New Phytologist

Research

Comparative phylogenetic analyses of RNA editing in fern plastomes suggest possible adaptive innovations

Blake D. Fauskee[1], Li-Yaung Kuo[2], Tracy A. Heath[3], Pei-Jun Xie[2,4] and Kathleen M. Pryer[1]

[1]Department of Biology, Duke University, Durham, NC 27708, USA; [2]Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, 300044, Taiwan; [3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA; [4]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14850, USA

Author for correspondence:
Blake D. Fauskee
Email: blake.fauskee@duke.edu

Received: *26 February 2025*
Accepted: *6 May 2025*

New Phytologist (2025)
doi: 10.1111/nph.70244

Key words: adaptation, chloroplast, neutral evolution, phylogenetics, plastome, RNA editing.

Summary

- RNA editing in plant organelles is widely regarded as a neutral corrective mechanism, yet it persists as a complex, energetically costly process, requiring numerous nuclear-encoded pentatricopeptide repeat proteins. Ferns are the most diverse lineage of land plants that uniquely retain both cytidine-to-uridine (C-to-U) and uridine-to-cytidine (U-to-C) RNA editing in their plastomes, offering a powerful system to investigate the evolutionary forces shaping both editing types.
- Two distantly related fern lineages – Hymenophyllaceae and Vittarioideae (Pteridaceae) – each containing sister sublineages with contrasting evolutionary rates, were selected for comparative analysis. Genomic and transcriptomic data were combined with bioinformatic and phylogenetic methods to identify RNA editing sites and investigate the evolutionary dynamics of both C-to-U and U-to-C editing in fern plastomes.
- Nonsynonymous edits were frequently lost, consistent with neutral evolution. By contrast, C-to-U edits at start codons and U-to-C edits at internal stop codons were evolutionarily conserved, displaying lower and more variable editing efficiencies that suggest these edits are regulated.
- C-to-U edits at start codons and U-to-C edits at internal stop codons are evolutionarily conserved and exhibit signatures of selective regulation, suggesting that they function as molecular checkpoints. These findings provide the strongest evidence to date that RNA editing in plants plays an adaptive role in modulating plastid gene expression.

Introduction

Across the tree of life, the central dogma holds that DNA is faithfully translated into proteins via RNA intermediates (Crick, 1958). Eukaryotes, however, have evolved mechanisms that add complexity to this process. The discovery of intron splicing in 1977 (Chow *et al.*, 1977; Berget & Sharp, 1977) revealed that DNA does not always directly predict protein sequences but requires introns to be removed from pre-mRNA during processing. Other modifications, like 5′ capping and 3′ polyadenylation, can further refine the mRNA transcript, affecting stability and translation efficiency, but without altering the coding sequence itself. RNA editing, by contrast, results in targeted modifications to specific nucleotide sequences within an RNA molecule, giving rise to transcripts that differ from their original DNA templates in discrete, often functionally significant ways (reviewed in Knoop, 2011). These edits may include base substitutions, insertions, or deletions, allowing organisms to fine-tune RNA stability, localization, or protein function posttranscriptionally.

Diverse examples of RNA editing can be found across eukaryotes. In trypanosomes, for example, RNA editing involves extensive uridine insertions and deletions to produce functional mitochondrial transcripts (Benne *et al.*, 1986), while in animals and plants, targeted base substitutions – such as adenosine-to-inosine (A-to-I) and cytidine-to-uridine (C-to-U) changes – allow fine-tuning of RNA function and stability. In animals, RNA editing can provide adaptive advantages by generating functionally specialized proteins. For instance, cephalopods exhibit extensive A-to-I editing on neural transcripts, allowing them to adjust protein function in response to changing water temperatures (Birk *et al.*, 2023).

In plants, however, RNA editing presents a different picture. Although widespread in both the plastid and mitochondrial genomes, editing does not result in new protein functions but instead restores evolutionarily conserved amino acid sequences in the mRNA transcript to faithfully yield functional organellar proteins (reviewed in Small *et al.*, 2020). The number of RNA editing sites varies greatly across plant lineages. Seed plants typically have 30 to 50 C-to-U edits in the plastome (rarely, higher counts are reported), while the lycophyte *Selaginella uncinata* contains more than 3000 editing sites (Oldenkott *et al.*, 2014; Fig. 1). Ferns, the sister lineage to seed plants, range from having hundreds of

editing sites in species like *Adiantum shastense* to none in *Equisetum hyemale* (Fauskee *et al.*, 2021; Knie *et al.*, 2016; Fig. 1). Plant RNA editing is orchestrated by a subclass of nuclear-encoded pentatricopeptide repeat (PPR) proteins – one of the largest gene families in land plants (Auborg *et al.*, 2000; O'Toole *et al.*, 2008; Li *et al.*, 2018; Schafran *et al.*, 2025). Each PPR protein typically targets a single editing site or, in some cases, a few sites by binding to a specific RNA segment located upstream of the editing site (Barkan & Small, 2014). This highly specific targeting of RNA editing sites by PPR proteins underscores the complexity of the plant RNA editing system, requiring a vast and specialized network of nuclear-encoded factors to maintain organellar function.

The need for RNA editing to 'correct' RNA sequences raises a unique question: Is there an evolutionary advantage that it confers over simply encoding the conserved amino acids correctly at the DNA level? Plant RNA editing appears to be more of a maintenance mechanism than an adaptive response (Lukeš *et al.*, 2011). The complexity of this process in plants is amplified by the fact that PPR proteins each typically target a single editing site (a separate PPR protein for each editing site). For RNA editing to succeed, both the nuclear-encoded PPR proteins and their specific upstream RNA binding sites must be conserved, which surely must require a substantial genetic investment. Mutations in either the PPR protein or the RNA binding site could disrupt editing, resulting in failed functional protein translation. Rather than encode the correct amino acids directly in the DNA, plant organellar gene expression appears to rely on an intricate interdependency where each edit (numbering in the thousands in some species; Fig. 1) depends on a match between a dedicated PPR protein and its unique RNA binding site. Is RNA editing in plants an overly complicated wrinkle in the gene expression process with no clear adaptive benefit, and, if so, why does it persist?

To explain why the evolution of RNA editing in plant organellar genomes apparently lacks an obvious adaptive benefit, Covello & Gray (1993) proposed a neutral framework, later formalized by Stoltzfus (1999) as the 'Constructive Neutral Evolution' (CNE) hypothesis. The CNE hypothesis suggests that the capability to edit organellar RNA arose in the nuclear genome, likely through the co-option of existing nuclear PPR proteins. Mutations at editable sites in organellar genomes were then corrected by RNA editing and became fixed through genetic drift. This editing capability is hypothesized to be maintained by natural selection, while additional RNA edits accumulate in the organellar DNA through a 'neutral evolutionary ratchet' (Covello & Gray, 1993; Lukeš *et al.*, 2011). Under this hypothesis, RNA editing quickly becomes entrenched, because losing RNA editing entirely would require each editing site to back-mutate to its pre-edited state – a highly improbable scenario once multiple RNA editing sites are established. Although CNE does not rule out potential adaptive benefits, it does not depend on them either. Importantly, these hypotheses have been largely based on seed-plant observations, where only C-to-U RNA editing is present, and very little is known about the evolution of U-to-C RNA editing in plants. Interestingly, recent research has shown that some RNA editing sites, particularly those that reinstate the start codon for some genes, appear to be developmentally regulated (Miyata & Sugita, 2004; Bentolila *et al.*, 2013; Li *et al.*, 2018) leading to the suggestion that RNA editing may act as a regulatory switch for organellar gene expression (Li *et al.*, 2018).

The central aim of this study was to investigate whether plant RNA editing might confer a selective advantage. We focus here on fern plastomes because ferns, in contrast to their sister group, the seed plants, not only exhibit higher levels of C-to-U RNA editing but also possess U-to-C editing, which is absent in seed plants (Fig. 1). This allows us to explore whether there are evolutionary pressures that differ between C-to-U and U-to-C editing, as has been suggested by Fauskee *et al.* (2021). Given the rapid evolution of RNA editing sites in fern plastomes (Guo *et al.*, 2015; Fauskee *et al.*, 2021), we mitigate potential saturation effects by focusing on two fern lineages, each with sister sublineages that exhibit significant substitution rate heterogeneity: the Vittarioideae subfamily within Pteridaceae (Rothfels & Schuettpelz, 2014) and the Hymenophyllaceae (Schuettpelz & Pryer, 2006). Within Vittarioideae, the genus *Adiantum* is sister to a clade of epiphytic ferns, the vittarioids, that have undergone a marked substitution rate acceleration relative to *Adiantum* (Rothfels & Schuettpelz, 2014). A similar pattern is exhibited in Hymenophyllaceae, where the sister subfamily Trichomanoideae displays a much higher substitution rate than the Hymenophylloideae (Schuettpelz & Pryer, 2006). By expanding the available data for these two lineages, we report here on the plastid RNA editing sites for 10 *Adiantum* species and eight vittarioids in Vittarioideae, and for four species in Hymenophylloideae and seven in Trichomanoideae within Hymenophyllaceae.

We employ a novel integrative approach, combining genomic and transcriptomic data with several rigorous phylogenetic methods to investigate whether selective pressures may be influencing the evolution of RNA editing in plastomes. To address these questions, we more than triple the available data on plastid RNA editing in ferns. In addition, because little is known about U-to-C RNA editing in natural systems, our work provides a significant opportunity to understand this unique process. Utilizing an integrative analysis framework, we explore whether the evolutionary patterns of RNA editing sites (both C-to-U and U-to-C) in ferns are consistent with expectations under CNE or whether certain subsets of edits may have been co-opted for adaptive roles. By examining RNA editing within a comparative phylogenetic context, we aimed to shed light on an atypical evolutionary phenomenon that has puzzled researchers for decades.

