Scientific Research Publishing

American Journal of Plant Sciences, 2023, 14, 1464-1479
https://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

# In-Silico Identification and Differential Analysis of Mitochondrial RNA Editing Events in *Helianthus* Genotypes/Species and Powdery Mildew Infected Variants

**Duruvasula Sree Lekha[1], Kandasamy Ulaganathan[1*], Mulpuri Sujatha[2*]**

[1]Centre for Plant Molecular Biology, Osmania University, Hyderabad, India
[2]ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad, India
Email: *kulaganathan123@gmail.com

**How to cite this paper:** Lekha, D.S., Ulaganathan, K. and Sujatha, M. (2023) In-Silico Identification and Differential Analysis of Mitochondrial RNA Editing Events in *Helianthus* Genotypes/Species and Powdery Mildew Infected Variants. *American Journal of Plant Sciences*, **14**, 1464-1479.
https://doi.org/10.4236/ajps.2023.1412099

**Received:** June 20, 2023
**Accepted:** December 26, 2023
**Published:** December 29, 2023



## Abstract

Sunflower is one of the most used commercial oilseed crops and suffers due to Powdery mildew. RNA sequence alteration occurs due to RNA editing which is a post transcriptional modification. It causes a deviation from the genomic DNA sequence resulting in RNA-DNA differences. Accurate study of RNA editing events in diverse species is possible by NGS based methods. Here, we performed RNA sequencing of 12 leaf transcriptomes, which include three genotypes of *Helianthus annuus* (2023B, TX16R and ID25), *H. debilis*, *H. niveus,* and *H. praecox* along with their respective powdery mildew pathogen infected variants and systematically analysed the mitochondrial RNA editing events using computational reference-based mapping approach. We discovered 687 editing sites, 220 editing events in the protein-coding regions, among all species and genotypes considered in this study. These included "C to U" and "U to C" RNA editing events. On further analysis, we observed that these editing events include 14 different types of amino acid changes that involve the creation of two stop codon events. The conserved editing sites identified were 247 accounting for ~36% of all the editing sites identified. This study provides a detailed picture of the Helianthus species' mitochondrial RNA editing status. We have identified and characterized for the first time, genotype-specific, species-specific, and stress-specific RNA editing events which may be useful as a potential source for stress-responsive studies in the future.

## Keywords

*Helianthus annuus*, RNA Editing, RNA-seq, *H. niveus*, *H. debilis*, *H. praecox,*

Biotic Stress, Powdery Mildew

## 1. Introduction

Sunflower (*Helianthus annuus* L.) is among the major commercial oil producing crops worldwide owing to its adaptability to varied climatic conditions and soils. It has a diploid genome and belongs to the family Asteraceae which is a large family of flowering plants. The crop is vulnerable to several biotic and abiotic stresses and outbreak of diseases such as powdery mildew affects crop growth and yield. *Golovinomyces orontii* is a fungal pathogen that is the causative agent for powdery mildew in sunflowers and is geographically distributed among all continents. A study on the impact of powdery mildew on sunflowers in India found that the disease's severity levels at 30% and 64% resulted in seed output losses of up to 20.5% and 52.6%, respectively [1]. Powdery mildew infection in sunflowers reduces photosynthesis, causes early senescence, and results in poor seed filling and causes stunting and about 81% reduction in yield [2].

Transcriptomic and proteomic diversity in eukaryotic organisms occurs due to post-transcriptional modifications of mRNA sequences [3] [4]. Plant mitochondria are known to exhibit diversity in terms of number of genes and structure. The plant mitochondrial genes range from 50 to 60 in number and have large intergenic regions. Specific post-transcriptional nucleotide modifications could be attributed to RNA editing, a characteristic of higher plant mitochondria. The RNA editing modifications include insertions, deletions, substitutions, and single base chemical modifications. Both protein-coding and non-protein-coding regions may be influenced by these changes [5]. Although, in plants, RNA editing primarily takes place in organelles, it can also occur in the nucleus and cytoplasm [6].

The occurrence of an RNA editing event is frequently due to a C to U conversion in both chloroplast and mitochondria of most land plants [7]. A reverse U to C type of editing is rarely observed and seems to be confined to ferns, lycophytes, and hornworts. However, a recent study in monilophytes suggests that "C to U" and "U to C" editing events in fern chloroplasts adhere to divergent evolutionary pathways as opposed to those that have previously been seen in flowering plants [8]. The RNA editing events occur in translated regions of organellar mRNA, untranslated regions, intervening sequences, and structural RNA [9]. The events in the coding regions modify the codons and help in producing functional proteins whereas the ones in non-coding regions are known to affect translation efficiency and splicing. Therefore, RNA editing is a supplementary proofreading mechanism and helps in the restoration of evolutionarily conserved amino acid residues. RNA editing sites are the locations of the specific RNA positions that are altered and the corresponding DNA locations. The editing events lead to the creation, deletion and substitution of start and stop codons. The RNA editing process could be a translational control process as trans-

lational initiation and translational termination codons can be introduced by RNA editing [10]. The RNA editing frequency ranges from 0 - 1000 sites across the plant kingdom [11]. The number of mitochondrial RNA editing sites ranges between 300 and 600 in flowering plants [12].

Though the first evidence of RNA editing in plant mitochondria of flowering plants was reported as early as 1989 [4] [13] [14] profiling RNA editing is made easy with the next generation sequencing technologies [15]. Previously, using NGS based methods, 401 RNA edit sites in *V. vinifera* [16], 491 edit sites in *O. sativa* [17], 357 edit sites in *B. vulgaris* [18], 1123 strand-specific mitochondrial RNA editing sites in *S. miltiorrhiza* were identified [19]. In this study, we performed transcriptome-wide identification of RNA editing events in both pathogens infected (*G. orontii*) genotypes/species and their respective controls of four *Helianthus* species (*H. annuus, H. debilis, H. praecox, H. niveus*) along with three genotypes of *H. annuus* (2023B, ID25, TX16R) using NGS based deep sequencing methods. Our work focuses on identification and differential mitochondrial RNA editing analysis in Sunflower species infected with Powdery mildew pathogen along with their respective negative controls. In this approach, we mapped RNA-seq reads to a known mitochondrial genome sequence to identify RNA editing sites. The complete cataloguing of RNA editing sites revealed eight unique pathogen stress-specific RNA editing sites. The RNA editing status among the species and genotypes may help in a greater understanding of the molecular mechanism(s) affecting processes like development, stress resistance and adaptive evolution.

