50	S17→L	61.3 ± 8.6	49.6 ± 8.8	58.9 ± 5.0	−7.1
rpl23	86055	C89	S30→L	82.3 ± 1.9	78.1 ± 0.9	78.5 ± 0.3	−4.0
rpoA	78691	C200	S67→F	58.9 ± 5.4	64.2 ± 4.4	61.7 ± 4.6	4.0
rpoB	25992	C338	S113→L	65.9 ± 16.6	70.4 ± 17.7	65.4 ± 8.3	1.9
rpoB	25779	C551	S184→L	38.1 ± 9.1	30.8 ± 7.7	27.2 ± 2.5	−9.2
rpoB	23898	C2432	S811→L	69.9 ± 7.0	70.0 ± 3.3	60.3 ± 4.9	−4.8
rpoC1	21806	C488	S163→L	20.8 ± 2.9	26.7 ± 1.2	27.4 ± 2.9	6.2
rps4	45095	C734	3′-UTR	3.2 ± 0.2	1.9 ± 0.5	2.5 ± 0.2	−1.0
rps12-intron	69553	C(i1 58)[e]	intron	19.5 ± 1.1	13.5 ± 3.2	18.6 ± 4.0	−3.5
rps14	37161	C80	S27→L	96.4 ± 0.4	95.1 ± 0.2	94.8 ± 0.3*	−1.4
rps14	37092	C149	P50→L	87.6 ± 2.3	91.0 ± 0.2	90.7 ± 0.4	3.2
ycf3-intron	43350	C(i2 174)[f]	intron	7.5 ± 0.9	4.5 ± 0.3*	7.8 ± 0.5	−1.4

The editing extents were determined with the whole-transcriptome RNA-seq method. The values (mean ± SE, n = 4) are given as the ratio of the edited transcript to edited plus unedited transcripts. The asterisk indicates significant difference between the mutant and the wild type (WT) (Student's *t*-test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

[a] The Arabidopsis plastid genome reference sequence (NCBI accession number NC_000932.1) was used to indicate the genome positions of the editing sites.

[b] TheC targets are numbered relative to the nucleotide A of the predicted translation initiation codon ATG, where A = 1.

[c] This column shows the average difference in editing extents between the wild type (WT) and *orrm6* mutants.

[d] The C target is numbered relative to the nucleotide A of ATG of the *ndhK* transcript, where A = 1.

[e] The target C is the 58th nucleotide of intron 1 (i1 58).

[f] The target C is the 174th nucleotide of intron 2 (i2 174).

whole-transcriptome RNA-Seq data are consistent with the results from Sanger sequencing, strand-and-transcript-specific PCR sequencing, and poisoned primer extension.[29]

Gene set enrichment pathway analysis revealed that expression of genes involved photosynthesis is significantly upregulated in the orrm6 mutants

It was previously found with quantitative reverse transcription (RT)-PCR that the transcript levels of photosystem II (PSII) genes *psbA*, *psbB*, and *psbI* (where psb refers to PSII) are significantly increased in the *orrm6–2* mutant (see Supplemental Fig. S3 in ref.[29]). To further investigate gene expression changes in *orrm6* mutants versus wild type, gene set enrichment analysis was performed using the Generally Acceptable Gene-set Enrichment (GAGE) software and the Gene Ontology biologic process gene sets.[34,35] Pathways were discovered by comparing gene sets in the Gene Ontology biologic process pathways for Arabidopsis using normalized counts for the wild type vs. a single mutant line and looking for significance. Significantly upregulated pathways are shared among the 2 *orrm6* mutants (Table 2). Of the shared upregulated pathways, 5 of them are gene sets for photosynthetic processes. These data are consistent with the previous quantitative RT-PCR analysis which showed increased transcript levels of PSII related genes despite decreased PSII efficiency.[29] The increased expression of photosynthetic genes might be a compensatory response to the reduced PSII electron transport rate in the *orrm6* mutants.

ORRM6 is a unique C-to-U RNA editing factor in the plastid

ORRM6 differs from ORRM1, the other plastid-targeted ORRM protein, in that it does not contain any RIP/MORF boxes. The 2 RIP/MORF boxes in ORRM1 are required for its

Table 2. Gene sets that were significantly upregulated in the *orrm6–1* and *orrm6–2* mutants.