Materials and Methods

Sampling, DNA extraction and sequencing, plastome assembly, and annotation

Genomic data were collected from frozen or silica-dried tissue for 10 *Adiantum* species, eight vittarioids, seven species of Trichomanoideae, and four from Hymenophylloideae (Supporting Information Table S1). Most of these data were newly generated as part of this study, but some plastome assemblies were leveraged from available data (but only when the specimen vouchers were

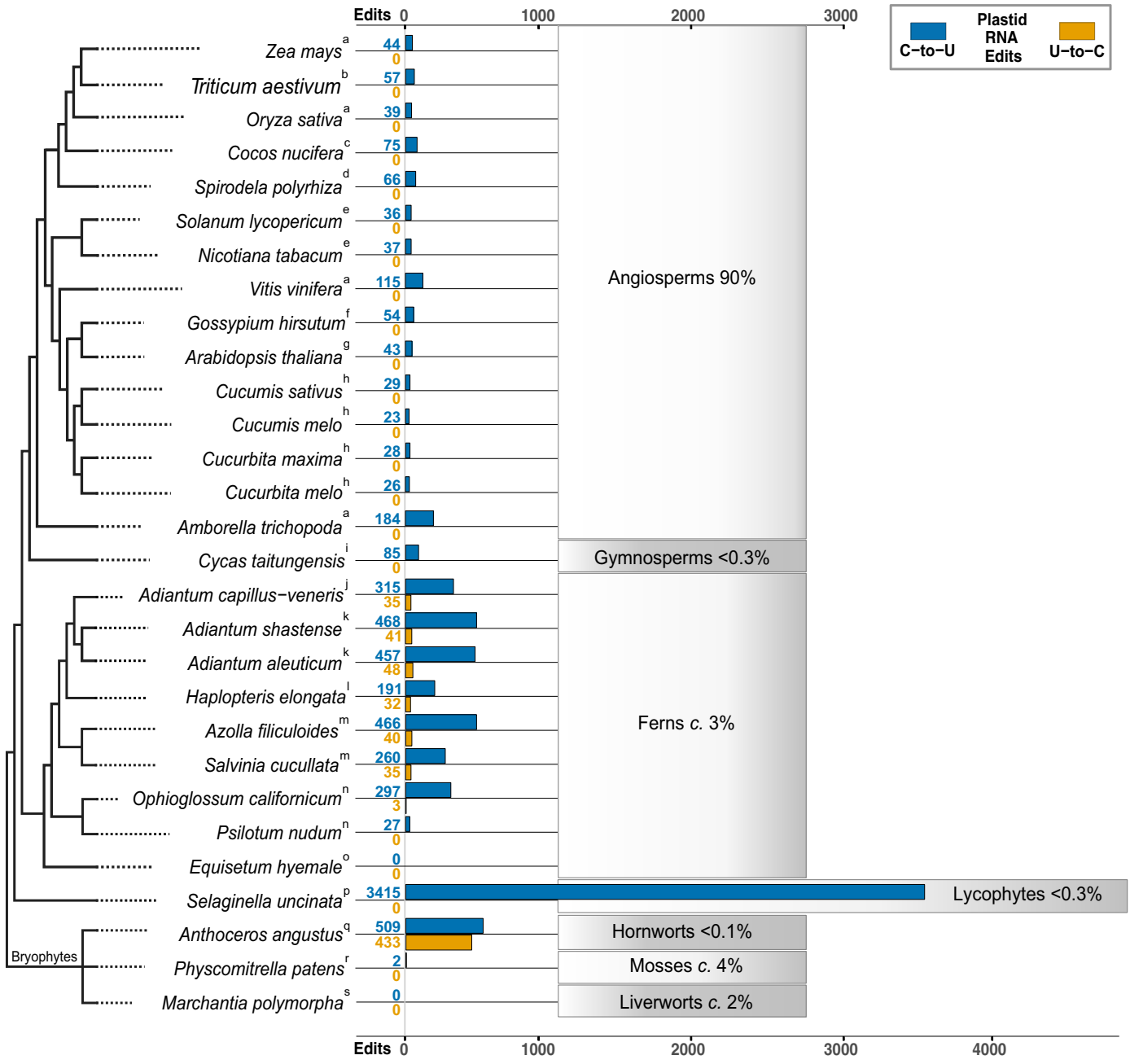

Fig. 1 Number of plastid RNA editing counts reported from across a representative selection of land plants. Phylogenetic relationships drawn to match previously published studies. Blue bars/numbers indicate the number of C-to-U) edits, and orange bars show the number of U-to-C edits reported for each taxon. Major land plant clades are outlined to the right of the phylogeny, each indicating a rough percentage estimate they represent of the entire *c.* 500 000 species that comprise land plants. Superscript letters following taxon names denote publications that report the RNA edit numbers for each taxon shown here: (a) Ishibashi *et al*. (2019), (b) Mirzaei *et al*. (2018), (c) Huang *et al*. (2013), (d) Wang *et al*. (2015), (e) Kahlau *et al*. (2006), (f) Jiang *et al*. (2012), (g) Ruwe *et al*. (2013), (h) Guzowska-Nowowiejska *et al*. (2009), (i) Chen *et al*. (2011), (j) Wolf *et al*. (2004), (k) Fauskee *et al*. (2021), (l) This study, (m) Li *et al*. (2018), (n) Guo *et al*. (2015), (o) Knie *et al*. (2016), (p) Oldenkott *et al*. (2014), (q) Kugita *et al*. (2003), (r) Miyata & Sugita (2004), and (s) Rüdinger *et al*. (2008).

the same as for our transcriptomic data, described later). Namely, the plastome assemblies of *Antrophyum semicostatum* Blume, *Scoliosorus ensiformis* (Hook) T. Moore, *Vaginularia junghuhnii* Mett., and *Vittaria graminifolia* Kaulf. were previously published by Robison *et al.* (2018). For all other samples, DNA was extracted using the E.Z.N.A. SP Plant and Fungal DNA Extraction Kit from Omega Bio-Tek (D5511, Norcross, GA, USA).

DNA was quantified on the Qubit 2 Fluorometer (Thermo Fisher Scientific Inc., Walden, MA, USA) using the Qubit dsDNA High Sensitivity Quantification Assay Kit (Q32851;

Thermo Fisher Scientific Inc.). DNA libraries for Illumina sequencing were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645; New England Biosciences, Ipswitch, MA, USA) following the manufacturer's default protocol for 200 base pair insert sizes. To enable multiplexing, each sample was given a unique barcode using the NEBNext Multiplex Oligos for Illumina (E6609; New England Biosciences). Libraries were then sequenced on the Illumina Novaseq X Plus by Novogene Co., Ltd (Beijing, China) using 150 base pair, paired-end chemistry. Demultiplexing was also performed by Novogene Co., Ltd.

Paired-end DNA reads were then uploaded to the Duke Compute Cluster (Duke University, Durham, NC, USA) where adapters and low-quality reads were trimmed using TRIMMOMATIC v.0.39 (Bolger *et al.*, 2014) with the following settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Plastomes were then assembled *de novo* from the trimmed reads using NOVOPLASTY v.4.3.3 (Dierckxsens *et al.*, 2017). *rbcL* sequences from closely related individuals were obtained from GenBank and used as seeds for the assemblies (Table S1). Draft plastome assemblies were then polished using PILON v.1.24 (Walker *et al.*, 2014). Plastomes were iteratively polished for two rounds or until Pilon suggested no further changes. Polished plastomes were then annotated in GENEIOUS v.2022.0.2 (Kearse *et al.*, 2012) using the BLAST (Altschul *et al.*, 1990) plugin, followed by manual adjustment. Here, annotations from previously published and closely related fully assembled plastomes were used as reference (Table S1).

Transcriptome sequencing

Transcriptome data were generated from the same specimen vouchers from which plastome sequences were generated (Table S1). RNA was extracted from green leaf tissue that was either fresh or flash-frozen. The RNA extractions were performed using the E.Z.N.A. Plant RNA Kit from Omega Bio-Tek (R6827, Norcross, GA, USA) or the CTAB-column approach (RNA spin columns from Spectrum Total Plant RNA Kit; Millipore Sigma, Darmstadt, Germany) described in Pelosi *et al.* (2024). During RNA extraction, samples were treated with Millipore Sigma DNase I to reduce DNA contamination (69 182; Millipore Sigma). The resulting RNA extractions were then quantified on the Qubit 2 Fluorometer (Thermo Fisher Scientific Inc) using the Qubit RNA High Sensitivity Quantification Assay Kit (Q32852; Thermo Fisher Scientific Inc).

RNA (cDNA) libraries were then constructed using one of two library preparation kits: the Zymo-Seq RiboFree Total RNA Library Kit (R3000; Zymo Research, Irvine, CA, USA) or the NEBNext Ultra II Library Prep (E7775; New England Biosciences) with ribosomal depletion probes designed for plant samples supplied by New England Biosciences as part of a beta test agreement. Library preps for both kits were done following suggested manufacturers' protocols. In each library preparation, unique barcodes were tagged to each sample to enable multiplexing. The resulting libraries were then sequenced by Novogene Co., Ltd (Beijing, China) and Genomics Corp. (New Taipei, Taiwan) using 150 base pair, paired-end chemistry on the Illumina HiSeq or the Illumina NovaSeq X Plus or HiSeq X Ten. Novogene and Genomics Corp. conducted all the library quality control, data quality control, and demultiplexing. Newly generated transcriptomic and genomic data for this project are available from the Sequence Read Archive (SRA), BioProject no.: PRJNA1216602 (specific BioSample and SRA accession numbers in Table S1).