## 2. Materials and Methods

### 2.1. Plant Material

The seeds of *Helianthus annuus* (2023B, TX16R, and ID25), *Helianthus niveus* (Accn No 1452), *Helianthus praecox*, and *Helianthus debilis* were soaked in water for a whole night. The seeds were decoated, placed in petri dishes and covered with moist filter paper. The seedlings, upon germination, were transferred to pots. At flowering stage, plants were transferred to the greenhouse (28˚C, 70% RH). Powdery mildew conidia were dusted on the leaves from the infected leaves of 2023B (Susceptible sunflower accession). Transcriptome profiling of infected leaves that were fixed at 24, 48, and 72 hours and control leaves (prior to infection) was performed.

### 2.2. Library Preparation and Sequencing

The lamina of both control and pathogen-treated plants (pooled samples of 24, 48, and 72 h post-infection) was collected and stored in "RNAlater" solution (Thermo Fisher Scientific) at −80˚C. The leaf samples were ground to a fine powder using liquid nitrogen in a mortar and pestle. RNA isolation was performed according to the defined kit protocol using RNeasy Plant kit. The RNA concentration and purity were evaluated using Nanodrop Spectrophotometer

(Thermo Scientific - 1000) and RNA integrity was analysed on a Bioanalyzer (Agilent, 2100). RNA samples with 7.9 and 8.2 RNA integrity numbers, respectively were used for library preparation. Library preparation was performed based on Illumina TruSeq RNA library protocol by Illumina Technologies (San Diego, CA). PolyA purification of mRNA was done using 1 μg of total RNA. Reverse transcription was carried out using SuperScriptIII Reverse transcriptase after fragmenting purified mRNA for 8 minutes at a temperature of 94˚C in the presence of divalent cations. The second strand cDNA was synthesized in the presence of DNA polymerase I and RNaseH. HighPrep PCR reagent (MAGBIO) was used for the cleaning of cDNA. Post end repair and an A base addition, Illumina adapters were ligated to the cDNA molecules and were followed by SPRI (solid-phase reversible immobilisation, Beckman Coulter) cleanup. The adapter ligated fragments were enriched by amplification of the library using 8 cycles of PCR. The library quality was verified on the High Sensitivity DNA Kit and quantified by using Qubit (Agilent). Illumina-HiSeq system (Illumina, San Deigo, CA) was used to carry out sequencing of ~80 million reads per sample. The estimated effective insert size estimated was ~130 - 380 bp.

### 2.3. Alignment of the RNA-seq Data

Initially, the transcriptomic reads of *H. annuus* (2023B, TX16R and ID25), *H. niveus*, *H. praecox*, and *H. debilis* along with their pathogen infected samples were aligned onto the indexed reference mitochondrial genome using bowtie2 (version 2.2.1) [20]. The reference mitochondrial genome of *H. annuus* (NC_023337.1) was downloaded from the NCBI (National Centre for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov). The output of the alignment was obtained in the Sequence alignment map (SAM) format. The SAM files obtained were converted into BAM (Binary alignment map) files using SAMtools (version 1.9) [21]. Further, the BAM files were sorted and used for further analysis.

### 2.4. Identification of RNA Editing Sites

RNA editing sites were identified using a pipeline comprising REDItools (version 1.2) [22]. The RNA editing events were identified by using REDItoolDenovo.py script using the parameters, coverage 5, frequency 0.10, and significant value 0.05 [23]. The -l parameter indicating "select significant sites" and -U parameter "use specific substitutions TC, CT" was considered to obtain edit sites.

### 2.5. Characterization of Differential RNA Editing Sites

The resultant REDItools files were used for comparison between each species/genotype separately. These sites were considered RNA editing sites (RES) for further studies. The annotation was done manually by downloading 26 mitochondrial genes from NCBI. The edit sites present in the coding region were filtered and the editing events that lead to synonymous and non-synonymous

amino acid changes were characterized. The intergenic regions were also categorized based on the genes in between they occur. In-silico validation of the RNA editing sites was performed using SNPEff (version 4.5) [24]. The edit sites that were common and unique between the pathogen-infected samples and their respective controls were enumerated in terms of different criteria such as those present and absent in coding and non-coding regions.

## 3. Results

### 3.1. Sequencing of Leaf-Specific Transcriptomes of Helianthus Species

RNA was isolated from the leaves of controls and pathogen-infected plants *i.e.*, genotypes of *H. annuus* (2023B, TX16R and ID25) and *H. debilis*, *H. niveus* and *H. praecox* and two independent libraries were prepared. The expected fragment distribution in the range of ~250 - 500 bp was observed in the cDNA libraries. Our sequencing resulted in 100 bp paired-end (PE) reads from the libraries using Illumina RNA-Seq technology with ~80 million reads per sample. The reads enumerated are recorded in Table 1.

### 3.2. Mapping of Helianthus Transcriptome Reads on to the Reference Mitochondrial Genome

A reference-based assembly approach was used in the identification of RNA editing sites in the mitochondrial transcripts of different genotypes of *Helianthus* species. The transcriptomic reads obtained were quality checked before they were used for Bowtie2 based assembly onto the reference mitochondrial genome of *H. annuus* (NCBI accession number: NC_023337.1). The results of the alignment ranged between 1,268,592 and 5,101,884 reads (Table 1). The average alignment percentage of both controls and infected samples is 3.38%.

Table 1. Summary of alignment percentages of Helianthus transcriptomes mapped to the mitochondrial genome of *H. annuus*.

| Name of the specie | Total number of reads | Total number of reads aligned to mitochondrial genome | Percentage of overall alignment |
|---|---|---|---|
| *H. annuus PS2023B control* | 96,122,236 | 1,268,592 | 1.31 |
| *H. annuus PS2023B pathogen infected sample* | 87,418,604 | 3,131,592 | 3.58 |
| *H. annuus ID25 control* | 82,174,332 | 2,740,596 | 3.33 |
| *H. annuus ID25 pathogen infected sample* | 91,162,936 | 4,173,904 | 4.57 |
| *H. annuus TX16R control* | 80,017,428 | 2,815,108 | 3.51 |
| *H. annuus TX16R pathogen infected sample* | 82,246,144 | 2,700,394 | 3.28 |
| *H. debilis control* | 87,529,646 | 3,338,856 | 3.81 |
| *H. debilis pathogen infected sample* | 82,408,066 | 3,114,186 | 3.77 |
| *H. niveus control* | 83,175,466 | 4,911,814 | 5.9 |
| *H. niveus pathogen infected sample* | 136,544,538 | 5,101,884 | 3.73 |
| *H. praecox control* | 89,159,016 | 2,218,732 | 2.48 |
| *H. praecox pathogen infected sample* | 89,434,750 | 1,175,668 | 1.31 |

### 3.3. Identification of RNA Editing Sites in Different Helianthus Species

We have considered sites with both "C to U" and "U to C" type of RNA editing. A total of 687 RNA edit sites were enumerated in all the species of *Helianthus* at a significance of 0.05. The overall localization of edit sites was lower in the protein-coding regions (220) in comparison to that of the non-coding regions (467). The average of mean frequency of edit sites of each gene in the coding region is 0.61. Here, the editing frequency refers to the number of reads edited/total number of reads covering that site.