WT *vs. orrm6–1*	*p* value	WT *vs. orrm6–2*	*p* value
DNA-templated transcription, elongation	0.020	DNA-templated transcription, elongation	0.012
Photosynthetic electron transport in PSII	0.030	Photosynthesis	0.021
Photosynthetic electron transport chain	0.031	Photosynthetic electron transport chain	0.025
Electron transport chain	0.032	Electron transport chain	0.025
PSII assembly	0.033	PSII assembly	0.025
Photosynthesis	0.045	Photosynthesis, light reaction	0.026
Photosynthesis, light reaction	0.049	Photosynthetic electron transport in PSII	0.028

Gene sets significantly upregulated in the *orrm6* mutants were discovered through the Generally Acceptable Gene-set Enrichment pathway analysis of the RNA-seq data, using the Gene Ontology biologic process gene sets.

interaction with PPR-DYW proteins.[17] Consistent with the absence of RIP/MORF boxes in ORRM6, this protein did not interact with PPR-DYW proteins LPA66 and RARE1 in reciprocal bimolecular fluorescence complementation (BiFC) assays.[29] LPA66 and RARE1 are required for editing at *psbF*-C77 and *accD*-C794, respectively.[26,27] Interestingly, ORRM6 was found to interact with RIP1/MORF8, RIP2/MORF2, and RIP9/MORF9, when transiently co-expressed in *Nicotiana benthamiana* leaves.[29] Therefore we hypothesize that the interaction between ORRM6 and RIP/MORF proteins may compensate for the absence of RIP/MORF boxes in ORRM6. ORRM6 was also found to interact with OZ1 and itself in BiFC assays.[29] Therefore, we propose that the editosome at *psbF*-C77 may require LPA66, ORRM6, RIP/MORF proteins (RIP1/MORF8, RIP2/MORF2, and/or RIP9/MORF9), and OZ1, and the editosome at *accD*-C794 may require RARE1, ORRM6, RIP/MORF proteins (RIP1/MORF8, RIP2/MORF2, and/or RIP9/MORF9), and OZ1 (Fig. 3).

ORRM6 is also unique due to the relatively small number of RNA editing sites (*psbF*-C77 and *accD*-C794) that are affected in the *orrm6* mutants. On the contrary, loss-of-function mutations in the *ORRM1* gene resulted in near complete loss of editing at 12 plastid sites,[17] and loss-of-function mutations in the *ORRM2, ORRM3*, and *ORRM4* genes resulted in decreased editing at 35, 32, and 262 mitochondrial sites, respectively.[15,20] Therefore, it is possible that different ORRM proteins have different editing site specificities.

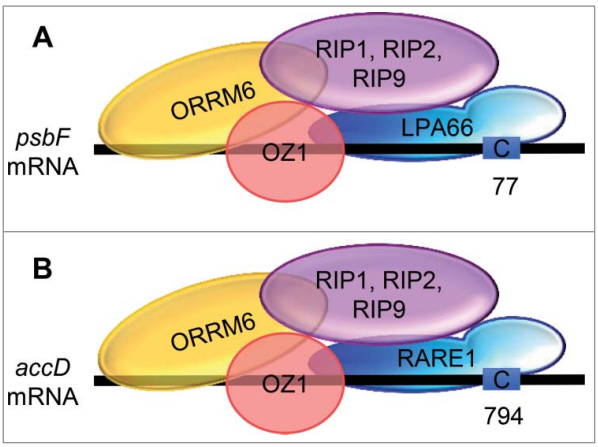

Figure 3. Tentative models for editosomes at the *psbF*-C77 and *accD*-C794 RNA editing sites based on protein-protein interaction data. (A) Model for the editosome at *psbF*-C77. (B) Model for the editosome at *accD*-C794. For simplicity, only one name is shown for proteins with multiple names (*e.g.*,"RIP1" for RIP1/MORF8). Black lines represent transcripts; the letter C in blue boxes represents the cytidine target.