RNA editing detection

Protein-coding genes were extracted from plastome assemblies together with 100 base pairs flanking the beginning and end of each gene sequence. For each species, all relevant protein-coding genes (and flanking sequences) were placed into one multi-fasta file. A custom RNA editing detection pipeline, modified from Edera & Sanchez-Puerta (2021) was constructed to identify putative RNA editing sites. An example script is available on GitHub (https://github.com/bfauskee/fauskee-fern-rna-editing-scripts). Briefly, RNA reads were trimmed using TRIMMOMATIC v.0.39 (Bolger *et al.*, 2014) with the following settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 to remove adapters and low-quality reads. RNA reads were then trimmed again in TRIMMOMATIC (Bolger *et al.*, 2014) in single-end mode to remove the first 13 bases in each read, where the GC content was nonuniform (HEADCROP:13). For each species, trimmed RNA reads were then mapped to the multi-fasta file containing the plastid protein-coding gene sequences using BOWTIE2 v.2.2.4 (Langmead & Salzburg, 2012). SAMTOOLS v.1.14 (Li *et al.*, 2009) was used to construct a BAM file that was used as input to BAM-READCOUNT v.0.8.0 (Khanna *et al.*, 2022) along with custom Linux commands that were then used to count the total number of RNA reads mapped to each site in each gene and the number of reads with each nucleotide mapped to each site. The output of this step is a TSV file showing the reference base, the number of reads mapped to the reference base, and the number of mapped RNA reads displaying each of the four nucleotides.

A custom R pipeline was developed to detect and characterize each RNA editing site and is available on GitHub (https://github.com/bfauskee/fauskee-fern-rna-editing-scripts). A site was recognized as an RNA editing site if it met the following criteria: At least 10 reads mapped to the site and at least three reads and 10% of the total mapped RNA reads displayed the edited base (a T mapped to a reference C for a C-to-U edit and vice versa for a U-to-C edit). Manual inspection was performed in rare regions with very low RNA read coverage. This R pipeline also outputs the type of edit (C-to-U vs U-to-C), its position, its codon position, the induced amino acid change, the editing efficiency, and several other features. RNA editing efficiency was calculated by dividing the number of mapped reads displaying the edited base (e.g. a T for a C-to-U edit) by the total number of mapped RNA reads.

Evolutionary and phylogenetic analyses of RNA editing

To investigate how RNA editing sites are conserved across species in each independent dataset (Vittarioideae and Hymenophyllaceae),

we created sequence alignments for each protein-coding gene using MAFFT v.7.505 (Katoh & Standley, 2013). Here, only the coding sequences were aligned; flanking regions and introns were excluded. Alignment position was then used to compare the presence or absence of each unique RNA editing site across all species within each dataset.

For each dataset, we estimated a maximum likelihood (ML) phylogeny from a concatenated supermatrix of all plastid protein-coding genes. This totaled 87 genes for Hymenophyllaceae and 85 genes for Vittarioideae, which lack *psaM* and *ycf66* that are present in Hymenophyllaceae. Based on previously published phylogenetic topologies across all ferns (Testo & Sundue, 2016; Nitta *et al.*, 2022), one outgroup for each dataset was also included: *Osmundastrum cinnamomeum* (L.) C. Presl for Hymenophyllaceae and *Myriopteris lanosa* (Michx.) Grusz & Windham for Vittarioideae. Gene alignments were carried out again to include outgroups using MAFFT v.7.505 (Katoh & Standley, 2013), and alignments were then concatenated using AMAS (Borowiec, 2016) and partitioned by gene. Maximum likelihood tree estimation was carried out in IQ-TREE2 v.2.2.2.7 (Minh *et al.*, 2020) with 1000 ultrafast bootstrap replicates. Substitution model selection and optimal partitioning schemes were automated using the MFP + MERGE option in IQ-TREE2, and the Bayesian information criterion (BIC) was the metric used to select the best models and partitioning schemes. Each tree was then rooted with the outgroups described above, and the outgroups were then pruned.

A phylogenetic generalized least squares (PGLS) analysis was conducted to explicitly test the relationship between the number of RNA edits of each type (C-to-U and U-to-C) and the branch length, expressed as the root-to-tip distance. Root-to-tip distance was extracted for each species from the ML plastome trees using the `distRoot` function in the ADEPHYLO R package (Jombart *et al.*, 2010). The PGLS analysis was then conducted in R using the following packages: APE (Paradis & Schliep, 2019), GEIGER (Pennell *et al.*, 2014), NLME (Pinheiro *et al.*, 2024), and PHYTOOLS (Revell, 2012).

For genes whose best-fitting nucleotide substitution model, as determined by BIC, contained an among-site rate heterogeneity parameter, site-specific substitution rates were calculated in IQ-TREE2 v.2.2.2.7 (Minh *et al.*, 2020). Using custom scripts, the rate category assigned to each RNA editing site was extracted. Rate categories were used to enable comparisons across genes. Rate categories for each unique edit were then color-coded to distinguish whether they were a nonsynonymous C-to-U edit, a nonsynonymous U-to-C edit, an edit that restores a start codon, or an edit that corrects an internal stop codon, and were plotted using the GGWAFFLE R package (Gilbey, 2022).

RevBayes estimation of RNA editing site transition rates

Using sequence alignments, binary presence/absence matrices were generated to catalog RNA editing sites across species in each dataset. Data were divided into four subsets: nonsynonymous C-to-U edits, nonsynonymous U-to-C edits, C-to-U edits restoring start codons, and U-to-C edits correcting internal stop codons. These matrices were then converted into NEXUS files using the APE package in R (Paradis & Schliep, 2019). The ML plastid phylogenies for each dataset, with outgroups removed, were transformed into ultrametric trees – where all extant taxa are equidistant from the root – using the `chronos` function in the ape R package (Paradis & Schliep, 2019). This ensures that the branch lengths represent relative time rather than genetic change, providing a starting tree for molecular clock-based analyses.

A combined analysis of molecular and RNA editing characters was conducted to estimate branch-specific rates of evolution on a fixed ML tree topology. The RNA editing site data consisted of nonsynonymous C-to-U RNA editing sites, while the molecular data comprised the alignment of the plastid gene *rbcL*, which is thought not to contain any RNA editing sites in most fern lineages and was confirmed to lack such sites in our datasets. All analyses were conducted in REVBAYES v.1.2.3 (Höhna *et al.*, 2016).

RNA editing data were modeled using a continuous-time Markov chain model with a Jukes–Cantor transition matrix (`fnJC(2)`) to specify equal substitution probabilities (Lewis, 2001). Among-site rate heterogeneity was accounted for by discretizing a gamma distribution with four rate categories, where the shape parameter was drawn from an exponential prior with a rate parameter of 1.0 corresponding to a mean of 1.0 (`dnExponential(1.0)`). RNA editing site branch rates were modeled using an uncorrelated lognormal (UCLN) relaxed clock (Drummond *et al.*, 2006). The mean and SD of the lognormal distribution were assigned exponential priors (`dnExponential(2.0)` for the mean, `dnExponential(3.0)` for the SD), and individual branch rates were sampled across the tree. RNA editing site data were clamped to this model.

We used the general time-reversible (GTR) model (Tavaré, 1986) with gamma-distributed rate variation across sites (Yang, 1994) and a proportion of invariant sites (Sidow *et al.*, 1992; Steel *et al.*, 2000; GTR + I + G) to model molecular evolution. Exchangeability rates and equilibrium frequencies were assigned Dirichlet priors (`dnDirichlet(1,1,1,1)`). The shape parameter for the gamma distribution was drawn from an exponential prior with a rate parameter of 1.0, corresponding to a mean of 1.0 (`dnExponential(1.0)`). The proportion of invariant sites was modeled as a probability value ranging from 0 to 1, drawn from a beta distribution with alpha and beta shape parameters both set to 1 (`dnBeta(1,1)`). This configuration creates a uniform distribution across the interval (0,1). The Beta (1,1) distribution reflects maximum uncertainty about the proportion of invariant sites, allowing the data to determine the final estimate. Branch-specific molecular rates were modeled under a UCLN clock, with the mean and SD of the lognormal distribution assigned exponential priors (`dnExponential(2.0)` and `dnExponential(3.0)`, respectively). The *rbcL* molecular dataset was clamped to this model.

A constant-rate birth-death process (Kendall, 1948; Nee *et al.*, 1994; Gernhard, 2008) was used to model the diversification dynamics of the phylogeny. Speciation and extinction rates were assigned exponential priors (`dnExponential(10)`), and the probability of sampling extant taxa (rho) was fixed at 0.03, roughly reflecting the proportion of sampled species in the

study. An arbitrary prior root age of 1 was assigned because our goal was not to date these trees, but to estimate informative branch lengths in relative time units for comparative analyses. The input ML tree topology was fixed, and only node ages were estimated using a time-scaling move. We ran two independent Markov chain Monte Carlo (MCMC) runs, each running for 2000 000 generations with 227 moves performed each generation, to estimate posterior distributions of model parameters. Trees were sampled every 100 generations and logged to a file for downstream analyses. Convergence was assessed using the effective sample sizes of parameters and visual inspection of trace plots in Tracer v.1.7.1 (Rambaut *et al.*, 2018).

Following the MCMC analysis, we summarized the posterior distribution of trees for each dataset by calculating a maximum clade credibility (MCC) tree. This tree was obtained from the posterior tree trace and represents the topology with the highest posterior probability of its clades. Node ages in the MCC tree were summarized using the mean heights from the posterior distribution. The resulting tree, now containing branch rates estimated by our models, was used for downstream estimates of RNA editing site transition rates.

We estimated transition rates for each of the four RNA editing subsets (nonsynonymous C-to-U edits, nonsynonymous U-to-C edits, C-to-U edits restoring start codons, and U-to-C edits correcting internal stop codons) on the MCC trees generated from our branch rate analysis. These trees, estimated in RevBayes using birth-death and relaxed-clock models, have biologically meaningful branch lengths such that the distance between two nodes represents the relative divergence time. For each dataset, two models were tested: a free-rate model (FreeK) and an equal-rate model.