Individually, 364 edit sites (*H. annuus* 2023B control), 390 edit sites (*H. annuus* 2023B pathogen infected), 375 edit sites (*H. annuus* ID25 control), 387 (*H. annuus* ID25 pathogen infected sample), 413 edit sites (*H. annuus TX16R* control), 407 edit sites (*H. annuus TX16R* pathogen infected), 370 edit sites (*H. debilis* control), 377 edit sites (*H. debilis* pathogen infected), 435 edit sites (*H. niveus* control), 436 edit sites (*H. niveus* pathogen infected sample), 371 edit sites (*H. praecox* control), 381 edit sites (*H. praecox* pathogen infected sample) (**Figure 1**) were identified in the *Helianthus* species and their respective pathogen infected samples. The average number of RNA editing sites observed in all species was 392.

Among all the 687 edit sites, 247 RNA editing events were commonly found in all the *Helianthus* species and their pathogen-infected samples depicting 36% of all the edit sites being conserved. We identified 467 RNA editing sites in the non protein coding regions accounting to ~67% of the total edit sites. There were no edit sites observed in the long non-coding RNAs. However, there were 4 edit sites in tRNAs and 15 in rRNAs and the rest were localized in the intergenic regions.

### 3.4. Characterization of RNA Editing Sites in Protein Coding Regions

We analysed the distribution of RNA editing sites in all the protein-coding regions. In total, 220 edit sites were observed in 15 protein-coding genes among all species and their respective pathogen-infected samples. Among all the species, 46 RNA edit sites that were enumerated in the *ccmB* gene were marked the highest accounting to 21% followed by 33 edit sites in *coxI* and 25 edit sites in *rps4* gene. Also, 19 and 18 RNA edit sites were observed in *nad3* and *cob* genes, respectively (**Figure 2**). Most RNA editing sites (53.63%) were at the second codon position (118) while 62 and 40 edit sites were found in the first and third codon positions respectively. Among all the editing events about 28.18%, 53.63% and 18.19% of editing changes occurred in first, second and third codon positions, respectively. This pattern was observed among all the genotypes of *H. annuus* and other *Helianthus* species.

RNA editing in the protein coding regions of *Helianthus* species resulted in 14 types of amino acid changes. The P > L amino acid transition occurred in most (43) of RNA editing sites. Second most prevalent amino acid transition was S > L followed by S > F observed in 39 and 31 edit sites, respectively (**Figure 3**). The

conversion of Serine to Leucine is a shift from hydrophilic to hydrophobic. Additionally, two creation of a stop codon events were observed.

