Chloroplast C-to-U RNA editing in vascular plants is adaptive due to its restorative effect: testing the restorative hypothesis

YUANGE DUAN, WANZHI CAI, and HU LI

Key Laboratory of Pest Monitoring and Green Management, Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193, China

ABSTRACT

The adaptiveness of nonsynonymous RNA editing (recoding) could be conferred by the flexibility of the temporal-spatially controllable proteomic diversity, or by its restorative effect which fixes unfavorable genomic mutations at the RNA level. These two complementary hypotheses, namely, the diversifying hypothesis and the restorative hypothesis, have distinct predictions on the landscape of RNA editing sites. We collected the chloroplast C-to-U RNA editomes of 21 vascular plants (11 angiosperms, four gymnosperms, and six ferns) from a previous study, aiming to testify whether the plant editomes typically conform to the restorative hypothesis. All predictions made by the restorative hypothesis are verified: (i) nonsynonymous editing sites are more frequent and have higher editing levels than synonymous sites; (ii) nonsynonymous editing levels are extremely high and show weak tissue-specificity in plants; (iii) on the inferred genomic sites with recent T-to-C mutations, nonsynonymous sites but not synonymous sites are compensated by C-to-U RNA editing. In conclusion, nonsynonymous C-to-U RNA editing in plants is adaptive due to its restorative effects. The recoding levels are high and are constantly required across the whole plant so that the recoding events could perfectly mimic DNA mutations. The evolutionary significance of plant RNA editing is systematically demonstrated at the genome-wide level.

Keywords: C-to-U RNA editing; vascular plants; recoding; adaptive; restorative hypothesis

INTRODUCTION

RNA editing is prevalent in all domains of lives. In animals, adenosine-to-inosine (A-to-I) RNA editing catalyzed by ADARs is the most abundant editing type (Palladino et al. 2000; Savva et al. 2012), while in plants, cytidine-to-uridine (C-to-U) RNA editing mediated by PPR proteins (Fig. 1A) occurs in transcripts of non-nuclear genes (Gray and Covello 1993; Yin et al. 2013). Since I is read as G and U is similar to T, both A-to-I(G) or C-to-U(T) editing types are able to cause amino acid substitutions, termed nonsynonymous editing (recoding) (Fig. 1B). The biological significance of nonsynonymous editing has been interpreted by two complementary hypotheses (Duan et al. 2022): the diversifying hypothesis (Gommans et al. 2009) and restorative hypothesis (Jiang and Zhang 2019).

The diversifying hypothesis considers that RNA editing increases the proteomic diversity in a temporal-spatial manner (Gommans et al. 2009). The flexibility of RNA editing is advantageous compared to the pleiotropic effect of DNA mutations. For example, if a DNA mutation is beneficial to adult flies but lethal to larva, then this mutation would be eliminated due to its antagonistic effect. However, for RNA editing, one could selectively recode the transcripts in adult flies while keeping the larva genes unedited (Graveley et al. 2011). This flexibility circumvents the pleiotropic effect of DNA mutations. The controllable manner of RNA editing is adaptive. The diversifying hypothesis also requires tissue- or developmental stage-specific RNA editing (Graveley et al. 2011) although the global editing level is not necessarily very high. Nevertheless, nonsynonymous editing sites should have higher editing levels than the neutrally evolving synonymous sites (Duan et al. 2017, 2018; Yablonovitch et al. 2017).

Corresponding author: duanyuange@cau.edu.cn

Abbreviations: C-to-U, cytidine-to-uridine; A-to-I, adenosine-to-inosine; PPR, pentatricopeptide repeat; ADAR, adenosine deaminase acting on RNA; CDS, coding sequence; Nonsyn, nonsynonymous; Syn, synonymous; SNP, single-nucleotide polymorphism

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079450.122.

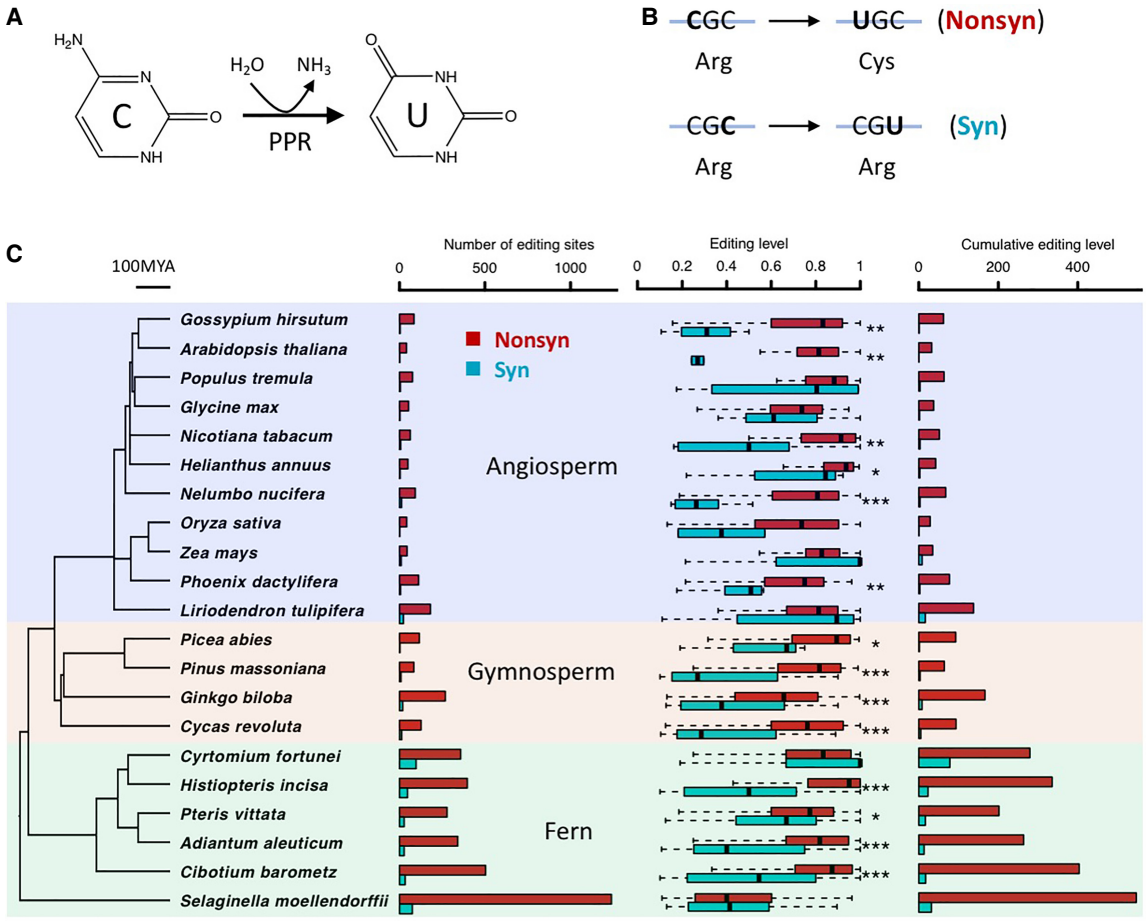


FIGURE 1. C-to-U RNA editing in vascular plants. (*A*) C-to-U RNA editing. (*B*) C-to-U editing could lead to nonsynonymous or synonymous mutations. (*C*) The phylogeny and C-to-U RNA editing landscape of 21 vascular plants. Eleven angiosperms, four gymnosperms, and six ferns were collected from a previous study. The numbers of editing sites, distribution of editing levels, and cumulative editing levels are shown next to each species. Nonsynonymous sites and synonymous sites are displayed separately, and their editing levels are compared to each other by Wilcoxon rank sum tests. (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$.