Materials and methods

Plant materials and growth conditions

The 2 Arabidopsis (*Arabidopsis thaliana*) T-DNA insertion lines (*orrm6–1* [SAIL_763_A05] and *orrm6–2* [WiscDsLox485–488P23], both in Columbia ecotype) were obtained from the Arabidopsis Biological Resource Center. Plants were grown in a growth chamber on a 12-h-light/12-h-dark photoperiod, as described previously.[36] The light intensity was 150 μmol photons m^{-2} s^{-1}, the temperature was 20°C, and the relative humidity was 50%.

Analysis of RNA editing by Sanger sequencing

The transcript regions surrounding Arabidopsis plastid RNA editing sites were amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and primers listed in Table 3. The PCR products were sequenced at the Michigan State University Genomics Facility, using the Sanger method and the primers listed in Table 3.

Analysis of RNA editing extents by whole-transcriptome RNA-Seq

Total leaf RNAs were extracted from the wild type and the *orrm6* mutants (4 biologic replicates per genotype) using the RNeasy Plant Mini Kit (QIAGEN), digested with the RNase-Free DNase I (QIAGEN), re-purified with the RNeasy Plant Mini Kit, and submitted to the Michigan State University Genomics Facility for downstream processes. rRNAs were depleted with the Ribo-Zero rRNA Removal kit for plant leaf (Illumina). Sequencing libraries were prepared using the TruSeq Stranded Total RNA Low Throughput Library Prep Kit (Illumina). The libraries were first run on a LabChip GX system (Perkin Elmer) to assess the size distribution of the library fragments and estimate a mean fragment size. The libraries were then quantified using both a Qubit fluorimetric assay and a Kapa Biosystems library quantification qPCR method to determine the molar concentration of each library and adjust each library to a standard concentration. After quality control and quantitation, the 12 libraries (3 genotypes—wild type, *orrm6–1*, and *orrm6–2*; 4 biologic replicates per genotype) were pooled on an equal equimolar amount basis and loaded on 2 lanes of an Illumina HiSeq 2500 Rapid Run flow cell (v1). Base calling was done by the Illumina Real Time Analysis (RTA) v1.17.21.3 software and output of RTA was demultiplexed and converted to the FastQ format by the Illumina Bcl2fastq v1.8.4 software.

Prior to analysis of RNA-Seq data, the adaptor sequences were removed using trimmomatic v 0.32 software (http://www.usadellab.org/cms/?page = trimmomatic).[37] The resulting data files were filtered by FASTQ Quality Filter in the FASTX-toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) so that

Table 3. Primers used to amplify and sequence 34 plastid RNA editing sites.

Primer	Sequence
accD_1_F	TTCATTTGTAGTGAAAGCGG
accD_1_R	TTTCGCCTACTACGGATCCC
accD_2_F	CTACTACCGGTGGAGTGACAGC
accD_2_R	AGAATCTGATCTAACAACAGGGAA
atpF_1_F	GAGTTTCGGATTTAATACCG
atpF_1_R	AGCTCCTTGTAAAGCTTGTTG
clpP_1_F	TTGGGTTGACATATACAACCG
clpP_1_R	TGAACCGCTACAAGATCAAC
matK_2_F	CGTTACCGGGTAAAAGATGC
matK_2_R	AGCGGCGTATCCTTTGTTGC
ndhB_1–3_F	TTTGCTTCTCTTCGATGGAAG
ndhB_1–3_R	ACGACTGGAGTGGGAGATCCTTC
ndhB_7–12_F	CGTATACGAAGGATCTCCCAC
ndhB_7–12_R	CCTGAGCAATCGCAATAATCG
ndhD_1–5_F	TTGAGTACGCGTTCTTTGGAC
ndhD_1–5_R	AATAGCTCCATTAAGTCCAGG
ndhF_2_F	AAAACCTTCGCCGCATGTGG
ndhF_2_R	GCATTCGCTGCAATAGGTCG
ndhG_1_F	ATGGATTTGCCTGGACCAATAC
ndhG_1_R	TTGATAAATGAATTCCTATTTGTTG
petL_2_F	AAATTTGGTAATTAACACGG
petL_2_R	ATTTCAATTGAAACTTAGGG
psbE_1_psbF_1_F	ACAGGAGAACGTTCTTTTGC
psbE_1_psbF_1_R	ATATAATCCATCCGAATGGG
psbZ_1_F	ATGAGATACGCGATCCAGTATAC
psbZ_1_R	TCAAGAGATAAGAGAATTAAGGATAC
rpl23_1_F	AAGAGGTGGAATAGAATAACCCG
rpl23_1_R	CAATTCCTACTGGATGCACGC
rpoA_1_F	GTAAGCGTCTTTATTATGGACGC
rpoA_1_R	CTTGATGAAGTGCTTCTTTAGGAG
rpoB_1_3_F	GAAAACCAGTAGGAATATGC
rpoB_1_3_R	GTCTCCAATTAATATTTCGGCG
rpoB_7_F	GAGGTGGGTTCAGAAAAAGG
rpoB_7_R	TATCTGTCCTACATTCATGCG
rpoC1_1_F	AGTTTTGTGAACAATGTGGAGTTG
rpoC1_1_R	TGAATGATGGGTCTCAACTCGG
rps12_1_F	CTTGTACAATTCACATTCTTTGGC
rps12_1_R	ACAAGACAGCCAATCCGAAAC
rps14_1–2_F	TTATAGGGAGAAGAAGAGGC
rps14_1–2_R	TACCAGCTTGATCTTGTTGC