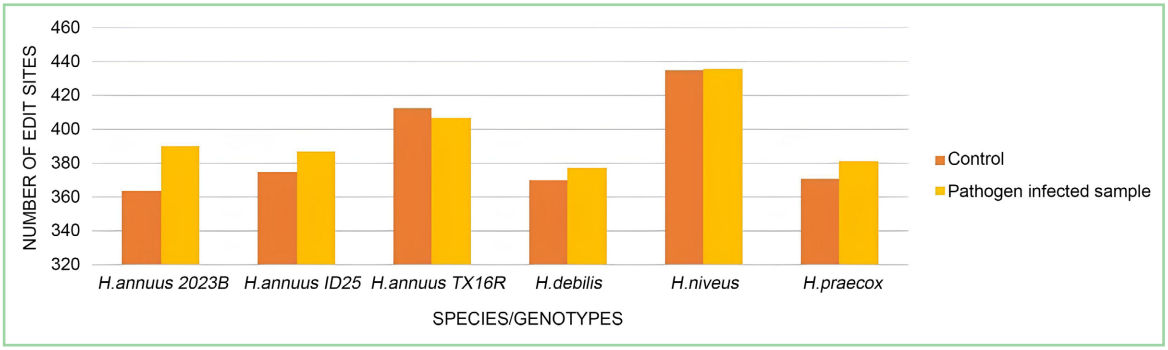


**Figure 1.** Distribution of RNA editing sites among all Helianthus Genotypes/Species.

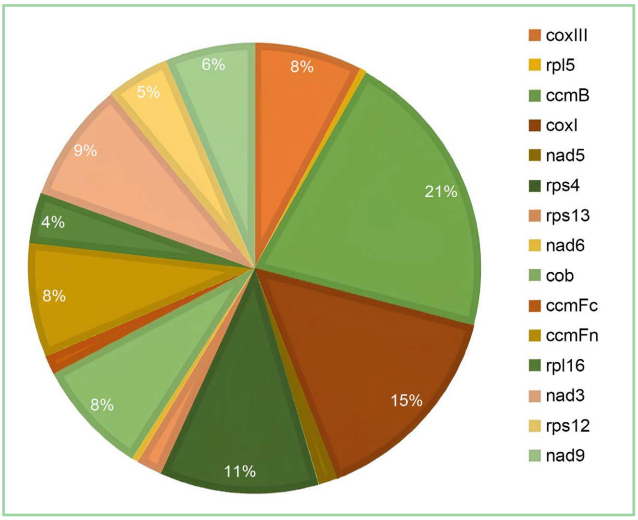


**Figure 2.** Distribution of RNA editing sites in the protein-coding regions of all genotypes/species of *Helianthus*.

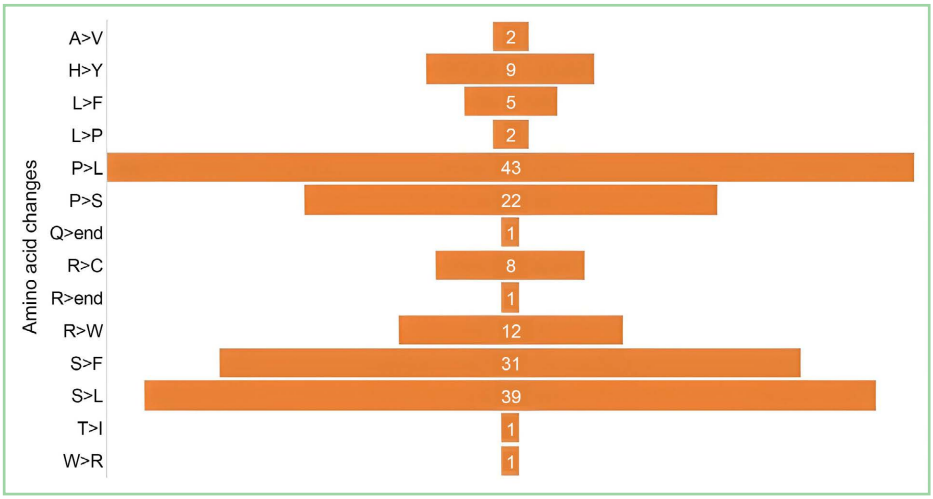


**Figure 3.** Distribution of Amino acid changes due to RNA editing events in Helianthus species/ genotypes.

The editing events that lead to the creation of stop codons are due to a C to U conversions. An editing event in the gene *rpl16-37* occurs in the first codon position (Cag > Tag). It occurred due to the conversion of the amino acid Glutamine to stop codon. It is conserved among all species/genotypes. The other editing event *rps4-991* also occurs in the first codon position (Cga > Tga) which lead to the conversion of Arginine to stop codon. This editing event is not conserved. The average editing levels in all the genes of the coding region were calculated based on the editing frequency at each site. In each genotype/specie, the average editing level is the mean value of the editing frequency of all sites present in a gene. Overall, the highest editing levels were observed in the *coxI* gene ranging between 0.975 in *H. annuus* ID25 pathogen-infected sample and 0.86 in *H. niveus* pathogen-infected sample (**Figure 4**). While lowest editing levels were observed in the *rpl5* gene (0.11) in both *H. debilis* control.

In our analysis 85% are these editing events are "C to U" type and the rest are reverse editing ("U to C") events. Among those protein-coding genes, *ccmB* has the most editing sites predicted (29) which agrees with our results of the gene with the highest number of edit sites (46 edit sites). Also, 33 edit sites were observed in *cox1* gene.

### 3.5. Genotype-Specific Differential RNA Editing Events in *H. annuus*

Not only species-specific variation in the number of editing sites but also genotype-specific variation was observed when three genotypes of *H. annuus* were analysed for mitochondrial genotype-specific RNA editing. The highest number of edit sites was found in *H. annuus* TX16R control (413) (**Table 2**). The average number of edit sites among the three genotypes and their respective pathogen-infected samples of *H. annuus* was found to be 389. The higher number of edit sites was observed in the pathogen-infected samples in comparison to the controls except in the case of *H. annuus* TX16R, where 413 edit sites were observed in the control in comparison to 407 in the pathogen-infected sample. The number of edit sites in the protein-coding regions was lower in all the genotypes in comparison to those in the non-coding regions. Also, the amino acid changes that resulted due to nucleotide conversions were mostly non-synonymous substitutions with their average percentage accounting for ~86.77% of edit sites in protein-coding regions. There is an increase in the number of edit sites in the non-coding regions of pathogen-infected samples in comparison to their respective controls. However, this pattern was not observed in the case of *H. annuus* TX16R genotype. There is a decrease in the non-synonymous substitutions and increase in the synonymous substitutions in the pathogen-infected samples w.r.t their controls in all the genotypes except in the case of H. *annuus* TX16R where it is reversed. In the case of *H. annuus* TX16R pathogen-infected sample, 153 edit sites showed non-synonymous substitutions in comparison to 149 in its control (**Table 2**). Similar pattern can be observed in the case of unique sites as well.