In the FreeK model, the rates of gain and loss of RNA editing sites were allowed to differ, with each transition rate drawn independently from an exponential prior. A rate matrix was constructed from these rates, and root state frequencies were drawn from a Dirichlet prior. In the equal-rate model, all transition rates were constrained to be equal, and a single rate parameter was drawn from the same exponential prior. RNA editing matrices were clamped to the fixed MCC trees for each dataset. Both models were run twice independently using MCMC for 1000 000 generations, sampling parameters every 100 generations. Convergence was assessed by visual inspection of trace plots in Tracer v.1.7.1 (Rambault *et al.*, 2018). Representative RevBayes scripts for all the above analyses are available on GitHub (https://github.com/bfauskee/fauskee-fern-rna-editing-scripts).

To compare the FreeK and equal-rate models for each dataset, we performed Bayes factor analyses (Jeffreys, 1961) using both stepping-stone sampling (Fan *et al.*, 2011; Xie *et al.*, 2011) and path sampling (Lartillot & Philippe, 2006; Baele *et al.*, 2012) to estimate marginal likelihoods. Stepping-stone sampling approximates the marginal likelihood by incrementally raising the posterior distribution to fractional powers (power posteriors) and integrating over these intermediate distributions. Path sampling follows a similar logic but uses a continuous path through the power posteriors to calculate the marginal likelihood. We applied both approaches in RevBayes, using 63 power posterior distributions with 1000 MCMC generations per power posterior after an initial burn-in of 2000 generations. Marginal likelihoods were calculated as log values for both models. The Bayes factor was then computed as the exponentiated difference between the log-marginal likelihoods of the FreeK and equal-rate models. A Bayes factor close to 1 indicates no strong evidence favoring one model over the other (Jefferys, 1961). A Bayes factor much > 1 suggests that the FreeK model provides a better fit to the data, while a Bayes factor much < 1 suggests the equal-rate model is more consistent with the data. Consistent marginal likelihood estimates between stepping-stone and path sampling were assessed to confirm the robustness of the Bayes factor calculations.

Results

Phylogenetic inference of Hymenophyllaceae and Vittarioideae

Our ML phylogenetic estimates were based on concatenated plastid supermatrices of 85 protein-coding genes shared across 18 Vittarioideae species and 87 genes shared across 11 Hymenophyllaceae species. We recovered robustly supported phylogenies for each lineage (Fig. 2) with topologies that are congruent with previous taxonomic studies (Schuettpelz *et al.*, 2016; Huiet *et al.*, 2018; Kuo *et al.*, 2018). Additionally, both phylogenies reveal a robust signal of nucleotide substitution rate heterogeneity, consistent with earlier findings for each of these lineages by Schuettpelz and Pryer (2006), Hymenophyllaceae, and Rothfels and Schuettpelz (2014), Vittarioideae, whereby species in the subfamily Trichomanoideae have longer branches than their sister subfamily Hymenophylloideae and vittarioids are found on longer branches than their sister *Adiantum* species (Fig. 2).

Plastid RNA editing analyses

Notable trends were observed in RNA editing patterns across both fern clades (Fig. 2). Within Hymenophyllaceae, the number of C-to-U edits is higher in the slower evolving Hymenophylloideae (HYM), ranging from 155 to 124 (*Hymenophyllum fujisanense* Nakai– *H. holochilum* (Bosch) C. Chr.), whereas the faster Trichomanoideae (TRI) show a lower number of C-to-U edits, ranging from 115 to 44 (*Abrodictyium obscurum* (Blume) Ebihara & K. Iwats.–*Didymoglossum tahitense* (Nadeaud) Ebihara & K. Iwats.). The lower number of U-to-C edits across the Hymenophyllaceae overlap, ranging from 69 to 48 in the Hymenophylloideae ((*H. pallidum* (Blume) Ebihara & K. Iwats.–*H. holochilum*) and from 59 to 32 in the Trichomanoideae (*A. obscurum*–*D. tahitense*). In Vittarioideae, the number of C-to-U edits is higher in the slower evolving *Adiantum* (ADI), ranging from 539 to 299 (*A. shastense* Huiet & A.R. Sm.–*A. davidii* Franch.), whereas the faster vittarioids (VIT) consistently show a lower number of C-to-U edits, ranging from 221 to 182 (*Haplopteris yakushimensis* C.W. Chen & Ebihara–*V. junghuhnii*). The lower number of U-to-C edits across the Vittarioideae is very similar, ranging from 43 to 31 in *Adiantum* (*Adiantum caudatum* L. –*A. tenerum* Sw.) and from 39 to 29 in the vittarioids (*H. yakushimensis*–*Vittaria lineata* (L.) Sm.).

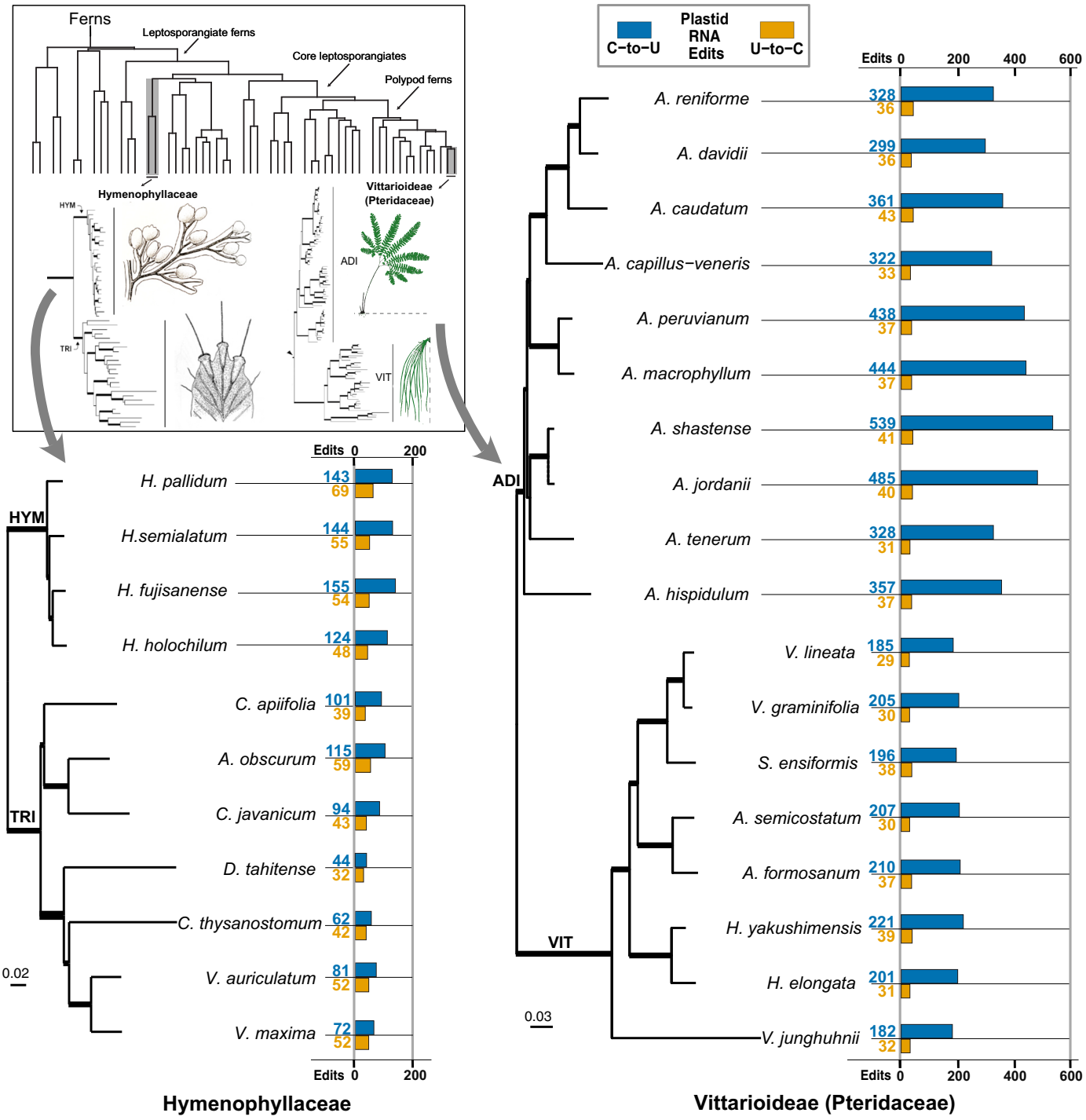

Fig. 2 Maximum likelihood phylogenetic estimates for Hymenophyllaceae and Vittarioideae (Pteridaceae) showing the distinct rate heterogeneity (slow vs fast) between the sister lineages within each respective group: Hymenophylloideae (slow; HYM) and Trichomanoideae (fast; TRI); *Adiantum* (slow; ADI) and the vittarioids (fast; VIT). Plastid RNA editing counts are represented as blue bars/numbers for cytidine-to-uridine (C-to-U) edits and orange bars/numbers for U-to-C edits. Trees in the top left corner show the phylogenetic position of each group in the broader fern topology and the rate heterogeneity patterns previously reported for Hymenophyllaceae (Schuettpelz & Pryer, 2006) and Vittarioideae (Rothfels & Schuettpelz, 2014). Branch lengths are in substitutions per site.