only reads with at least 85% of base calls with a quality score of 20 were kept. The kept reads were aligned with the Tophat mapping software,[32] using the TAIR10 Arabidopsis genome released by Ensembl[38] as the reference genome. The Integrative Genomics Viewer (http://software.broadinstitute.org/software/igv/)[33] was used to navigate to RNA editing sites and editing extents were calculated by dividing T base calls by C + T base calls.[17] The 4 replicates per line were averaged and a student's 2-tailed *t*-test was performed between the wild type and the *orrm6* mutants.

Gene expression and pathway analyses of the RNA-seq data

Following alignment of reads as described above, differential gene expression analysis was performed using a combination of SAMtools (http://samtools.sourceforge.net/),[39] htseq-count (http://www-huber.embl.de/HTSeq/doc/overview.html),[40] and edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html)[41] programs and a previously developed protocol.[32] Pathway analysis was performed with the Generally Applicable Gene-set Enrichment (GAGE, http://bioconductor.org/packages/release/bioc/html/gage.html) program,[35] using the Gene Ontology for biologic process gene sets. During the GAGE pathway analysis, the "pseudo.counts" expression counts from the RNA-seq data were used because they accounted for the transformations performed to normalize read counts by library size. A paired *t*-test was performed between the wild type and the *orrm6* mutants; the resulting gene sets with statistically significant changes were outputted to text files using the geneData function.

Exon coverage analysis of the *ORRM6* gene

The start and end regions of the 4 exons in the *ORRM6* gene were defined using the TAIR Ensembl 10 genome annotation file.[38] Exon regions were compiled into a BED file for analysis with the program BEDtools.[42] The BEDtools coverage tool was used to examine read depth at the defined exons in the wild type, *orrm6–1* and *orrm6–2* (n = 4 for each genotype). The resulting read counts were combined into an Excel file and the ratios of exon counts to total read counts were compared. Visual inspection of exon coverage was done by navigating Integrative Genomics Viewer to the *ORRM6* gene.[33]

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Ryan L. Wessendorf (Western Michigan University [WMU]) for plant care, Christopher D. Jackson (WMU) for growth chamber management, John Kapenga and Elise de Doncker (WMU) for access to the WMU High Performance Computational Science Laboratory, and Cheng Peng (Michigan State University) for advices on RNA-seq data analysis.

Funding

This work was supported by the US. National Science Foundation under Grant Number MCB-1244008.