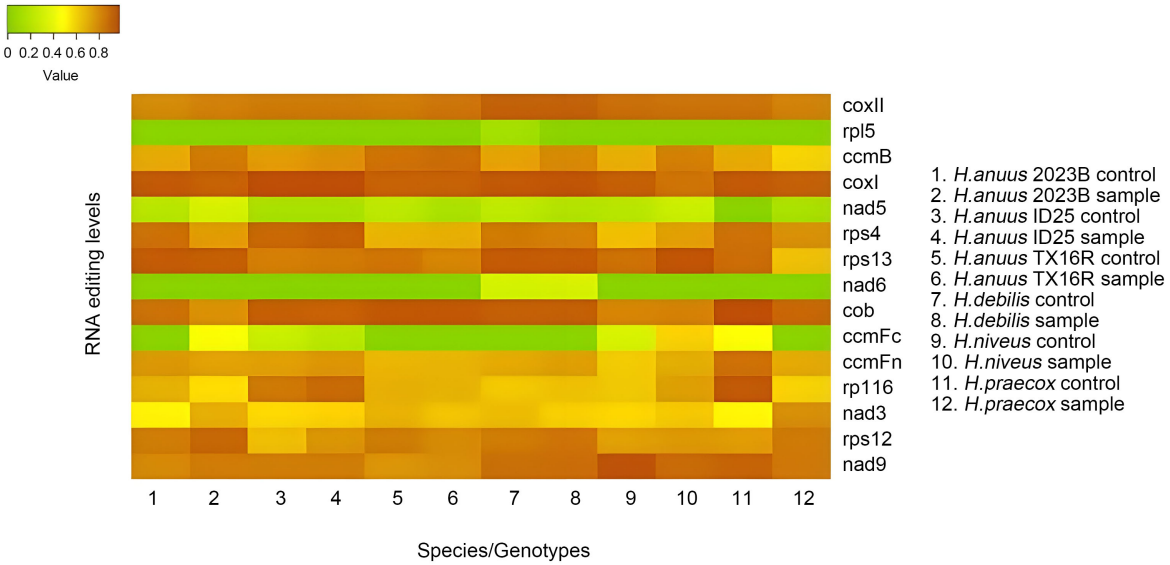


**Figure 4.** Distribution of RNA editing levels in the protein-coding regions of all Helianthus species.

**Table 2.** Summary of characterisation of RNA editing events in *Helianthus annuus genotypes*.

| Location/type | *H. annuus* 2023B | | Common sites in 2023B | Unique sites in 2023B | | *H. annuus* ID25 | | Common sites in ID25 | Unique sites in ID25 | | *H. annuus* TX16R | | Common sites in TX16R | Unique sites in TX16R | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | I | | C | I | C | I | | C | I | C | I | | C | I |
| Protein coding transcripts | 170 | 170 | 158 | 12 | 12 | 168 | 168 | 166 | 2 | 2 | 175 | 177 | 169 | 6 | 8 |
| First codon position | 53 | 52 | 50 | 3 | 2 | 53 | 52 | 52 | 1 | 0 | 50 | 52 | 50 | 0 | 2 |
| Second codon position | 97 | 94 | 90 | 7 | 4 | 97 | 96 | 96 | 1 | 0 | 99 | 101 | 96 | 3 | 5 |
| Third codon position | 20 | 24 | 18 | 2 | 6 | 18 | 20 | 18 | 0 | 2 | 26 | 24 | 23 | 3 | 1 |
| Creation of start codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Substitution of start codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Creation of stop codon | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| Substitution of a stop codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-synonymous substitutions | 149 | 145 | 139 | 10 | 6 | 149 | 147 | 147 | 2 | 0 | 149 | 153 | 146 | 3 | 7 |
| Synonymous substitutions | 21 | 25 | 19 | 2 | 6 | 19 | 21 | 19 | 0 | 2 | 26 | 24 | 23 | 3 | 1 |
| Non-protein coding transcripts | 194 | 220 | 176 | 18 | 44 | 207 | 219 | 192 | 15 | 27 | 238 | 230 | 207 | 31 | 23 |
| tRNA | 3 | 3 | 3 | 0 | 0 | 3 | 3 | 3 | 0 | 0 | 3 | 3 | 3 | 0 | 0 |
| rRNA | 3 | 3 | 2 | 1 | 1 | 3 | 2 | 2 | 1 | 0 | 5 | 3 | 2 | 3 | 1 |
| Long noncoding RNA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intergenic region | 188 | 214 | 171 | 17 | 43 | 201 | 214 | 187 | 14 | 27 | 230 | 224 | 202 | 28 | 22 |
| Total no of sites | 364 | 390 | 334 | 30 | 56 | 375 | 387 | 358 | 17 | 29 | 413 | 407 | 376 | 37 | 31 |

### 3.6. Species-Specific RNA Editing Events and Their Differential Status among Helianthus Species and Their Respective Pathogen Infected Samples

The highest number of edit sites were identified in *H. niveus* pathogen infected sample (436). The average number of edit sites found in the protein coding regions was enumerated as 174 and is marked higher than that recorded in *H. annuus* genotypes. In all the species, the number of edit sites was higher in the non-protein coding regions. Also, the editing events that occur due to substitutions in the second codon position were marked highest among all the species like that of *H. annuus* genotypes. In the protein-coding regions, the non-synonymous amino acid substitutions were clearly higher (87.57%) than the synonymous substitutions (Table 3). The other gain of stop codon event which was not conserved was prevalent in *H. annuus* 2023B pathogen-infected sample, *H. niveus* control and *H. niveus* sample.

Table 3. Summary of characterised mitochondrial RNA editing sites in *H. debilis*, *H. niveus* and *H. praecox*.

| Location/type | *H. debilis* | | Common sites in *H. debilis* | Unique site in *H. debilis* | | *H. niveus* | | Common sites in *H. niveus* | Unique site in *H. niveus* | | *H. praecox* | | Common sites in *H. praecox* | Unique sites in *H. praecox* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | I | | C | I | C | I | | C | I | C | I | | C | I |
| Protein coding transcripts | 175 | 175 | 173 | 2 | 2 | 173 | 174 | 160 | 13 | 14 | 170 | 179 | 167 | 3 | 12 |
| First codon position | 54 | 54 | 54 | 0 | 0 | 56 | 55 | 52 | 4 | 3 | 55 | 56 | 55 | 0 | 1 |
| Second codon position | 102 | 100 | 100 | 2 | 0 | 99 | 93 | 90 | 9 | 3 | 98 | 99 | 96 | 2 | 3 |
| Third codon position | 19 | 21 | 19 | 0 | 2 | 18 | 26 | 18 | 0 | 8 | 17 | 24 | 16 | 1 | 8 |
| Creation of start codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Substitution of start codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Creation of stop codon | 1 | 1 | 1 | 0 | 0 | 2 | 2 | 2 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| Substitution of a stop codon | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-synonymous substitutions | 156 | 154 | 154 | 2 | 0 | 154 | 147 | 141 | 13 | 6 | 152 | 153 | 150 | 2 | 3 |
| Synonymous substitutions | 19 | 21 | 19 | 0 | 2 | 19 | 27 | 19 | 0 | 8 | 18 | 26 | 17 | 1 | 9 |
| Non-protein coding transcripts | 195 | 202 | 185 | 10 | 17 | 262 | 262 | 203 | 59 | 59 | 201 | 202 | 169 | 32 | 33 |
| tRNA | 3 | 3 | 3 | 0 | 0 | 3 | 3 | 2 | 1 | 1 | 3 | 3 | 3 | 0 | 0 |
| rRNA | 1 | 0 | 0 | 0 | 0 | 3 | 13 | 3 | 0 | 10 | 3 | 3 | 2 | 1 | 1 |
| Long noncoding RNA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Intergenic region | 191 | 199 | 182 | 9 | 17 | 256 | 246 | 198 | 58 | 48 | 195 | 196 | 164 | 31 | 32 |
| Total no of sites | 370 | 377 | 358 | 12 | 19 | 435 | 436 | 363 | 72 | 73 | 371 | 381 | 336 | 35 | 45 |

In case of the total number of edit sites enumerated, highest number of edit sites were observed in pathogen-infected sample of *H. niveus*. In all the species there is an increase in the number of edit sites in the pathogen-infected samples in comparison to their controls. An increased number of RNA editing sites were observed in the non-coding regions of pathogen-infected samples in comparison to their respective controls except in case of *H. niveus* where equal number of edit sites were observed (262). The edit sites found among all the genotypes/species were considered as conserved. In the protein coding regions 131 out of 247 edit sites were conserved accounting to 53%. Among them, highest number of edit sites were observed in the gene *ccmB* (41) followed by 19 in *nad3*. The conserved edit sites were completely absent in the *rpl5*, *nad5*, *nad6* and *ccmC* genes.

Individually, common edit sites between pathogen-infected samples and their respective controls were enumerated and it was observed that *H. annuus* TX16R showed the highest number of edit sites with 376 common edit sites. While a comparison between other three species showed that *H. niveus* (363) showed the highest number of common edit sites. The number of common edit sites in the protein coding regions of *H. debilis* (173) was higher in *H. debilis* despite the lesser number of common edit sites (358). Unique edit sites were also recorded. The total number of unique edit sites identified was 157 and 15 were observed in the coding regions (**Figure 5**). The unique edit sites were completely absent in *H. annuus* 2023B control, its pathogen infected sample, *H. annuus* ID25 control and *H. debilis* sample.