In both Hymenophyllaceae and Vittarioideae, C-to-U edits primarily impacted second-codon positions, while U-to-C edits mostly occurred at first-codon positions, and a small proportion of both editing types acted on third-codon positions (Fig. 3). Most C-to-U edits resulted in nonsynonymous amino acid changes, with a small proportion of these edits resulting in

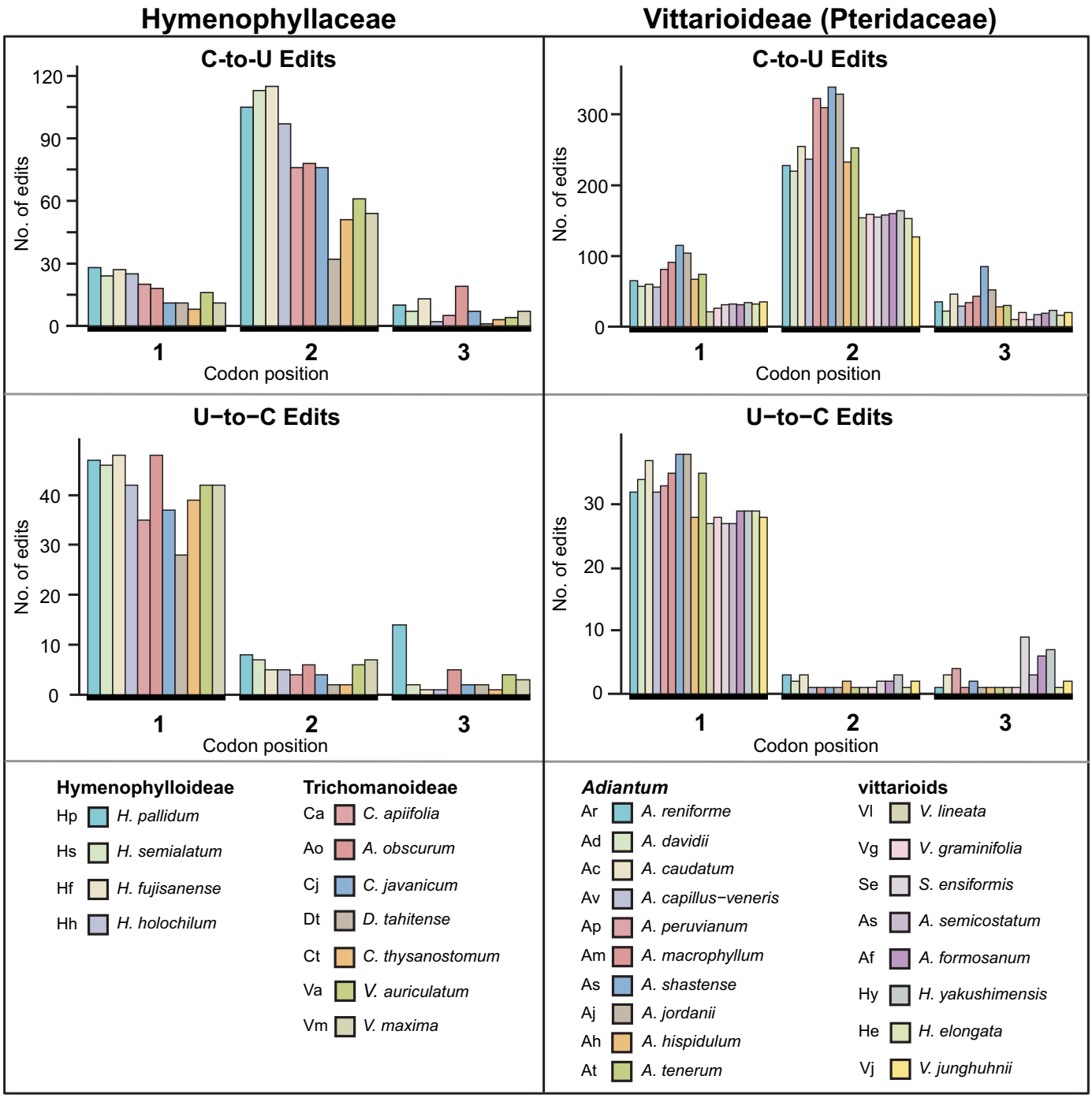

Fig. 3 RNA editing by codon position in Hymenophyllaceae (left) and Vittarioideae (right). Colored bars depict the number of edits for each species at each codon position. Species names are abbreviated at the bottom. Cytidine-to-uridine (C-to-U) edits are shown on the top and U-to-C edits are shown on the bottom.

restoring a start codon (Fig. 4). Most of the U-to-C edits corrected an internal stop codon to a sense codon (usually an arginine or glutamine) in each species studied here (Fig. 4). Relatively fewer U-to-C edits resulted in nonsynonymous amino acid changes, and even fewer in synonymous changes (Fig. 4). U-to-C edits that corrected internal stop codons occurred primarily in the 5′ half of the transcript sequence, whereas other U-to-C edits were more evenly distributed across the transcript (Fig. 4). For both datasets, a Mann–Whitney test revealed that U-to-C edits correcting internal stop codons to sense codons are significantly biased toward the 5′-end of the transcript compared with both nonsynonymous and synonymous edits, whereas no significant difference in transcript position was observed between nonsynonymous and synonymous edits (Fig. 4).

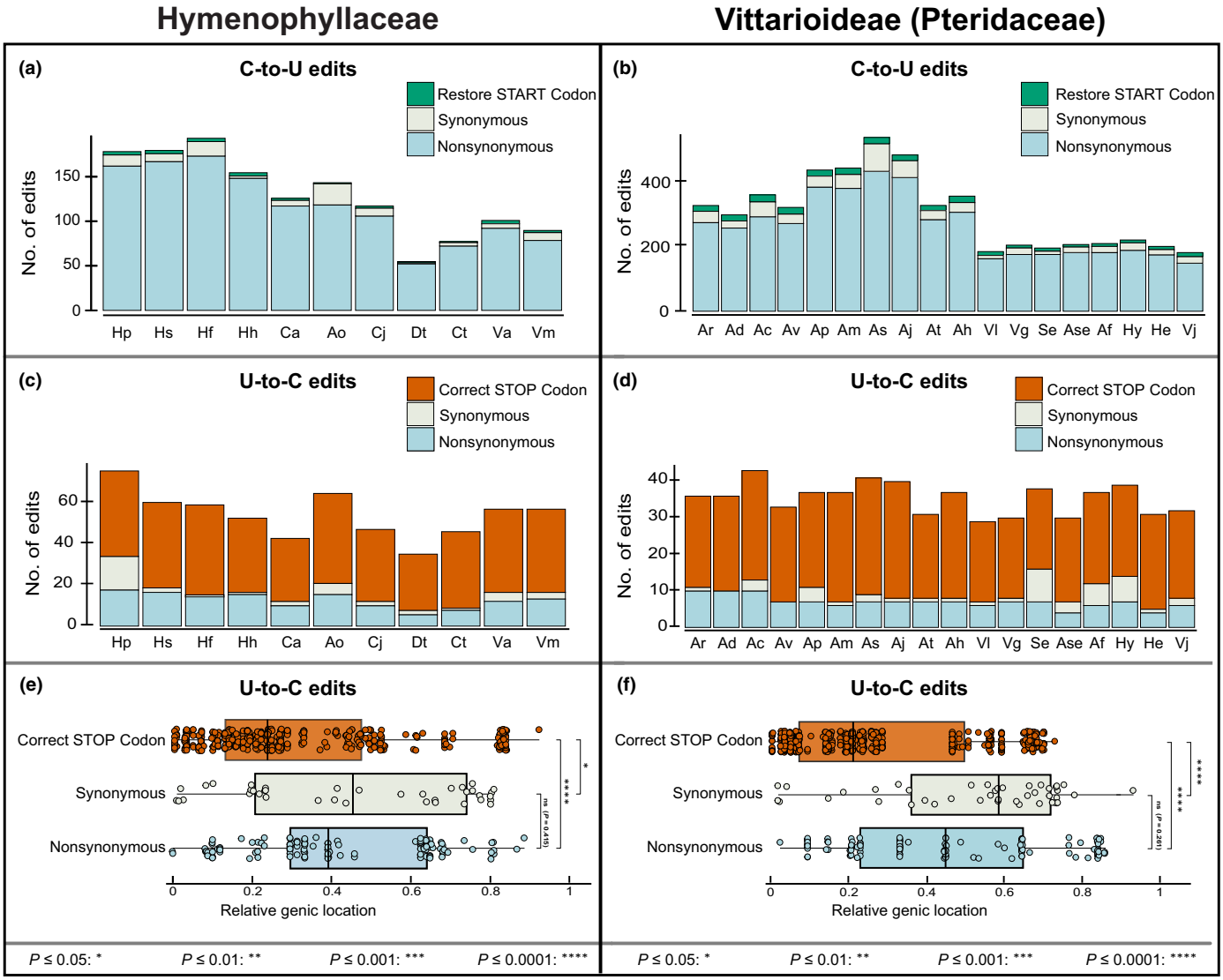

Fig. 4 Proportion of RNA editing sites resulting in nonsynonymous (light blue) and synonymous (gray) codon changes, as well as cryptic start codon restoration (green, cytidine-to-uridine (C-to-U) only) and internal stop codon correction (orange, U-to-C only). Hymenophyllaceae species are shown on the left and Vittarioideae on the right. The top panels (a, b) show the proportions of C-to-U edits by codon change type, while the middle panels (c, d) show the proportions of U-to-C edits. The bottom panels (e, f) display the relative genic locations of U-to-C edits, grouped by codon change type. Middle lines in each box represent the median, while the left-most and right-most lines represent the 25th and 75th percentiles, respectively. Whiskers show the lowest and highest values within 1.5× the interquartile range. Significant differences in relative genic locations between codon change classes are marked with asterisks denoting increasing *P*-value thresholds, and nonsignificant *P*-values are reported. Species abbreviations are as follows: Ac, *A. caudatum*; Ad, *A. davidii*; Af, *A. formosanum*; Ah, *A. hispidulum*; Aj, *A. jordanii*; Am, *A. macrophyllum*; Ao, *A. obscurum*; Ap, *A. peruvianum*; Ar, *A. reniforme*; As, *A. shastense*; Ase, *A. semicostatum*; At, *A. tenerum*; Av, *A. capullis-veneris*; Ca, *C. apiifolia*; Cj, *C. javanicum*; Ct, *C. thysanostomum*; Dt, *D. tahitense*; He, *H. elongata*; Hf, *H. fujisanense*; Hh, *H. holochilum*; Hp, *H. pallidum*; Hs, *H. semialatum*; Hy, *H. yakushimensis*; Se, *S. ensiformis*; Va, *V. auriculatum*; Vg, *V. graminifolia*; Vj, *V. junghuhnii*; Vl, *V. lineata*; Vm, *V. maxima*.