The restorative hypothesis proposes that RNA editing reverses deleterious DNA mutations and restores the ancestral allele (Jiang and Zhang 2019). Although the original study explained that the edited allele (RNA) at present is no fitter than the ancestral DNA allele (Jiang and Zhang 2019), it is undeniable that in the present species the edited allele is better than the unedited allele so that editing confers adaptiveness by restoration. The recoding event (the editing mechanism) itself is apparently adaptive as it serves as a compensation for unfavorable DNA mutations, but we also respect the interpretation that restorative editing is nonadaptive (by comparing the edited allele versus the ancestral allele). Under the restorative hypothesis, it is conceivable that the whole body should be recoded without tissue-specificity and that the editing levels should be as high as possible to mimic DNA mutation (in other words, to dilute the deleterious alleles in transcriptome). Then, the restorative hypothesis should predict that the nonsynonymous editing (for instance, C-to-U) but not synonymous editing should be observed on the sites with recent T-to-C DNA mutations. Here, "recent" means the time from an ancestral node to the extant species. The ancestral node may be far away according to the species of interest, but we still use "recent" for simplicity to describe the DNA mutations taken place at a time from the ancestral node to present.

Note that both diversifying and restorative hypotheses predict higher editing levels on recoding sites compared to synonymous sites. Therefore, to distinguish between the two situations, additional analyses apart from the recoding-synonymous comparison should be performed. In this article, we collected the chloroplast C-to-U RNA editomes of 21 vascular plants (11 angiosperms, four gymnosperms, and six ferns) from a previous study (Zhang et al. 2022). These 21 species covered a wide range of representative clades in the plant kingdom. We aim to testify whether the plant editomes typically conform to the restorative hypothesis. Although the restorative effect of a few

C-to-U RNA editing sites has been proposed by early studies before the "Omics era" (Shields and Wolfe 1997), it remains unclear whether the global editome follows this rule. Evolutionary biology particularly focuses on the genome-wide pattern instead of special cases. In this work, we found that the plant editomes nicely conform to the restorative hypothesis in many ways: (i) Nonsynonymous editing sites are more frequent and have higher editing levels than synonymous sites; (ii) editing levels are extremely high and show weak tissue-specificity in plants; and (iii) on the inferred genomic sites with recent T-to-C mutations, nonsynonymous sites but not synonymous sites are compensated by C-to-U RNA editing. By clarifying these features, we unravel the evolutionary trajectory of plant RNA editing, and deepen the understanding on the biological significance and function of this editing mechanism.

RESULTS

Nonsynonymous editing is more prevalent than synonymous editing in vascular plants

We profiled the chloroplast C-to-U RNA editing sites in 21 collected vascular plants including 11 angiosperms, four gymnosperms, and six ferns (Fig. 1C). We classified nonsynonymous and synonymous editing sites according to the functional consequence (Table 1). From the codon table we know that all C-to-T mutations on the second codon position are nonsynonymous. All C-to-T mutations on the third codon position are synonymous. Most C-to-T mutations on the first codon position are nonsynonymous and but there are two exceptions: CTA > TTA and CTG > TTG are synonymous changes and both cases encode Leucine. Therefore, classifying editing sites according to codon positions may not accurately reflect the functional consequence of editing events.

In our data, we found that the numbers of nonsynonymous RNA editing sites are much greater than the numbers of synonymous editing sites in all 21 species (Fig. 1C). For the editing level (the percentage of edited molecules at a particular site) comparison, nine out of 11 angiosperms show higher nonsynonymous editing levels than synonymous editing levels, six of which are statistically significant. Similarly, all four gymnosperms and four out of six ferns show significantly higher nonsynonymous editing levels than synonymous editing levels (Fig. 1C). We then defined cumulative editing level (the sum of editing levels) to represent the level-weighted number of editing sites. Nonsynonymous (recoding) sites have remarkably higher cumulative editing levels than synonymous sites in all 21 species (Fig. 1C). These results all indicate the adaptiveness of recoding events over the neutrally evolving synonymous sites.

TABLE 1. The relationship between codon position of C-to-U(T) editing sites and the functional consequence

Codon position	First	Second	Third
Nonsynonymous	14	16	0
Synonymous	2	0	16

Note that the coding sequences intrinsically have more nonsynonymous sites than synonymous sites. The potential bias caused by this nature should be considered. In the chloroplast genes of *A. thaliana*, if we manually change all the Cs to Ts, we will obtain 9570 nonsynonymous mutations and 4081 synonymous mutations. Thus, the expected Nonsyn/Syn ratio under neutral evolution is 9570/4081 = 2.35. However, the observed Nonsyn/Syn for C-to-U RNA editing in *A. thaliana* is 41/2 = 20.5, which is significantly higher than the random expectation 2.35 ($P = 7.3 \times 10^{-5}$, Fisher's exact test). This suggests that the enrichment of nonsynonymous editing is not caused by artifact. In fact, the Nonsyn/Syn ratios of C-to-U editing in 21 vascular plants are as high as 12.6 ± 1.7 (mean ± standard error). Virtually no coding sequence could produce such high intrinsic Nonsyn/Syn ratios due to codon degeneracy.

Moreover, when we calculated the cumulative editing levels of orthologous genes among these 21 plants, we found quite a few genes that are heavily edited (nonsynonymous) in all six fern species (Fig. 2A), suggesting that the need for recoding these genes is evolutionarily conserved. Interestingly, we also found that the strength of nonsynonymous editing on each gene (indicated by the cumulative editing level per gene, nonsynonymous minus synonymous) is positively correlated with the gene expression level in all plant species (Fig. 2B). This is the supporting evidence for adaptive editing because the beneficial effects of recoding, if any, would be amplified in the highly expressed genes. Inversely, if recoding is deleterious, then we should expect recoding events to be avoided in highly expressed genes.

Considering that the power of detecting RNA editing sites increases with gene expression level, highly expressed genes are biased toward high cumulative editing level. We should cancel this bias by calculating mean editing level instead of cumulative editing level of each gene. We robustly found that the mean editing level of each gene (nonsynonymous minus synonymous) is positively correlated with gene expression level, reinforcing the evidence of adaptive recoding (Fig. 2C).

However, one should note that the prevalence of nonsynonymous editing over synonymous editing does not distinguish between the diversifying hypothesis and restorative hypotheses. Both hypotheses would expect such patterns. Therefore, we need to find evidences from other aspects to determine the adaptive nature of C-to-U RNA editing.