## 4. Discussion

The extreme RNA and protein diversity in eukaryotes maybe attributed to modifications like splicing / alternate splicing, RNA editing by which nucleotides are inserted, deleted, or substituted resulting in RNA-DNA differences [25] [26] [27]. The type of base transition that occurs most commonly due to RNA editing in most land plants is a C to U RNA editing event. The reverse U to C editing event is considered relatively occasional but is abundantly found in ferns, hornworts, and lycophytes in flowering plants RNA editing analysis of organellar transcriptomes show abundant C to U type of editing [8]. Therefore, we considered both types of editing events in our study. Plants show wide range of variation in number of mitochondrial RNA editing sites. A total of 569 C-to-U editing sites in the mitochondria-encoded open reading frames (ORFs) of Rice [28]. The very high number of mitochondrial RNA edit sites were found in plants like lycophyte *Isoetes engelmannii* (1782), *Cycas taitungensis* (1084) and *Liriodendron* (755) [29] [30] [31]. Whereas transcriptomic studies in *Physcomitrella patens* revealed only 11 sites in nine mitochondrial genes although RNA editing follows similar patterns as other land plants [32]. However, total number of edit sites identified among *Helianthus* species considered in this study agrees with earlier investigations that suggest 300 - 600 editing events occur in plant mitochondria [18]. In our analysis, the average number of RNA editing sites observed in all species was 392.

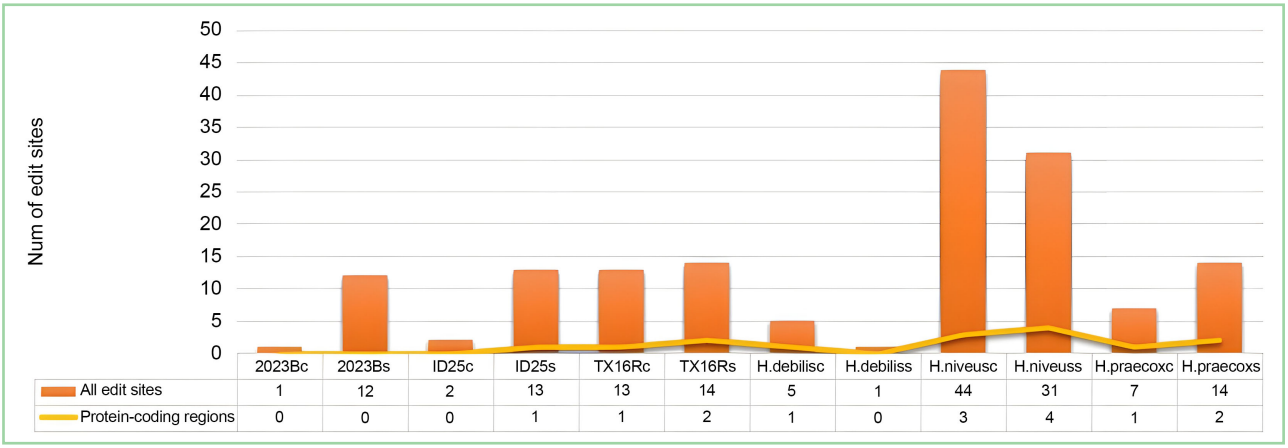


**Figure 5.** Unique RNA editing sites in all the species/genotypes.

The pattern of editing changes in first, second and third codon positions were 28.18%, 53.63% and 18.19% respectively among all the genotypes of *H. annuus* and other *Helianthus* species. It agrees with most of the studies that show maximum editing events due to changes in the second codon position [33]. A study on RNA editing status in 17 angiosperm genera showed the lowest conservation of edit sites in the third codon position [34]. The most frequent type of alterations that occur due to RNA editing in mitochondria is C to U changes [35] and in our analysis 85% are these editing events and the rest are reverse editing (U to C) events. In Arabidopsis the distribution of C to U editing events is 456 in mRNA, 441 in ORFs, 8 in introns [36], 401 in the coding region and 44 in the non-coding regions of Grape [16]. According to recent studies in the mitochondrial genome of *S. glauca,* 216 RNA editing sites in 26 protein-coding genes were found. Among those protein-coding genes, *ccmB* has the most editing sites predicted (29) which agrees with our results of the gene with the highest number of edit sites (46 edit sites). However, cox*1* lacks editing sites in *S. glauca* [37] in contrary to 33 edit sites observed in our analysis.

RNA editing plays an important role in essential physiological processes and affects stress responses. Plant growth, development, and fertility are adversely impacted by altering some editing sites, suggesting the significance of RNA editing in plant organellar gene expression [38]. According to a study, SLG1 which is a PPR protein localized in the mitochondria of Arabidopsis affects RNA editing and the SLG1 mutants exhibit impairment of NADH dehydrogenase activity as the mitochondrial transcript nad3 is abolished [39]. According to a study in Arabidopsis thaliana, the "C to U" RNA editing rates were significantly reduced under heat and cold stress suggesting the role of RNA editing in stress response [40].

In this study, we observed 85 edit sites that are specific to pathogenic stress-infected variants and these could be called stress-specific edit sites. It was noted that the highest number of edit sites were observed in *H. niveus* sample (31). The unique edit sites of pathogen-infected samples in the coding region are

*ccmFc-114* in *H. annuus ID25 sample* (Ala > Ala, gcC > gcT and 3rd codon position); *coxI-1148* (Ser > Phe, tCc > tTc and 2nd codon position) and *ccmFn-264* (Pro > Pro, ccC > ccT and 3rd codon position) in *H. annuus* TX16R pathogen infected sample; *coxI-1524* (Phe > Phe, ttC > ttT and 3rd codon position), *ccmFn-1104* (Ala > Val, gCt > gTc and 2nd codon position), *ccmFn-1107* (Leu > Leu, ctC > ctT and 3rd codon position) and *rps12-275* (Pro > Ser, Ccg > Tcg and 1st codon position) in *H. niveus* sample; *coxIII-593* (Ser > Phe, tCt > tTt and 2nd codon position) and *rps4-772* (Leu > Leu, Cta > Tta and 1st codon position) in *H. praecox* sample. The role of these edit sites in stress response must be deeply studied further.

According to a species-wide comparative study of four Populus species, ~69% (238 out of 355 edit sites) of mitochondrial RNA editing sites are conserved [41]. Though our analysis showed an overall ~36% of conserved sites, ~53% of the edit sites in the coding region are conserved. The basis for variations in the levels of editing event conservation among different land plants must be studied as there are very limited species-wide comparative studies.