RNA editing efficiency (defined as the proportion of edited RNA transcripts) was calculated separately for each edit and species. In both datasets, synonymous edits occurred with a significantly lower efficiency than nonsynonymous edits, as did edits that restore start codons (C-to-U) or correct internal stop codons (U-to-C; Fig. 5). For U-to-C edits, internal stop codon edits occurred with a significantly lower efficiency than nonsynonymous edits in both datasets (Fig. 5). For C-to-U edits, start codon edits occurred with a significantly lower efficiency than nonsynonymous edits in Vittarioideae, but this difference was not significant in Hymenophyllaceae (Fig. 5).

Comparative phylogenetic analyses of RNA editing sites

A PGLS analysis was used to test the relationship between evolutionary rate (as the root-to-tip distance on the ML plastid phylogeny) and the number of RNA edits of each type (C-to-U and U-to-C). A significant negative relationship between the

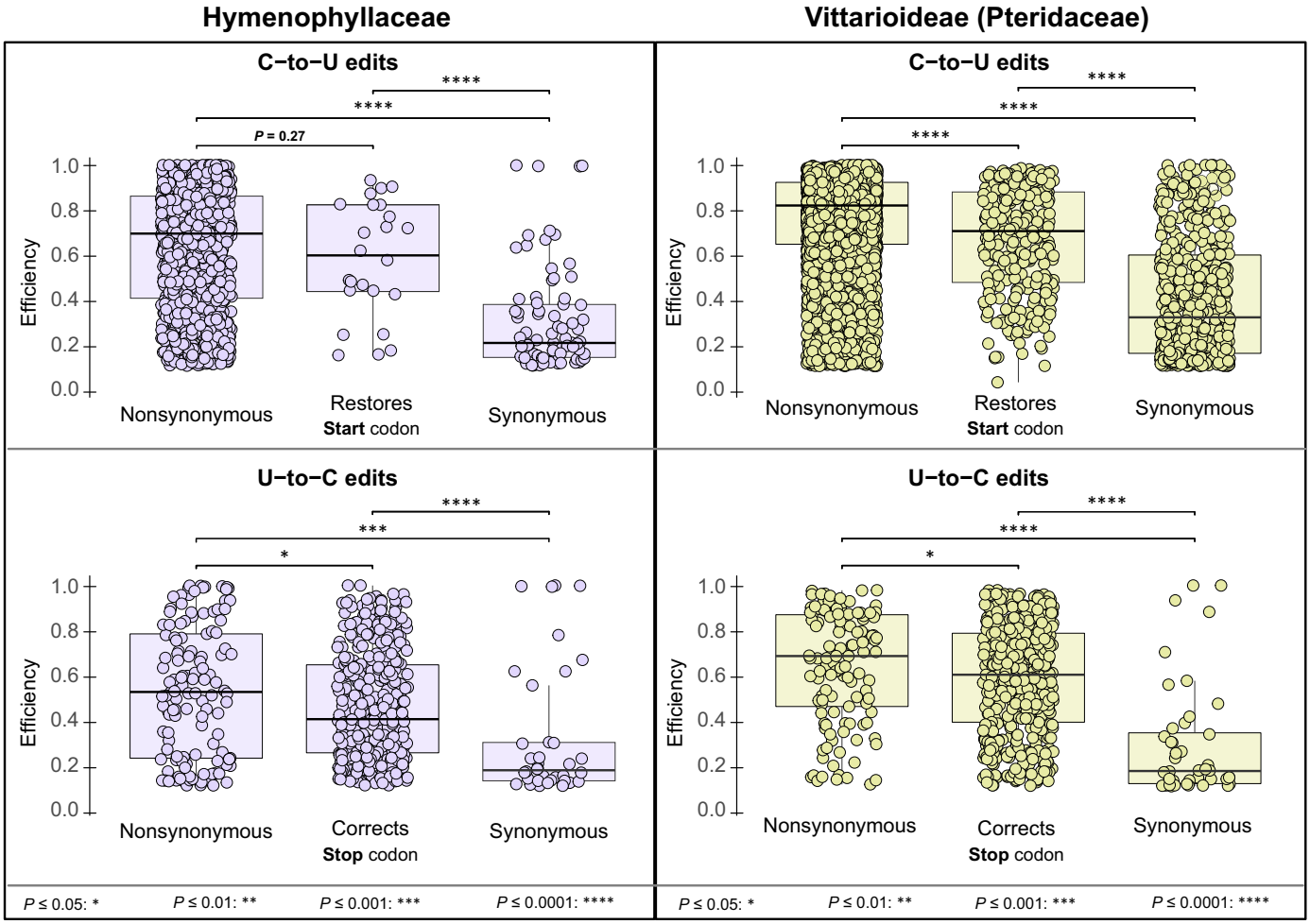

Fig. 5 Summary of RNA editing efficiencies across Hymenophyllaceae (left, lilac) and Vittarioideae (right, yellow). Efficiencies are grouped by codon change type. Cytidine-to-uridine (C-to-U) edits are shown on the top and U-to-C on the bottom. The middle lines in each box plot represent the median, while the top and bottom lines represent the 25th and 75th percentiles, respectively. The whiskers of each box plot extend to the highest and lowest values within 1.5× the interquartile range. Significant differences in efficiencies between codon change classes are marked with asterisks denoting increasing *P*-value thresholds. Nonsignificant *P*-values are reported.

root-to-tip distance and the number of C-to-U edits was detected in both fern lineages (Fig. 6; $P = 0.001$ for Vittarioideae and $P = 0.002$ for Hymenophyllaceae). A negative relationship was also observed between the number of U-to-C edits and root-to-tip distance in Hymenophyllaceae ($P = 0.043$); however, a similar correlation was not supported for Vittarioideae ($P = 0.42$; Fig. 6).

Sequence alignments for each gene were used to identify shared RNA editing sites across species within the independent datasets. To systematically compare the presence or absence of RNA editing sites across taxa, each RNA editing site was assigned a unique identifier by combining the gene name with the alignment position where an edit occurs. This unique tag allowed for a direct comparison between RNA editing site presence and absence across species. The distribution of shared editing sites was visualized using UpSet plots, providing a detailed overview of editing site conservation within each dataset. The analyses revealed that nonsynonymous C-to-U RNA editing sites were predominantly taxon-specific, with few edits shared across multiple species in either dataset (Fig. 7). Nonsynonymous U-to-C edits showed slightly higher conservation, particularly in Vittarioideae, but the majority were not shared by many taxa (Fig. 7). Notably, C-to-U edits that restore start codons and U-to-C edits that correct internal stop codons were widely shared across many species within each dataset (Fig. 7).

To investigate the rate at which RNA editing sites evolve, site-specific substitution rates were calculated for each site in genes where the best-fitting substitution model (according to BIC) contained an among-site rate heterogeneity parameter. Out of 85 genes for Vittarioideae, 73 had best-fitting models accommodating an among-site rate heterogeneity parameter, while 68 out of 87 genes for Hymenophyllaceae qualified. For applicable genes, the rate category for each unique edit site was plotted. Rate

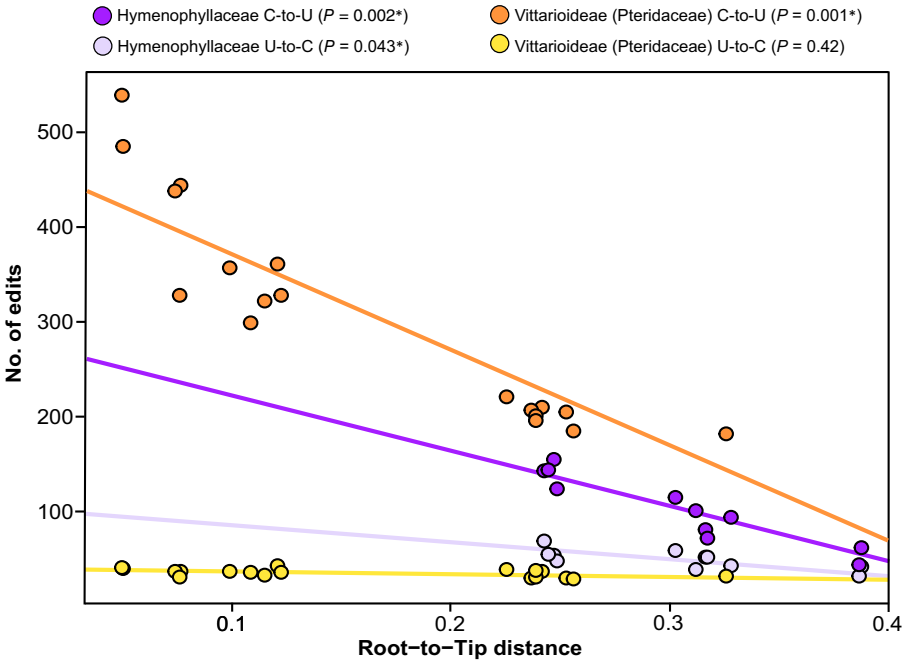

Fig. 6 Relationship between the number of RNA edits and the nucleotide substitution rate expressed as the root-to-tip branch length for each species. Vittarioids are shown in orange/yellow, and Hymenophyllaceae are shown in dark purple/light purple. Lighter colors represent uridine-to-cytidine (U-to-C) edits, and darker colors represent C-to-U edits. Trend lines are taken from a phylogenetic generalized least squares analysis. *P*-values are shown at the top of the figure, and significant *P*-values are indicated by asterisks.

categories were used to enable comparisons across genes within datasets. These analyses reveal that most RNA editing sites fall into high-rate categories (Fig. 8). However, a large proportion of U-to-C edits, both nonsynonymous edits and edits on internal stop codons, fall into low-rate categories (Fig. 8). Additionally, a significant proportion of C-to-U edits at internal stop codons occupy low-rate categories for Vittarioideae, while all three unique start codon edits (within genes with suitable models for this analysis) for Hymenophyllaceae occupy the highest rate category (Fig. 8).