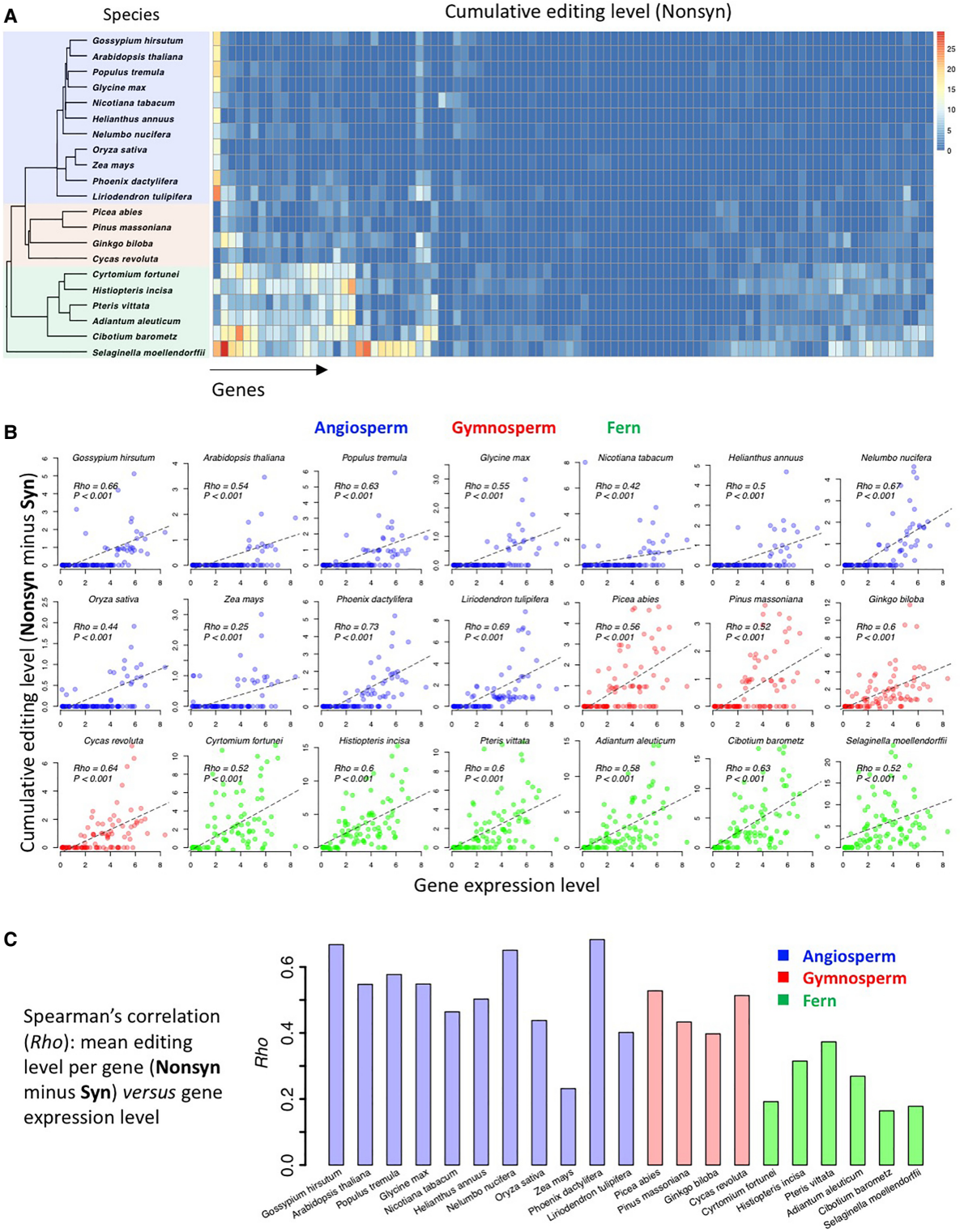


FIGURE 2. Cumulative editing levels of genes. (*A*) The cumulative nonsynonymous editing levels in each edited gene of all species. (*B*) Spearman's correlation (Rho) between cumulative editing level per gene (nonsynonymous minus synonymous) and gene expression level. Angiosperms are colored in blue; gymnosperms are colored in red; ferns are colored in green. (*C*) Spearman's correlation (Rho) between mean editing level per gene (nonsynonymous minus synonymous) and gene expression level. For the 21 Rho's, the geometric mean of the 21 corresponding *P*-values is 1.4×10^{-6}.

Editing levels in plants are extremely high and show low tissue-specificity

The diversifying hypothesis stresses the flexibility of temporal-spatial control of RNA editing, and thus predicts that RNA editing is tissue-specific. But the editing level is not necessarily very high under the diversifying hypothesis. In contrast, the restorative hypothesis should require a constantly high editing level in all tissues in order to fully reverse (or dilute) the deleterious allele in transcripts. Thus, the restorative hypothesis does not require tissue-specificity of editing activity (see Discussion for the limitation of this tissue-specificity explanation).

We first compared the C-to-U editing levels in plants *versus* the A-to-I editing levels in fly (*Drosophila melanogaster*) brains (Duan et al. 2017) and honeybee (*Apis mellifera*) heads (Duan et al. 2021) which we previously published. Only nonsynonymous sites were considered. We found that the nonsynonymous editing levels in three major plant clades are extremely high (median ~ 0.8) while the nonsynonymous editing levels in flies and bees are relatively low (median < 0.2) (Fig. 3A). Since the A-to-I RNA editing in flies and bees is known to be adaptive due to the diversifying effect (Duan et al. 2017, 2021), it becomes very likely (by sharp contrast) that the RNA editing landscape in plants conforms to the restorative hypothesis because high editing levels are intuitively considered to mimic DNA mutations.

Next, we set out to study the tissue-specificity of RNA editing events. We retrieved the transcriptomes of shoots and roots of *Arabidopsis thaliana* (Hsu et al. 2016) and calculated the C-to-U editing levels in chloroplast genes (shoot refers to the part above ground and root refers to the part underground). Meanwhile, we obtained the A-to-I editing levels in heads, thoraxes, and abdomens of honeybees from our previous study (Duan et al. 2021). Strikingly, nonsynonymous editing levels are comparably high in shoots, roots (Hsu et al. 2016), and leaves (this study) of *A. thaliana*, while in honeybees, editing is dominant in heads but not in thoraxes or abdomens (Fig. 3B). For the levels of individual C-to-U editing sites, the Pearson correlation coefficient between *A. thaliana* roots and shoots is 0.89 (Fig. 3C, also see Discussion section for the implication). In bees, since A-to-I editing is almost absent in nonhead tissues, there is no point to calculate the cross-tissue correlation of editing levels.