ORCID

Justin B. Hackett http://orcid.org/0000-0001-6532-4522
Yan Lu http://orcid.org/0000-0002-3374-7376

References

1. Covello PS, Gray MW. On the evolution of RNA editing. Trends Genet 1993; 9(8):265-8; PMID:8379005; https://doi.org/10.1016/0168-9525(93)90011-6
2. Brennicke A, Marchfelder A, Binder S. RNA editing. FEMS Microbiol Rev 1999; 23(3):297-316; PMID:10371035; https://doi.org/10.1111/j.1574-6976.1999.tb00401.x
3. Takenaka M, Zehrmann A, Verbitskiy D, Hartel B, Brennicke A. RNA editing in plants and its evolution. Annu Rev Genet 2013; 47:335-52; PMID:24274753; https://doi.org/10.1146/annurev-genet-111212-133519
4. Bock R. Sense from nonsense: How the genetic information of chloroplasts is altered by RNA editing. Biochimie 2000; 82(6–7):549-57; PMID:10946106; https://doi.org/10.1016/S0300-9084(00)00610-6
5. Chateigner-Boutin AL, Small I. Plant RNA editing. RNA Biol 2010; 7 (2):213-9; PMID:20473038; https://doi.org/10.4161/rna.7.2.11343

6. Shikanai T. RNA editing in plants: Machinery and flexibility of site recognition. Biochim Biophys Acta 2015; 1847(9):779-85; PMID:25585161; https://doi.org/10.1016/j.bbabio.2014.12.010
7. Fey J, Weil JH, Tomita K, Cosset A, Dietrich A, Small I, Marechal-Drouard L. Role of editing in plant mitochondrial transfer RNAs. Gene 2002; 286(1):21-4; PMID:11943456; https://doi.org/10.1016/S0378-1119(01)00817-4
8. Delannoy E, Le Ret M, Faivre-Nitschke E, Estavillo GM, Bergdoll M, Taylor NL, Pogson BJ, Small I, Imbault P, Gualberto JM. Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNAArg(ACG) and is essential for efficient chloroplast translation. Plant Cell 2009; 21(7):2058-71; PMID:19602623; https://doi.org/10.1105/tpc.109.066654
9. Karcher D, Bock R. Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. RNA 2009; 15(7):1251-7; PMID:19460869; https://doi.org/10.1261/rna.1600609
10. Zhou W, Karcher D, Bock R. Identification of enzymes for adenosine-to-inosine editing and discovery of cytidine-to-uridine editing in nucleus-encoded transfer RNAs of Arabidopsis. Plant Physiol 2014; 166(4):1985-97; PMID:25315605; https://doi.org/10.1104/pp.114.250498
11. Chateigner-Boutin A-L, Small I. A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. Nucleic Acids Res 2007; 35(17):e114; PMID:17726051; https://doi.org/10.1093/nar/gkm640
12. Bentolila S, Oh J, Hanson MR, Bukowski R. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. PLos Genet 2013; 9(6):e1003584; PMID:23818871; https://doi.org/10.1371/journal.pgen.1003584
13. Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB. Arabidopsis chloroplast quantitative editotype. FEBS Lett 2013; 587(9):1429-33; PMID:23523919; https://doi.org/10.1016/j.febslet.2013.03.022
14. Bentolila S, Heller WP, Sun T, Babina AM, Friso G, van Wijk KJ, Hanson MR. RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc Nat Acad Sci U S A 2012; 109(22):E1453-61; PMID:22566615; https://doi.org/10.1073/pnas.1121465109
15. Shi X, Germain A, Hanson MR, Bentolila S. RNA recognition motif-containing protein ORRM4 broadly affects mitochondrial RNA editing and impacts plant development and flowering. Plant Physiol 2016; 170(1):294-309; PMID:26578708; https://doi.org/10.1104/pp.15.01280
16. Takenaka M, Zehrmann A, Verbitskiy D, Kugelmann M, Härtel B, Brennicke A. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Nat Acad Sci U S A 2012; 109(13):5104-9; PMID:22411807; https://doi.org/10.1073/pnas.1202452109
17. Sun T, Germain A, Giloteaux L, Hammani K, Barkan A, Hanson MR, Bentolila S. An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc Nat Acad Sci U S A 2013; 110(12):E1169-78; PMID:23487777; https://doi.org/10.1073/pnas.1220162110
18. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 2014; 65:415-42; PMID:24471833; https://doi.org/10.1146/annurev-arplant-050213-040159
19. Shi X, Hanson MR, Bentolila S. Two RNA recognition motif-containing proteins are plant mitochondrial editing factors. Nucleic Acids Res 2015; 43(7):3814-25; PMID:25800738; https://doi.org/10.1093/nar/gkv245
20. Sun T, Shi X, Friso G, Van Wijk K, Bentolila S, Hanson MR. A zinc finger motif-containing protein is essential for chloroplast RNA editing. PLos Genet 2015; 11(3):e1005028; PMID:25768119; https://doi.org/10.1371/journal.pgen.1005028
21. Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16(8):2089-103; PMID:15269332; https://doi.org/10.1105/tpc.104.022236
22. Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J 2013; 74(6):935-45; PMID:23521509; https://doi.org/10.1111/tpj.12180
23. Boussardon C, Avon A, Kindgren P, Bond CS, Challenor M, Lurin C, Small I. The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. New Phytol 2014; 203(4):1090-5; PMID:25041347; https://doi.org/10.1111/nph.12928
24. Hayes ML, Dang KN, Diaz MF, Mulligan RM. A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. J Biol Chem 2015; 290(16):10136-42; PMID:25739442; https://doi.org/10.1074/jbc.M114.631630
25. Faivre-Nitschke SE, Grienenberger JM, Gualberto JM. A prokaryotic-type cytidine deaminase from Arabidopsis thaliana gene expression and functional characterization. Eur J Biochem 1999; 263(3):896-903; PMID:10469156; https://doi.org/10.1046/j.1432-1327.1999.00591.x
26. Cai W, Ji D, Peng L, Guo J, Ma J, Zou M, Lu C, Zhang L. LPA66 is required for editing psbF chloroplast transcripts in Arabidopsis. Plant Physiol 2009; 150(3):1260-71; PMID:19448041; https://doi.org/10.1104/pp.109.136812
27. Robbins JC, Heller WP, Hanson MR. A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast accD transcript. RNA 2009; 15(6):1142-53; PMID:19395655; https://doi.org/10.1261/rna.1533909
28. Härtel B, Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M. MEF10 is required for RNA editing at nad2-842 in mitochondria of Arabidopsis thaliana and interacts with MORF8. Plant Mol Biol 2013; 81(4–5):337-46; PMID:23288601; https://doi.org/10.1007/s11103-012-0003-2
29. Hackett JB, Shi X, Kobylarz AT, Lucas MK, Wessendof RL, Hines KM, Bentolila S, Hanson MR, Lu Y. An organelle RNA recognition motif protein is required for photosystem II subunit *psbF* transcript editing. Plant Physiol 2017; 173:2278-93; PMID:28213559; https://doi.org/10.1104/pp.16.01623
30. Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, Mackay JP. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. J Mol Biol 2011; 407(2):273-83; PMID:21256132; https://doi.org/10.1016/j.jmb.2010.12.041
31. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 2007; 76:51-74; PMID:17352659; https://doi.org/10.1146/annurev.biochem.76.050106.093909
32. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. Count-based differential expression analysis of RNA sequencing data using R and bioconductor. Nat Protoc 2013; 8(9):1765-86; PMID:23975260; https://doi.org/10.1038/nprot.2013.099
33. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotech 2011; 29(1):24-6; PMID:21221095; https://doi.org/10.1038/nbt.1754
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25(1):25-9; PMID:10802651; https://doi.org/10.1038/75556
35. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: Generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 2009; 10:161; PMID:19473525; https://doi.org/10.1186/1471-2105-10-161
36. Clark TJ, Lu Y. Analysis of loss-of-function mutants in aspartate kinase and homoserine dehydrogenase genes points to complexity in the regulation of aspartate-derived amino acid contents. Plant Physiol 2015; 168(4):1512-26; PMID:26063505; https://doi.org/10.1104/pp.15.00364
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics 2014; 30(15):2114-20; PMID:24695404; https://doi.org/10.1093/bioinformatics/btu170
38. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. Ensembl

genomes 2016: More genomes, more complexity. Nucleic Acids Res 2016; 44:D574-80; PMID:26578574; https://doi.org/10.1093/nar/gkv1209

39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25(16):2078-9; PMID:19505943; https://doi.org/10.1093/bioinformatics/btp352
40. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 2015; 31 (2):166-9; PMID:25260700; https://doi.org/10.1093/bioinformatics/btu638
41. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26(1):139-40; PMID:19910308; https://doi.org/10.1093/bioinformatics/btp616
42. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26(6):841-2; PMID:20110278; https://doi.org/10.1093/bioinformatics/btq033