Transition rates for RNA editing sites were estimated using comparative phylogenetic models in RevBayes (Höhna *et al.*, 2016). Both an equal-rate model (where rates of gain and loss are the same) and a two-rate (FreeK) model (where rates of gain and loss differ) were implemented and compared using Bayes factors. Our analyses showed that the two-rate model was preferred for nonsynonymous edits (both C-to-U and U-to-C) and internal stop codon edits (Fig. 9; Table S2). By contrast, the equal-rate model was favored for RNA edits on cryptic start codons. These patterns were consistent across both datasets. Nonsynonymous edits (both C-to-U and U-to-C) were found to be lost at a significantly higher rate than they were gained (Fig. 9). Start codon edits, however, were gained and lost at the same rate. By contrast, U-to-C edits at internal stop codons were gained at a significantly higher rate than lost, suggesting that these edits accumulate over evolutionary time.

Discussion

The role of RNA editing in plant organellar genomes has long been a fascinating evolutionary puzzle (Knoop, 2011; Lukeš *et al.*, 2011; Small *et al.*, 2020). Despite its high level of complexity, plant RNA editing has largely been viewed as a compensatory mechanism. This process enables the correction of specific deleterious mutations by restoring evolutionarily conserved amino acid sequences posttranscriptionally, without generating novel protein functions. The mutational buffer that RNA editing can provide may then allow these otherwise deleterious mutations to become fixed in populations through genetic drift. However, this raises critical questions as to why plants rely on an ostensibly overly complex and resource-intensive system for organellar gene expression. The scale of this investment is underscored in lineages such as ferns and hornworts, where up to 10% of expressed protein-coding genes encode for PPR RNA editing factors (Gutmann *et al.*, 2020), highlighting the substantial genetic and metabolic cost required to maintain this process. The CNE hypothesis (Covello & Gray, 1993; Lukeš *et al.*, 2011) provides a compelling explanation for how such a costly and complex system could become entrenched without a direct adaptive benefit, but is it truly the only explanation? Is RNA editing a purely neutral process or might it be more dynamic with many sites evolving neutrally, while some are evolutionarily conserved due to an as-yet undescribed adaptive function?

To explore these possibilities, we focused on two independent fern plastome datasets from the subfamily Vittarioideae (Pteridaceae) and the family Hymenophyllaceae. Ferns present a unique opportunity to investigate the evolutionary dynamics of plant RNA editing because, unlike seed plants, they retain both C-to-U and U-to-C editing–allowing for an explicit comparison of the evolutionary dynamics of both types of editing. Moreover, both fern lineages include sister subgroups with markedly different molecular evolutionary rates. This allows us to infer the directionality of RNA editing site evolution by comparing fast-evolving sublineages (vittarioids in Vittarioideae and

Trichomanoideae in Hymenophyllaceae) to their slower evolving sister taxa (*Adiantum* in Vittarioideae and Hymenophylloideae in Hymenophyllaceae). Specifically, we sought to determine whether certain edits were more evolutionarily conserved than others, potentially reflecting an adaptive significance that contrasts with the expectation that most RNA edits evolve neutrally.

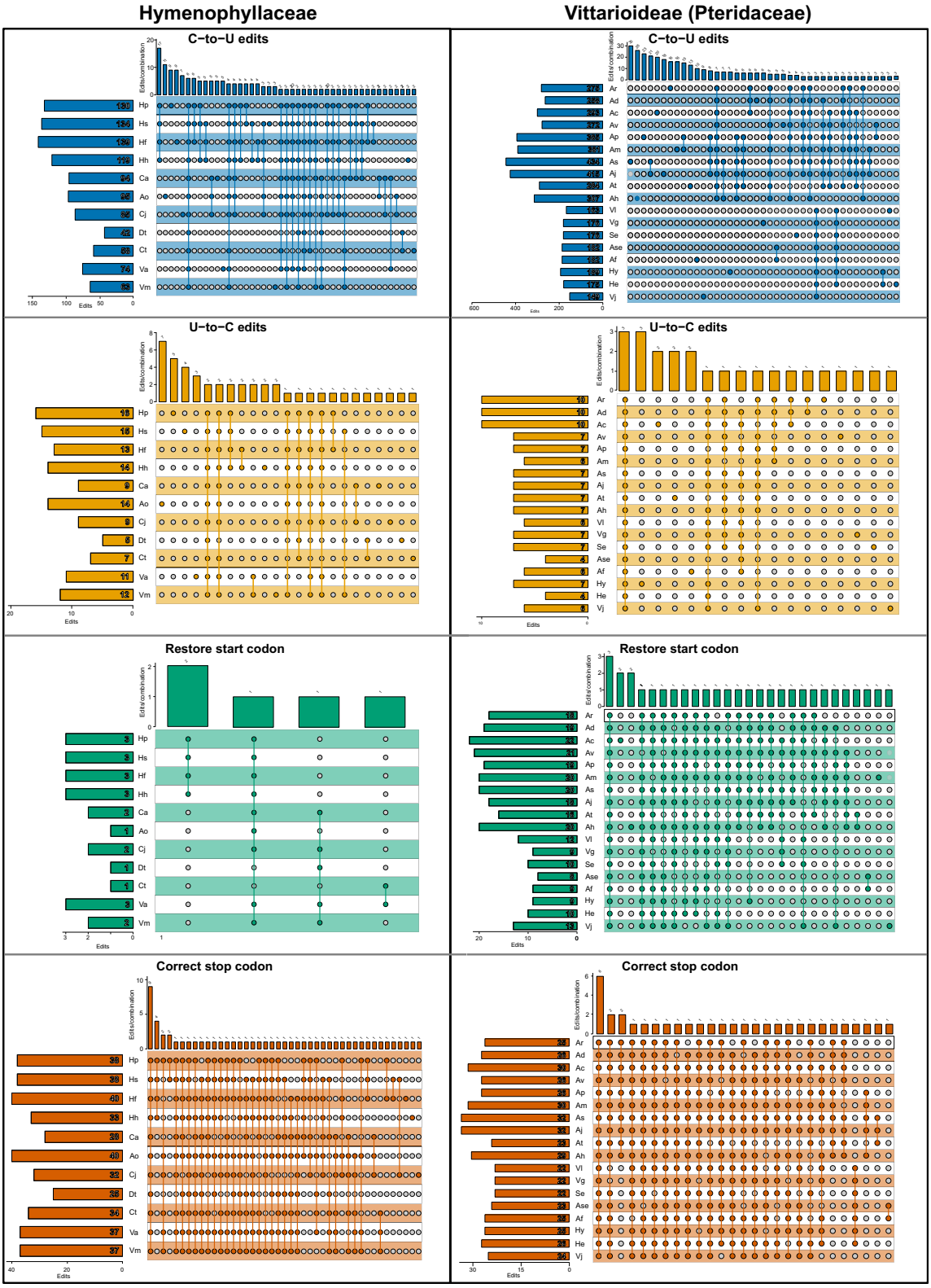

Fig. 7 Proportion of shared RNA editing sites for Hymenophyllaceae (left) and Vittarioideae (Pteridaceae; right). Nonsynonymous cytidine-to-uridine (C-to-U) edits are shown in blue (top panel), nonsynonymous U-to-C edits in yellow (second panel), C-to-U edits at start codons in green (third panel), and U-to-C edits at internal stop codons in dark orange (bottom panel). Horizontal bars to the left of species abbreviations show the number of edits present in each species. Vertical bars denote how many edits are shared by an exclusive group of taxa, defined by the connected dots. For Vittarioideae, sets of taxa sharing two or fewer edits are omitted, and sets of taxa sharing one or fewer edits are omitted for Hymenophyllaceae. Species abbreviations are as follows: Ac, *A. caudatum*; Ad, *A. davidii*; Af, *A. formosanum*; Ah, *A. hispidulum*; Aj, *A. jordanii*; Am, *A. macrophyllum*; Ao, *A. obscurum*; Ap, *A. peruvianum*; Ar, *A. reniforme*; As, *A. shastense*; Ase, *A. semicostatum*; At, *A. tenerum*; Av, *A. capillis-veneris*; Ca, *C. apiifolia*; Cj, *C. javanicum*; Ct, *C. thysanostomum*; Dt, *D. tahitense*; He, *H. elongata*; Hf, *H. fujisanense*; Hh, *H. holochilum*; Hp, *H. pallidum*; Hs, *H. semialatum*; Hy, *H. yakushimensis*; Se, *S. ensiformis*; Va, *V. auriculatum*; Vg, *V. graminifolia*; Vj, *V. junghuhnii*; Vl, *V. lineata*; Vm, *V. maxima*.