The facts that C-to-U RNA editing levels in plants are high and lack tissue-specificity suggest the restorative purpose of RNA editing events. The transcripts in the whole plant need to be edited at a sufficiently high level to reduce the deleterious effect of the genomic allele. Notably, the lack of tissue-specificity of C-to-U editing might be largely due to the ubiquitous expression of the editing enzymes and target genes. Many of the C-to-U edited genes in plants are housekeeping genes like ribosome

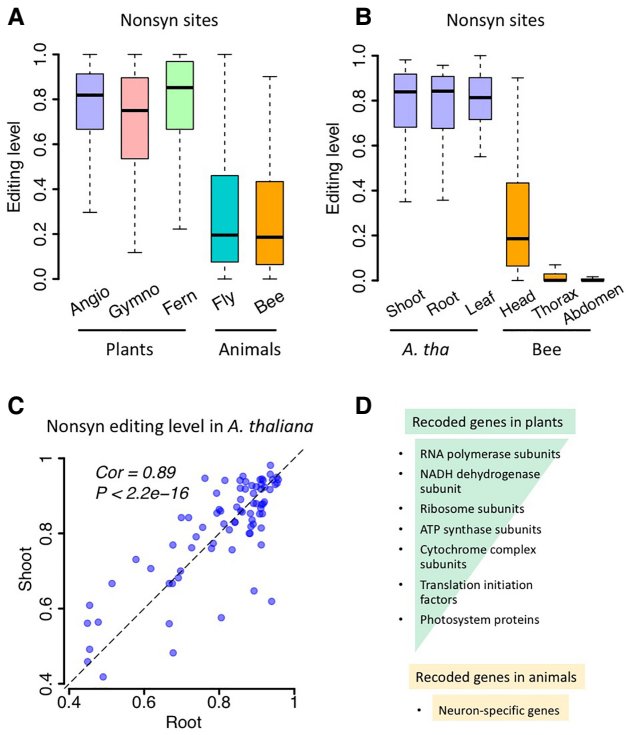


FIGURE 3. Comparison of editing levels. (*A*) Nonsynonymous editing levels in plants (angiosperms, gymnosperms, and ferns used in this study) and animals. Fly, *Drosophila melanogaster*; Honeybee, *Apis mellifera*. Plants have considerably higher editing levels than animals, supporting the restorative hypothesis. (*B*) Nonsynonymous editing levels in shoots, roots, and leaves of *Arabidopsis thaliana*, and in different tissues of honeybee. *A. thaliana* obviously has less tissue-specificity than honeybee, supporting the restorative hypothesis. (*C*) Correlation of nonsynonymous editing levels between roots and shoots. (*D*) Genes enriched with nonsynonymous RNA editing sites. Genes are ranked with decreasing numbers of nonsynonymous editing sites.

protein subunits (RPL and RPS), translation initiation factor, RNA polymerase subunits, NADH dehydrogenase subunits, and ATP synthase subunits (Fig. 3D). These housekeeping genes should be ubiquitously expressed in all plant tissues and developmental stages. The genomic allele in housekeeping genes, if deleterious, should suffer from strong purifying selection so that it needs to be fully recoded at RNA level. In contrast, the A-to-I recoded genes in animals are mainly neuron-specific genes (Fig. 3D). Accordingly, ADARs are also mainly expressed in animal nerve systems (Savva et al. 2012). Therefore, not all tissues in animals have the demand for high recoding levels.

The above comparison between plants versus insects editing sites has revealed distinct biological significance of RNA editing. Recoding events in insects, which are known to follow the diversifying hypothesis (Duan et al. 2017, 2021), are neuron-specific and the editing levels of which are low (Fig. 3). For plants, we found that the recoding sites lack tissue-specificity and their levels are

extremely high (Fig. 3). All these observations support the restorative purpose of nonsynonymous RNA editing in plants.

Nonsynonymous C-to-U editing is likely to take place following recent T-to-C DNA mutations

While the high editing levels and low tissue-specificity of nonsynonymous C-to-U editing serve as evidence to support the restorative hypothesis, the most straightforward evidence to confirm this hypothesis is to directly observe nonsynonymous C-to-U editing events taking place on the genomic cytidines that have recently mutated from an ancestral T to the current C (Fig. 4A). If the ancestral (nonsynonymous) DNA mutation from T to C is deleterious, then the C-to-U editing events will reverse the deleterious effect (Fig. 4A). In contrast, no such demand is needed on the neutral synonymous mutation sites. Here, "recent" refers to the time from a given ancestral node to the extant species of interest.

We set out to look for potential sites that have undergone recent T-to-C DNA mutations followed by present C-to-U RNA editing. For all the 21 species used in our study, a fern species *S. moellendorffi* is the outgroup of all other 20 species (including 15 seed plants and five "other ferns") (Fig. 4B). If a genomic T is seen in *S. moe* and in at least one other 20 species (15 seed plants and five "other ferns"), then the ancestral sequence of seed plants and "other ferns" is likely to be T (Fig. 4B). Next, if one of the 20 species has a C at this position, then it probably underwent a recent T-to-C DNA mutation. At such a particular site, we could count how many species underwent T-to-C DNA mutations and how many of them acquired C-to-U RNA editing to restore the ancestral DNA sequence T (Fig. 4B). For example, if five species experienced recent T-to-C DNA mutations, then "fully restored" means all five species are C-to-U edited at this site; "partially restored" means only a fraction of species are C-to-U edited; "none restored" means no species has C-to-U editing after recent T-to-C DNA mutation (Fig. 4B).

We fully take advantage of the data provided by the previous study (Zhang et al. 2022). Gene *atpA* is heavily edited in these plant species. We totally obtained 33 nonsynonymous sites and three synonymous sites that met the criteria of identifying a recent T-to-C DNA mutation (Fig. 4B). Among the 33 nonsynonymous T-to-C mutation sites, 31 sites are fully restored by C-to-U editing events and two sites are partially restored. In contrast, among the three synonymous T-to-C mutation sites, two sites are partially restored and one site belongs to "none restored" (Fig. 4B). The different restorative tendencies between nonsynonymous and synonymous sites are significant (Fig. 4B). Moreover, for the partially restored sites, the two nonsynonymous sites have obviously more species possessing C-to-U editing compared to the two synonymous sites (Fig. 4B). All these patterns support the restorative nature of nonsynonymous C-to-U editing.