## 5. Conclusion

Our study is the first comprehensive analysis of mitochondrial RNA editing sites in Helianthus sp. using deep transcriptome sequencing. We have only considered C to U and U to C type of editing events. RNA editing events that arise due to other base substitutions should be studied further. Also, RNA editing in the intergenic regions was recorded. The genotype, species-specific and stress-specific RNA editing events could be used as potential sources of information for genetic manipulation in sunflowers.

## Statements and Declarations

### Acknowledgements

This work is supported by the Department of Science and Technology, New Delhi, India. DS has received DST-INSPIRE fellowship. MS has received grant for the project (SERB/SR/SO/PS/24/2012).

### Author Contributions

DS was involved in RNA editing analysis and draft manuscript preparation; MS was involved in conceiving the idea, work plan, RNA isolation, RNA sequencing and transcriptome analysis; KU was involved in work plan, bioinformatic analysis, data interpretation and manuscript preparation. All authors read and approved the final manuscript.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References


[1] Dudhe, M.Y., Meena, H.P., Sujatha, M., Varaprasad, K.S. and Ranganatha, A.R.G. (2015) Sunflower Genetic Resources Management at DOR.

[2] Carter, J.F. (1978) Sunflower Science and Technology. American Society of Agronomy, Madison. https://doi.org/10.2134/agronmonogr19

[3] Gott, J.M. (2003) Expanding Genome Capacity via RNA Editing. *Comptes Rendus Biologies*, **326**, 901-908. https://doi.org/10.1016/j.crvi.2003.09.004

[4] Hiesel, R., Wissinger, B., Schuster, W. and Brennicke, A. (1989) RNA Editing in Plant Mitochondria. *Science*, **246**, 1632-1634. https://doi.org/10.1126/science.2480644

[5] Steinhauser, S., Beckert, S., Capesius, I., Malek, O. and Knoop, V. (1999) Plant Mitochondrial RNA Editing. *Journal of Molecular Evolution*, **48**, 303-312. https://doi.org/10.1007/PL00006473

[6] Bowe, L.M. and de Pamphilis, C.W. (1996) Effects of RNA Editing and Gene Processing on Phylogenetic Reconstruction. *Molecular Biology and Evolution*, **13**, 1159-1166. https://doi.org/10.1093/oxfordjournals.molbev.a025680

[7] Chateigner-Boutin, A.L. and Small, I. (2007) A Rapid High-Throughput Method for the Detection and Quantification of RNA Editing Based on High-Resolution Melting of Amplicons. *Nucleic Acids Research*, **35**, e114. https://doi.org/10.1093/nar/gkm640

[8] Knie, N., Grewe, F., Fischer, S. and Knoop, V. (2016) Reverse U-to-C Editing Exceeds C-to-U RNA Editing in Some Ferns—A Monilophyte-Wide Comparison of Chloroplast and Mitochondrial RNA Editing Suggests Independent Evolution of the Two Processes in Both Organelles. *BMC Evolutionary Biology*, **16**, Article No. 134. https://doi.org/10.1186/s12862-016-0707-z

[9] Börner, G.V., Mörl, M., Wissinger, B., Brennicke, A. and Schmelzer, C. (1995) RNA Editing of a Group II Intron in *Oenothera* as a Prerequisite for Splicing. *Molecular and General Genetics MGG*, **246**, 739-744. https://doi.org/10.1007/BF00290721

[10] Hoch, B., Maier, R.M., Appel, K., Igloi, G.L. and Kössel, H. (1991) Editing of a Chloroplast mRNA by Creation of an Initiation Codon. *Nature*, **353**, 178-180. https://doi.org/10.1038/353178a0

[11] Sun, Y., Li, X., Wu, D., Pan, Q., Ji, Y., Ren, H. and Ding, K. (2016) RED: A Java-MySQL Software for Identifying and Visualizing RNA Editing Sites Using Rule-Based and Statistical Filters. *PLOS ONE*, **11**, e0150465. https://doi.org/10.1371/journal.pone.0150465

[12] Liu, Y.J., Xiu, Z.H., Meeley, R. and Tan, B.C. (2013) Empty Pericarp5 Encodes a Pentatricopeptide Repeat Protein That Is Required for Mitochondrial RNA Editing and Seed Development in Maize. *The Plant Cell*, **25**, 868-883. https://doi.org/10.1105/tpc.112.106781

[13] Covello, P.S. and Gray, M.W. (1989) RNA Editing in Plant Mitochondria. *Nature*, **341**, 662-666. https://doi.org/10.1038/341662a0

[14] Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H. and Grienenberger, J.M. (1989) RNA Editing in Wheat Mitochondria Results in the Conservation of Protein Sequences. *Nature*, **341**, 660-662. https://doi.org/10.1038/341660a0

[15] Giudice, C.L., Hernández, I., Ceci, L.R., Pesole, G. and Picardi, E. (2019) RNA Editing in Plants: A Comprehensive Survey of Bioinformatics Tools and Databases. *Plant Physiology and Biochemistry*, **137**, 53-61. https://doi.org/10.1016/j.plaphy.2019.02.001

[16] Picardi, E., Horner, D.S., Chiara, M., Schiavon, R., Valle, G. and Pesole, G. (2010) Large-Scale Detection and Analysis of RNA Editing in Grape mtDNA by RNA Deep-Sequencing. *Nucleic Acids Research*, **38**, 4755-4767. https://doi.org/10.1093/nar/gkq202

[17] Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Kadowaki, K., *et al.* (2002) The Complete Sequence of the Rice (*Oryza sativa* L.) Mitochondrial Genome: Frequent DNA Sequence Acquisition and Loss during the Evolution of Flowering Plants. *Molecular Genetics and Genomics*, **268**, 434-445. https://doi.org/10.1007/s00438-002-0767-1

[18] Mower, J.P. and Palmer, J.D. (2006) Patterns of Partial RNA Editing in Mitochondrial Genes of $\beta$ Vulgaris. *Molecular Genetics and Genomics*, **276**, 285-293. https://doi.org/10.1007/s00438-006-0139-3

[19] Wu, B., Chen, H., Shao, J., Zhang, H., Wu, K. and Liu, C. (2017) Identification of Symmetrical RNA Editing Events in the Mitochondria of *Salvia miltiorrhiza* by Strand-Specific RNA Sequencing. *Scientific Reports*, **7**, Article No. 42250. https://doi.org/10.1038/srep42250