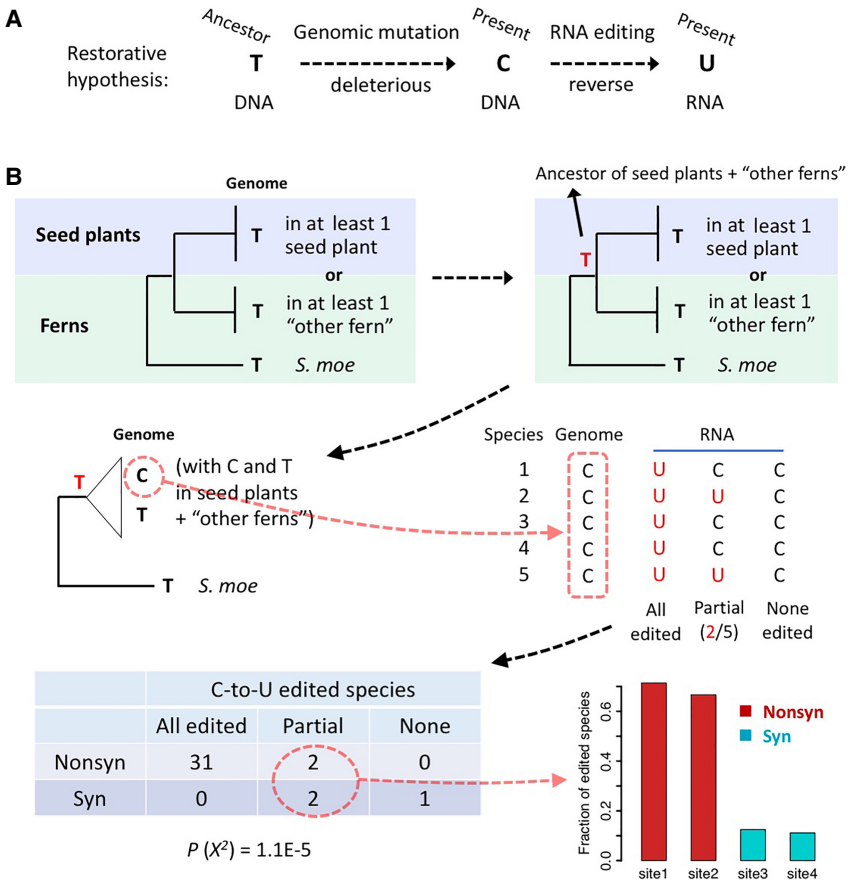


FIGURE 4. Direct evidence for the restorative hypothesis. (*A*) Nonsynonymous C-to-U editing is expected to take place following recent T-to-C DNA mutation. (*B*) Inference of ancestral state of 15 seed plants and five "other ferns." "Other ferns" refer to the fern species used in this study excluding *S. moellendorffi*. *S. moe* is the outgroup of seed plants and "other ferns." *P*-value *under* the statistic table was calculated using χ^2 test.

The splicing-related role of PPR proteins explains the prevalence and significance of restored start codons by C-to-U editing

The endosymbiotic origin of chloroplast and mitochondrial genomes determines that (i) these non-nuclear genes are subjected to post-

transcriptional splicing to be separated from the polycistron; (ii) most internal genes have no UTRs so that the start codons are the closest to splicing sites.

When we examined the relative position of C-to-U editing sites in the gene, we found a striking enrichment in the second nucleotide of the gene: that is, the second position of the start codon. Among the genes with identified editing sites, the fraction of editing in the start codon is significantly higher than the fraction of editing sites in the remaining CDS bodies (Fig. 5A). Among the 21 plant species, 16 species have C-to-U editing events in the start codons of 57 unique genes (Fig. 5B). Intriguingly, this means that the genomically annotated start codon for these genes is ACG and it is converted to the canonical start codon AUG by C-to-U editing. This type of mutation is typically named "start codon gained" mutation. Given that many of these edited genes are essential genes like ribosome protein subunits (Fig. 5B), this C-to-U conversion on start codon (which is required by RNA translation) should be mandatory for the proper function of cells.

One of these "start codon gained" C-to-U events is located in gene *atpA*. This site is only edited in two fern species *Histiopteris incisa* and *Cibotium barometz* while the other species (including the outgroup species *Selaginella moellendorffii*) have a T in the genome (Fig. 5C). This confirms the restorative role of this C-to-U RNA editing event. It remains to be seen whether this "start codon restored" pattern holds true for other edited genes. However, since the majority of the 21 species have the correct version of those genes/proteins, it could be inferred that: (i) in most species the genomic start codon is ATG(AUG); (ii) the start codon is recently genomically mutated to ACG in only a few species; (iii) then C-to-U RNA editing restores the canonical start codon AUG.

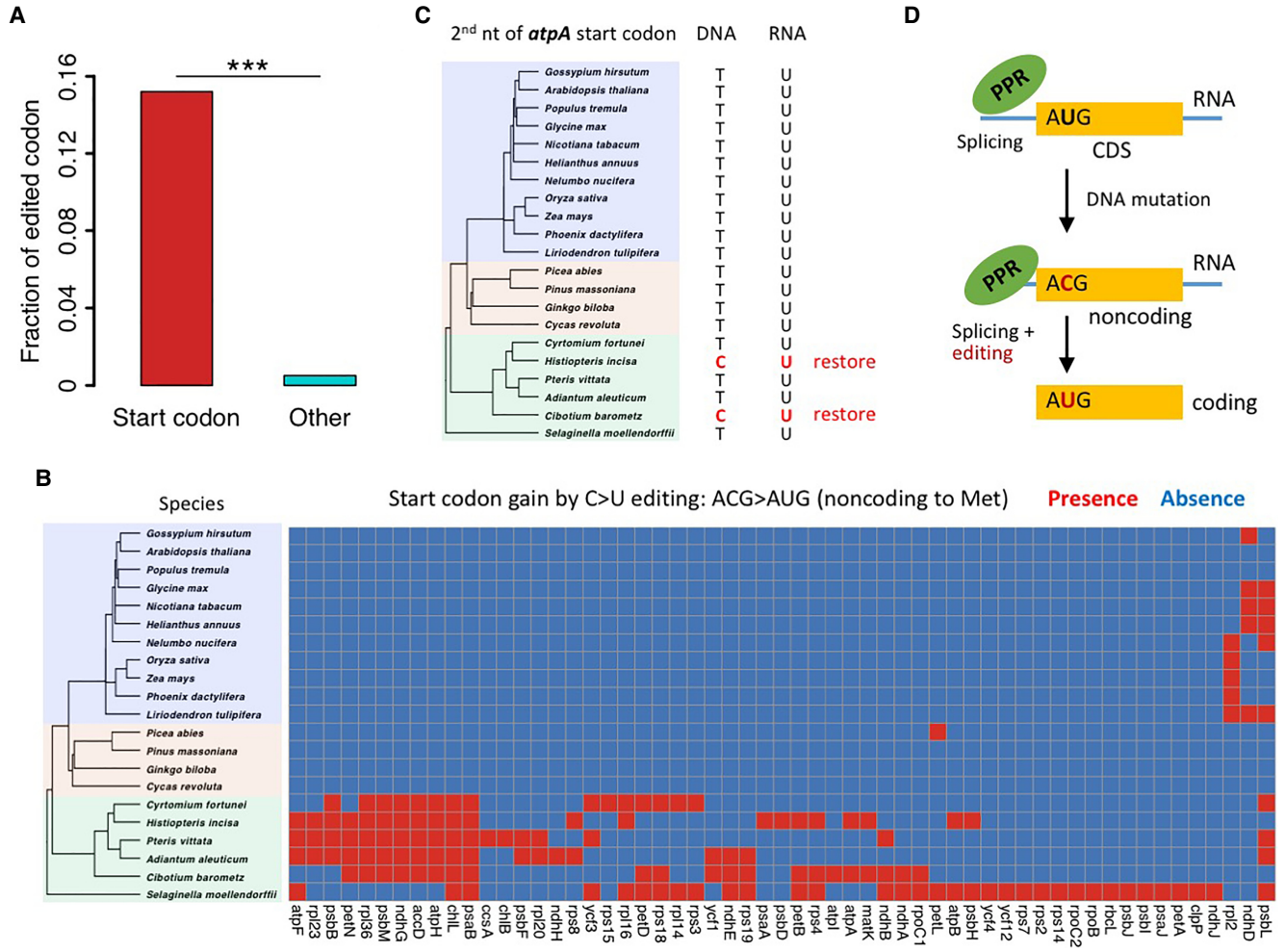


FIGURE 5. C-to-U RNA editing restores canonical start codon AUG. (*A*) Barplot showing the fraction of C-to-U editing on start codon or other codons. *P*-value was calculated by Fisher's exact test. (***) $P < 0.001$. (*B*) Heatmap showing the presence or absence of C-to-U editing on start codons. Each column is an orthologous gene across 21 species. (*C*) C-to-U editing restores start codon AUG in two fern species, whereas in other species the DNA encodes ATG. (*D*) Explanation for the prevalence of restorative C-to-U editing in start codons. PPR proteins are responsible for both RNA splicing and editing.

The loss of the canonical start codon by point mutation should be highly deleterious but it is corrected by C-to-U RNA editing. However, it is mechanistically unclear why should PPR proteins have a strong preference on editing the C in the initiator ACG instead of other internal Cs. Interestingly, it is known that PPR proteins participate in RNA splicing as well (Wang et al. 2022). The start codon should be the nearest codon to the splicing sites in the polycistron. Normally, PPR proteins splice RNAs (Fig. 5D). When the start codon is accidentally mutated to ACG in a species, then PPR proteins could simultaneously exert splicing and also edit the start codon to restore AUG (Fig. 5D). The high accessibility of start codon to PPR proteins determines that PPR might serve as the security guard to ensure the proper function of the start codon.

Notably, regardless of the ancestral state, the function of a "start codon gained" mutation could hardly be explained by the diversifying hypothesis because the mutation will create a completely unknown protein, which might be toxic to the organism. The "start codon gained" mutation is only meaningful (or tolerated) if this protein already existed in the ancestral node (Fig. 5D). Therefore, this ACG-to-AUG RNA editing only fits the restorative hypothesis.

DISCUSSION

Summary of main findings and advances

From multiple aspects, we verified that C-to-U RNA editing in vascular plants confers its adaptiveness by reversing genomic nucleotides. The evidences supporting our conclusion are: (i) nonsynonymous editing sites are more frequent and have higher editing levels than synonymous sites; (ii) C-to-U editing levels are extremely high in plants and show weak tissue-specificity; (iii) on the inferred genomic sites with recent T-to-C DNA mutations, nonsynonymous sites but not synonymous sites are compensated by C-to-U RNA editing.

Although the restorative effect of plant C-to-U RNA editing has already been proposed by previous studies (Shields and Wolfe 1997), there are still critical gaps in this field: (i) no systematic analyses were done at multiple-clade and genome-wide level to test the restorative hypothesis; (ii) in addition to reversing DNA mutation, the restorative hypothesis will predict many other "concomitant" patterns such as globally high recoding levels and the lack of tissue-specificity. These patterns are largely untested; (iii) the restorative hypothesis is only meaningful when it is compared to the diversifying hypothesis in the light of evolution. Case studies on the restorative role of single sites do not make sense because the low-throughput observations could not preclude the diversifying hypothesis at all. For example, without observing low editing levels in insects (Duan et al. 2017, 2021), one could not confirm the restorative role of plant editing despite its high editing level.

Seed plants have fewer editing sites than ferns but the editing levels are constantly high

Although the seed plants especially the angiosperms diverged later than the early split ferns, we should not claim that angiosperms are higher plants compared to ferns. In fact, all extant species are "equally ancient" since they just followed different evolutionary routes. The fewer C-to-U RNA editing sites identified in seed plants compared to ferns (Fig. 1C) could not be interpreted as a decrease in the number of editing sites "during evolution" because the direction of evolution is not from the extent ferns to the extant angiosperms. It is just an observation that the extant seed plants having less C-to-U editing sites than extant ferns may be due to the higher demand for C-to-U restoration in ferns (for unknown reasons).

Interestingly, no matter how few editing sites are detected in a particular angiosperm species, once a C is edited, it must be edited at a sufficiently high level (in order to reverse the genomic nucleotide). This observation could only be explained by the restorative hypothesis but not the diversifying hypothesis. Regarding the diversifying hypothesis, if the demand for diversifying the proteome has decreased in angiosperms compared to ferns, then not only the numbers of editing sites but also the editing levels should decrease in angiosperms. Mechanistically, the presence/absence of editing events and the editing levels are usually determined by similar *cis* or *trans* features (at least this is true for A-to-I editing in animals) (Zhang et al. 2017) so that there is no mechanism to solely reduce the number of editing sites while keeping the editing levels constantly high. The high editing levels should be maintained by purifying selection. Only the restorative hypothesis could explain this observation: different plant species have different demands for restoration, once a site needs to be reversed by C-to-U recoding, then it must be corrected at a high level. Suboptimal context leading to a low C-to-U recoding level should be eliminated by purifying selection. This selection force promotes nonsynonymous editing to mimic DNA mutations.

What, how, and why

Understanding a biological phenomenon requires the knowledge of "What," "How," and "Why." Now that we have observed the "What": C-to-U RNA editing levels are constantly very high in plants but the numbers of editing events vary widely across different species. This phenomenon resembles a terminology known as "all or none" (Duan et al. 2017) that describes the pleiotropic effect of DNA mutations. But the "How" behind this phenomenon remains unclear: What is the molecular mechanism underlying this "all or none" property of C-to-U RNA editing? We guess that the sequence context, coupled with the preference of the editing enzymes (PPR proteins), accounts for the extremely high editing levels (and the "all or none" property).

While the molecular mechanism (How) remains speculative, we almost know the "Why": Why should C-to-U editing have the "all or none" property? Because the aim of C-to-U editing is to restore the genomic nucleotide (which is similar to DNA mutations), then it must be "all or none" to fully restore the genomic sequence.

Not all editing in animals is diversifying

While we have compared the restorative C-to-U RNA editing in plants with the diversifying A-to-I RNA editing in a few insect species, we did not claim that all A-to-I editing sites play a diversifying role in all animal species. For example, a few recoding sites in *D. melanogaster* might be used for restoration (Tian et al. 2008). More importantly, in mammals like humans, <1% of the A-to-I editing sites are nonsynonymous sites according to the databases (Ramaswami and Li 2014; Picardi et al. 2017) or literatures (Liang and Landweber 2007; Mazloomian and Meyer 2015; Costa Cruz et al. 2020; Adetula et al. 2021). It is hard to believe that mammalian editing sites are designed for diversifying the proteome at the genome-wide level. In fact, the majority of human RNA editing events take place in *Alu* repeats which label endogenous RNAs as "self" (to distinguish self RNAs from non-self RNAs like exogenous viral RNAs) (Liddicoat et al. 2015). The prevalent editing in human *Alu*s is a mechanism that prevents unnecessary immune responses. Even for the A-to-I editing sites in human CDS, no signals of adaptation are observed at the global level (Xu and Zhang 2014, 2015), although there might be particular nonsynonymous editing sites with crucial functions (Sommer et al. 1991; Higuchi et al. 1993). Therefore, the RNA editing pattern in humans does not belong to the scenario predicted by either the diversifying or restorative hypothesis because both hypotheses will require a globally adaptive editing profile (where nonsynonymous editing should be more prevalent than synonymous editing). In plants, at least for the 21 species we examined, the editing profiles nicely fit the restorative hypothesis. It is interesting to think why nonsynonymous RNA editing in plants is restorative, but in insects (as least in *Drosophila*) is diversifying. Possibly, animals (insects) are mobile because they will encounter a more changeable environment, which makes rapid adaptation essential. RNA editing provides such a mechanism to recode the genetic information in a controllable manner, facilitating rapid adaptation. However, this mobility explanation is imperfect because plants may also face seasonal environmental changes. More evidences and mechanisms are needed to unravel the evolutionary driving forces of RNA editing in plants and animals.