[20] Langmead, B. and Salzberg, S.L. (2012) Fast Gapped-Read Alignment with Bowtie 2. *Nature Research*, **9**, 357-359. https://doi.org/10.1038/nmeth.1923

[21] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map Format and SAMtools. *Bioinformatics*, **25**, 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

[22] Lo Giudice, C., Tangaro, M.A., Pesole, G. and Picardi, E. (2020) Investigating RNA Editing in Deep Transcriptome Datasets with REDItools and REDIportal. *Nature Protocols*, **15**, 1098-1131. https://doi.org/10.1038/s41596-019-0279-7

[23] Picardi, E. and Pesole, G. (2013) REDItools: High-Throughput RNA Editing Detection Made Easy. *Bioinformatics*, **29**, 1813-1814. https://doi.org/10.1093/bioinformatics/btt287

[24] Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Ruden, D.M., *et al.* (2012) A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila melanogaster* Strain w1118; iso-2; iso-3. *Fly*, **6**, 80-92. https://doi.org/10.4161/fly.19695

[25] Benne, R., van den Burg, J., Brakenhoff, J.P., Sloof, P., van Boom, J.H. and Tromp, M.C. (1986) Major Transcript of the Frameshifted coxII Gene from Trypanosoma Mitochondria Contains Four Nucleotides That Are Not Encoded in the DNA. *Cell*, **46**, 819-826. https://doi.org/10.1016/0092-8674(86)90063-2

[26] Knoop, V. (2011) When You Can't Trust the DNA: RNA Editing Changes Transcript Sequences. *Cellular and Molecular Life Sciences*, **68**, 567-586. https://doi.org/10.1007/s00018-010-0538-9

[27] Farajollahi, S. and Maas, S. (2010) Molecular Diversity through RNA Editing: A Balancing Act. *Trends in Genetics*, **26**, 221-230. https://doi.org/10.1016/j.tig.2010.02.001

[28] Zheng, P., Wang, D., Huang, Y., Chen, H., Du, H. and Tu, J. (2020) Detection and Analysis of C-to-U RNA Editing in Rice Mitochondria-Encoded ORFs. *Plants*, **9**, Article 1277. https://doi.org/10.3390/plants9101277

[29] Grewe, F., Herres, S., Viehöver, P., Polsakiewicz, M., Weisshaar, B. and Knoop, V. (2011) A Unique Transcriptome: 1782 Positions of RNA Editing Alter 1406 Codon Identities in Mitochondrial mRNAs of the Lycophyte *Isoetes engelmannii*. *Nucleic Acids Research*, **39**, 2890-2902. https://doi.org/10.1093/nar/gkq1227

[30] Chaw, S.M., Chun-Chieh Shih, A., Wang, D., Wu, Y.W., Liu, S.M. and Chou, T.Y. (2008) The Mitochondrial Genome of the Gymnosperm *Cycas taitungensis* Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites. *Molecular Biology and Evolution*, **25**, 603-615. https://doi.org/10.1093/molbev/msn009

[31] Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J. and Palmer, J.D. (2013) The "Fossilized" Mitochondrial Genome of *Liriodendron tulipifera*: Ancestral Gene Content and Order, Ancestral Editing Sites, and Extraordinarily Low Mutation Rate. *BMC Biology*, **11**, Article No. 29. https://doi.org/10.1186/1741-7007-11-29

[32] Rüdinger, M., Funk, H.T., Rensing, S.A., Maier, U.G. and Knoop, V. (2009) RNA Editing: Only Eleven Sites Are Present in the *Physcomitrella patens* Mitochondrial Transcriptome and a Universal Nomenclature Proposal. *Molecular Genetics and Genomics*, **281**, 473-481. https://doi.org/10.1007/s00438-009-0424-z

[33] Mulligan, R.M., Chang, K.L. and Chou, C.C. (2007) Computational Analysis of RNA Editing Sites in Plant Mitochondrial Genomes Reveals Similar Information Content and a Sporadic Distribution of Editing Sites. *Molecular Biology and Evolution*, **24**, 1971-1981. https://doi.org/10.1093/molbev/msm125

[34] Edera, A.A., Gandini, C.L. and Sanchez-Puerta, M.V. (2018) Towards a Comprehensive Picture of C-to-U RNA Editing Sites in Angiosperm Mitochondria. *Plant Molecular Biology*, **97**, 215-231. https://doi.org/10.1007/s11103-018-0734-9

[35] Takenaka, M., Verbitskiy, D., Zehrmann, A., Härtel, B., Bayer-Csaszar, E., Glass, F. and Brennicke, A. (2014) RNA Editing in Plant Mitochondria—Connecting RNA Target Sequences and Acting Proteins. *Mitochondrion*, **19**, 191-197. https://doi.org/10.1016/j.mito.2014.04.005

[36] Giegé, P. and Brennicke, A. (1999) RNA Editing in Arabidopsis Mitochondria Effects 441 C to U Changes in ORFs. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 15324-15329. https://doi.org/10.1073/pnas.96.26.15324

[37] Cheng, Y., He, X., Priyadarshani, S.V.G.N., Wang, Y., Ye, L., Shi, C., Qin, Y., *et al.* (2021) Assembly and Comparative Analysis of the Complete Mitochondrial Genome of *Suaeda glauca*. *BMC Genomics*, **22**, Article No. 167. https://doi.org/10.1186/s12864-021-07490-9

[38] Hammani, K. and Giegé, P. (2014) RNA Metabolism in Plant Mitochondria. *Trends in Plant Science*, **19**, 380-389. https://doi.org/10.1016/j.tplants.2013.12.008

[39] Yuan, H. and Liu, D. (2014) Genetic Evidence That Arabidopsis ALTERED ROOT ARCHITECTURE Encodes a Putative Dehydrogenase Involved in Homoserine Biosynthesis. *Plant Cell Reports*, **33**, 25-33. https://doi.org/10.1007/s00299-013-1509-z

[40] Chu, D. and Wei, L. (2019) The Chloroplast and Mitochondrial C-to-U RNA Editing in Arabidopsis Thaliana Shows Signals of Adaptation. *Plant Direct*, **3**, e00169. https://doi.org/10.1002/pld3.169

[41] Brenner, W.G., Mader, M., Müller, N.A., Hoenicka, H., Schroeder, H., Zorn, I., Kersten, B., *et al.* (2019) High Level of Conservation of Mitochondrial RNA Editing Sites among Four *Populus* Species. *G3*: *Genes*, *Genomes*, *Genetics*, **9**, 709-717. https://doi.org/10.1534/g3.118.200763