There is a potential concern whether A-to-I editing in animals and C-to-U editing in plants are comparable due to their distinct *cis* and *trans* features: nuclear genes versus non-nuclear genes, ADARs versus PPRs, and other differences. This concern could not be eased at this stage but currently there is no evidence to show that RNA editing mechanisms in animals and plants are not comparable.

Does evolutionary tinkering fit the restorative hypothesis of C-to-U RNA editing?

Under the restorative hypothesis, it remains to be investigated why it is easier to acquire C-to-U editing than acquiring a C-to-T DNA mutation. Acquisition of an editing site (of high editing level) during evolution usually requires the optimization of the sequence context, the process of which entails numerous DNA mutations. Given that only a single C-to-T mutation would be enough to correct the unwanted cytidine, why should the organisms design the entire sequence context to create a C-to-U editing site? In sharp contrast, for the diversifying A-to-I editing in animals, there are no alternative ways to achieve the transcriptomic plasticity (at least not by A-to-G DNA mutation). Thus, A-to-I editing is the only way to diversify the proteome in a flexible manner: it could be obtained by "continuous probing" (evolutionary tinkering) (Gommans et al. 2009) given a long-enough time.

Limitations of the tissue-specificity explanation

Note that we have observed that plant C-to-U editing sites have weaker tissue-specificity than animal A-to-I editing sites. We did not claim that every C-to-U editing site has exactly the same editing level across all plant tissues (as it is technically impossible to observe this). In our data, the Pearson correlation coefficient of editing levels of root versus shoot is 0.89 (but not 1.0). This means that we could always find an outlier that deviates from the $y = x$ diagonal. However, a few outliers do not represent the global pattern. Not surprisingly, it was reported that some C-to-U editing sites exhibit tissue-specificity (Bock et al. 1993; Miyata and Sugita 2004; Guo et al. 2015), but they only represent a few sites in a limited number of plant genes. Moreover, the so-called tissue-specificity (variability) of C-to-U editing in plants is much weaker than what we saw in insects where A-to-I editing almost solely exists in neural tissues.

Next, in theory, no tissue-specificity does not completely preclude the diversifying hypothesis. We could envision an extreme case where all tissues need to be diversified and then the diversifying editing sites would be ubiquitously expressed. However, according to the Occam's razor principle, the restorative hypothesis is still more natural to explain the lack of tissue-specificity of C-to-U editing. If all tissues need to be diversified by RNA editing, why not directly introduce a heterozygous SNP and let it be maintained by balancing selection?

Conclusion

In conclusion, nonsynonymous C-to-U RNA editing (recoding) in plants is adaptive due to its restorative effects. The

recoding levels are high and are required across the whole plant. The long-term puzzle underlying the significance of plant RNA editing is systematically unraveled.

MATERIALS AND METHODS

Data availability

The C-to-U RNA editing sites and levels in 21 vascular plants were downloaded from the supplemental materials of a previous study (Zhang et al. 2022). All these editing sites were generated from leaves of each plant species. Then, the transcriptomes of roots and shoots of *Arabidopsis thaliana* were retrieved from an earlier literature (Hsu et al. 2016). A-to-I editing sites in *Drosophila melanogaster* (fruitfly) and *Apis mellifera* (honeybee) were obtained from our previous work (Duan et al. 2017, 2021).

Phylogeny of 21 vascular plants

We uploaded the Latin names of the 21 vascular plants to the TimeTree website (http://www.timetree.org/) and obtained their phylogeny (nwk format). The phylogenetic tree is visualized by FigTree (http://tree.bio.ed.ac.uk/software/).

Mapping and variant calling in *Arabidopsis thaliana*

We retrieved the TAIR10 version of *Arabidopsis thaliana* genome and mapped the transcriptome to the CDS of chloroplast genes (gene ID beginning with ATCG) using bowtie2 (Langmead and Salzberg 2012). Variants were called using samtools (Li et al. 2009). C-to-T variants, which were dominant across the total variants, were regarded as C-to-U RNA editing sites. Annotation of nonsynonymous and synonymous sites was done by comparing the amino acids before and after C-to-T alteration.

Cumulative editing level

Cumulative editing level is defined as the sum of editing levels of all sites of interest (Alon et al. 2015; Duan et al. 2021). Sites of interest could be the sites within the same gene or all sites of a species. The cumulative editing level is an indicator of total editing activity in a sample (or can be otherwise understood as the level-weighted number of editing sites).

Gene expression level

The original study (Zhang et al. 2022) has provided the sequencing depth on each editing site. For each gene, it is presumed that the mRNA-seq reads should be evenly covered within a gene. Therefore, the average sequencing depth of editing sites within the same gene was used as the gene expression level.

Statistical tests

All statistical tests were performed in R language (version 3.6.3). The graphical works were also done in R environment.

DATA DEPOSITION

The C-to-U RNA editing sites and levels in 21 vascular plants were downloaded from the supplemental materials of a previous study (Zhang et al. 2022).

ACKNOWLEDGMENTS

We thank the 2115 Talent Development Program of China Agricultural University for financial support. This study is financially supported by China Postdoctoral Science Foundation 2022M710221, the National Natural Science Foundation of China (no. 31922012), and the 2115 Talent Development Program of China Agricultural University.

Author contributions: Conceptualization and supervision: Y.D., W.C., and H.L. Data analysis: Y.D. Writing—original draft: Y.D. Writing—review and editing: Y.D., W.C., and H.L.

Received September 16, 2022; accepted October 30, 2022.

REFERENCES

Adetula AA, Fan X, Zhang Y, Yao Y, Yan J, Chen M, Tang Y, Liu Y, Yi G, Li K, et al. 2021. Landscape of tissue-specific RNA Editome provides insight into co-regulated and altered gene expression in pigs (*Sus-scrofa*). *RNA Biol* **18:** 439–450. doi:10.1080/15476286.2021.1954380

Alon S, Garrett SC, Levanon EY, Olson S, Graveley BR, Rosenthal JJ, Eisenberg E. 2015. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *Elife* **4:** e05198. doi:10.7554/eLife.05198

Bock R, Hagemann R, Kossel H, Kudla J. 1993. Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids–a new regulatory mechanism? *Mol Gen Genet* **240:** 238–244. doi:10.1007/BF00277062

Costa Cruz PH, Kato Y, Nakahama T, Shibuya T, Kawahara Y. 2020. A comparative analysis of ADAR mutant mice reveals site-specific regulation of RNA editing. *RNA* **26:** 454–469. doi:10.1261/rna.072728.119

Duan Y, Dou S, Luo S, Zhang H, Lu J. 2017. Adaptation of A-to-I RNA editing in *Drosophila*. *PLoS Genet* **13:** e1006648. doi:10.1371/journal.pgen.1006648

Duan Y, Dou S, Zhang H, Wu C, Wu M, Lu J. 2018. Linkage of A-to-I RNA editing in metazoans and the impact on genome evolution. *Mol Biol Evol* **35:** 132–148. doi:10.1093/molbev/msx274

Duan Y, Dou S, Porath HT, Huang J, Eisenberg E, Lu J. 2021. A-to-I RNA editing in honeybees shows signals of adaptation and convergent evolution. *iScience* **24:** 101983. doi:10.1016/j.isci.2020.101983

Duan Y, Tang X, Lu J. 2022. Evolutionary driving forces of A-to-I editing in metazoans. *Wiley Interdiscip Rev RNA* **13:** e1666. doi:10.1002/wrna.1666

Gommans WM, Mullen SP, Maas S. 2009. RNA editing: a driving force for adaptive evolution? *Bioessays* **31:** 1137–1145. doi:10.1002/bies.200900045

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471:** 473–479. doi:10.1038/nature09715

Gray MW, Covello PS. 1993. RNA editing in plant-mitochondria and chloroplasts. *FASEB J* **7:** 64–71. doi:10.1096/fasebj.7.1.8422976

Guo W, Grewe F, Mower JP. 2015. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS ONE* **10:** e0117075. doi:10.1371/journal.pone.0117075

Higuchi M, Single FN, Kohler M, Sommer B, Sprengel R, Seeburg PH. 1993. RNA editing of AMPA receptor subunit GluR-B—a base-paired intron-exon structure determines position and efficiency. *Cell* **75:** 1361–1370. doi:10.1016/0092-8674(93)90622-W

Hsu PY, Calviello L, Wu HYL, Li FW, Rothfels CJ, Ohler U, Benfey PN. 2016. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc Natl Acad Sci* **113:** E7126–E7135. doi:10.1073/pnas.1614788113

Jiang D, Zhang J. 2019. The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. *Nat Commun* **10:** 5411. doi:10.1038/s41467-019-13275-2

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Liang H, Landweber LF. 2007. Hypothesis: RNA editing of microRNA target sites in humans? *RNA* **13:** 463–467. doi:10.1261/rna.296407

Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, Li JB, Seeburg PH, Walkley CR. 2015. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349:** 1115–1120. doi:10.1126/science.aac7049

Mazloomian A, Meyer IM. 2015. Genome-wide identification and characterization of tissue-specific RNA editing events in *D. melanogaster* and their potential role in regulating alternative splicing. *RNA Biol* **12:** 1391–1401. doi:10.1080/15476286.2015.1107703

Miyata Y, Sugita M. 2004. Tissue- and stage-specific RNA editing of *rps14* transcripts in moss (*Physcomitrella patens*) chloroplasts. *J Plant Physiol* **161:** 113–115. doi:10.1078/0176-1617-01220

Palladino MJ, Keegan LP, O'Connell MA, Reenan RA. 2000. *dADAR*, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* **6:** 1004–1018. doi:10.1017/S1355838200000248

Picardi E, D'Erchia AM, Lo Giudice C, Pesole G. 2017. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* **45:** D750–D757. doi:10.1093/nar/gkw767

Ramaswami G, Li JB. 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res* **42:** D109–D113. doi:10.1093/nar/gkt996

Savva YA, Rieder LE, Reenan RA. 2012. The ADAR protein family. *Genome Biol* **13:** 252. doi:10.1186/gb-2012-13-12-252

Shields DC, Wolfe KH. 1997. Accelerated evolution of sites undergoing mRNA editing in plant mitochondria and chloroplasts. *Mol Biol Evol* **14:** 344–349. doi:10.1093/oxfordjournals.molbev.a025768

Sommer B, Kohler M, Sprengel R, Seeburg PH. 1991. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67:** 11–19. doi:10.1016/0092-8674(91)90568-J

Tian N, Wu X, Zhang Y, Jin Y. 2008. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. *RNA* **14:** 211–216. doi:10.1261/rna.797108

Wang X, Wang J, Li S, Lu C, Sui N. 2022. An overview of RNA splicing and functioning of splicing factors in land plant chloroplasts. *RNA Biol* **19:** 897–907. doi:10.1080/15476286.2022.2096801

Xu G, Zhang J. 2014. Human coding RNA editing is generally nonadaptive. *Proc Natl Acad Sci* **111:** 3769–3774. doi:10.1073/pnas.1321745111

Xu G, Zhang J. 2015. In search of beneficial coding RNA editing. *Mol Biol Evol* **32:** 536–541. doi:10.1093/molbev/msu314

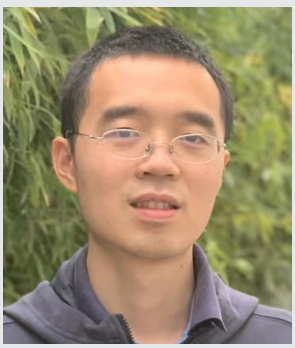
Yablonovitch AL, Deng P, Jacobson D, Li JB. 2017. The evolution and adaptation of A-to-I RNA editing. *PLoS Genet* **13:** e1007064. doi:10.1371/journal.pgen.1007064

Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, et al. 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504:** 168–171. doi:10.1038/nature12651

Zhang R, Deng P, Jacobson D, Li JB. 2017. Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. *PLoS Genet* **13:** e1006563. doi:10.1371/journal.pgen.1006563

Zhang A, Fang J, Zhang X. 2022. Diversity of RNA editing in chloroplast transcripts across three main plant clades. *Res. Square* doi:10.21203/rs.3.rs-1838292/v1

MEET THE FIRST AUTHOR



Yuange Duan

***Meet the First Author(s)* is an editorial feature within *RNA*, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of *RNA* and the RNA research community. Yuange Duan is first author of this paper, "Chloroplast C-to-U RNA editing in vascular plants is adaptive due to its restorative effect: testing the restorative hypothesis." Yuange is an Associate Professor at the China Agricultural University. Research in the laboratory focuses on the evolutionary genomics of insects and plants.**

Continued

What are the major results described in your paper and how do they impact this branch of the field?

Nonsynonymous C-to-U RNA editing in plants would reverse DNA mutations. This is why the editing events are adaptive and are maintained to a high editing level with low tissue-specificity. This study systematically tested one of the two major hypotheses that explain the adaptive (functional) RNA editing, namely, the diversifying hypothesis and the restorative hypothesis. Now we know that the behavior of plant RNA editing supports the restorative hypothesis.

What led you to study RNA or this aspect of RNA science?

I have always focused on RNA editing. The adaptive RNA editing in many insect species has been revealed by my previous work so this time I turned to plant RNA editing.

If you were able to give one piece of advice to your younger self, what would that be?

Don't be afraid if other scientists disagree with you. If this happens, that means your work is influential.

Are there specific individuals or groups who have influenced your philosophy or approach to science?

Sure, that would be George Zhang from the University of Michigan. Much of my previous work has followed his philosophy (such as the bisection method in studying the adaptiveness of a feature during evolution).

What are your subsequent near- or long-term career plans?

I plan to study the evolutionary genomics of insects. If necessary, I will continue to focus on A-to-I RNA editing.