C-to-U and U-to-C RNA editing in plastids display distinct evolutionary patterns

Levels of C-to-U RNA editing in slow-evolving groups (*Adiantum* and Hymenophylloideae) far exceeded those of their faster evolving sister clades (Vittarioideae and Trichomanoideae, respectively; Fig. 2). Notably, some Hymenophyllaceae species possess among the lowest numbers of C-to-U edits known for ferns, particularly *Didymoglossum tahitense* (44 C-to-U edits) and *Crepidomanes thysanostomum* (Makino) Ebihara & K. Iwats. (62 C-to-U edits). While very early diverging eusporangiate fern lineages like *Equisetum* and *Psilotum* have even fewer C-to-U edits (0 and 27, respectively; Fig. 1), they lack U-to-C editing entirely (Guo *et al.*, 2015; Knie *et al.*, 2016). This positions Hymenophyllaceae as the leptosporangiate fern family with the lowest reported number of C-to-U edits across ferns that retain both editing types. U-to-C editing, however, followed a different pattern: U-to-C edit numbers remained relatively stable among the fast and slow-evolving groups within each dataset, but Hymenophyllaceae exhibited higher overall U-to-C editing levels than Vittarioideae (Fig. 2). Thus, it initially appears that C-to-U and U-to-C RNA editing may not evolve in concert, as the plastomes of Hymenophyllaceae retain far fewer C-to-U edits but significantly more U-to-C edits than Vittarioideae (Fig. 2).

Another indication that C-to-U and U-to-C edits are shaped by different evolutionary pressures is their distinct codon position biases, functional consequences, and variability across species. C-to-U edits primarily occur at second-codon positions (Fig. 3), producing mostly nonsynonymous amino acid changes, although a subset restores cryptic ACG start codons to AUG (Fig. 4). By contrast, U-to-C edits occur almost exclusively at first-codon positions (Fig. 3) and predominantly correct internal stop codons to sense codons (arginine or glutamine; Fig. 4). Additionally, C-to-U edits show considerable variation in number across species, while U-to-C edits are more stable in frequency within each lineage but occur at higher overall levels in Hymenophyllaceae than in Vittarioideae (Fig. 2).

Levels of editing efficiency suggest RNA editing is regulated at start and internal stop codons

For both C-to-U and U-to-C editing, most nonsynonymous edits occurred with relatively high efficiency, whereas synonymous edits – those that produce no amino acid change – took place with significantly lower efficiency (Fig. 5). Notably, U-to-C edits that correct internal stop codons and C-to-U edits that restore start codons occurred with lower efficiencies than nonsynonymous edits of the same type, although the difference in restoring start codon edits for Hymenophyllaceae was not significant (Fig. 5). This unanticipated pattern challenges the notion that RNA editing functions solely as a compensatory corrective mechanism. If editing served purely to restore function, one would expect the sites with the most profound functional impact, such as edits to restore start codons and to correct premature stop codons, to occur with the highest efficiency. Instead, the lower and more variable efficiencies of these edits, compared with nonsynonymous edits, suggest they are regulated, potentially acting as a gene regulatory 'switch', as has been previously proposed for start codon edits (Miyata & Sugita, 2004; Bentolila *et al.*, 2013; Li *et al.*, 2018). Our findings extend this concept to edits correcting internal stop codons. That these edits are significantly positioned toward the 5′-end of the transcript further supports our hypothesis that they may act as gene regulatory checkpoints (Fig. 4). These types of edits (restoring start codons and correcting stop codons) may enable the nucleus to exert fine control over plastid gene expression, ensuring proper coordination between nuclear and plastid components – an essential aspect of cellular homeostasis given the interdependence of these systems in plant cells (Figs 4, 5).

Start and stop codon edits are evolutionarily conserved whereas nonsynonymous edits are progressively lost

Across both datasets, the number of C-to-U edits was strongly and inversely correlated with substitution rate, even when accounting for phylogenetic relatedness (Fig. 6), indicating an evolutionary trend of progressive loss of C-to-U editing sites. This aligns with previous findings in angiosperms, where C-to-U edits have been shown to also be progressively lost over time (Mower, 2008; Ishibashi *et al.*, 2019). By contrast, U-to-C edits, which predominantly correct internal stop codons, did not show consistent patterns relative to evolutionary rate (Fig. 6). No significant correlation between U-to-C editing and evolutionary rate was detected (Fig. 6) in Vittarioideae; however, in Hymenophyllaceae, a significant inverse correlation was observed between U-to-C editing and branch length ($P = 0.043$; Fig. 6), though very close to the 0.05 threshold for significance. This discrepancy may stem from the types of U-to-C edits present. Hymenophyllaceae harbors more nonsynonymous U-to-C edits than Vittarioideae (Fig. 4). Fauskee *et al.* (2021) proposed that U-to-C edits that

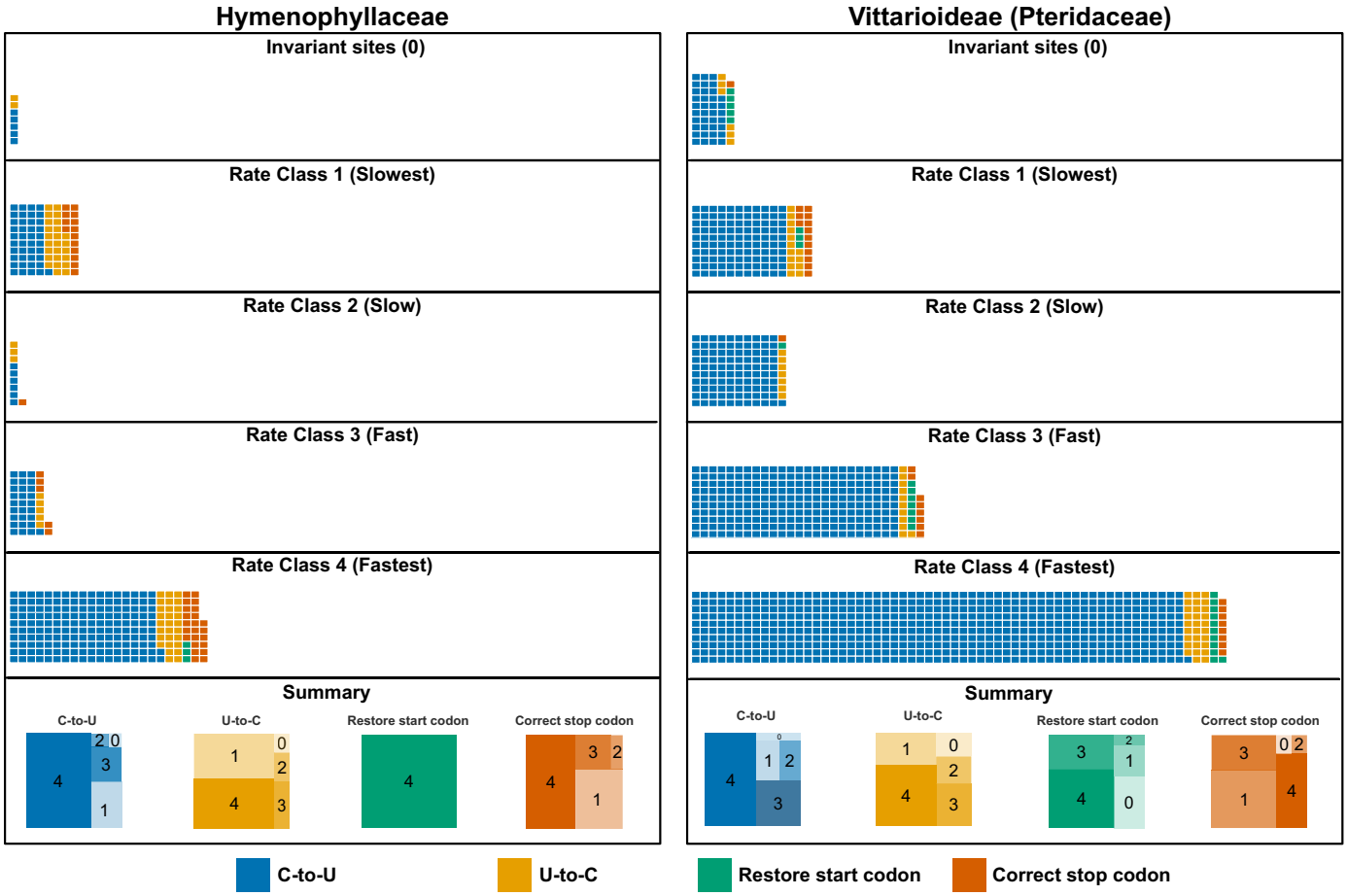

Fig. 8 Distribution of RNA editing sites across five rate categories in Hymenophyllaceae (left) and Vittarioideae (Pteridaceae; right). Rate categories 1–4 were estimated using either a discretized gamma distribution or a free rateheterogeneity model where rates are inferred independently without an underlying parametric distribution. Among these, rate Category 1 represents the slowest evolving sites, with rates increasing up to Category 4, which represents the fastest evolving sites. Category 0 corresponds to invariant sites. Each box represents a unique RNA editing site. Nonsynonymous cytidine-to-uridine (C-to-U) edits are shown in blue, nonsynonymous U-to-C edits in yellow, C-to-U edits at start codons in green, and U-to-C edits at internal stop codons in dark orange. The lower panel displays summary treemaps, illustrating the overall proportion of rate categories occupied by RNA editing subeset.

correct stop codons are selectively retained, whereas nonsynonymous U-to-C edits are far less conserved. The significant correlation in Hymenophyllaceae may be driven by a progressive loss of nonsynonymous U-to-C edits that are more abundant in this group than in Vittarioideae.

To further investigate differences in evolutionary histories within each type of editing (C-to-U and U-to-C), we categorized edits into four groups: nonsynonymous C-to-U, nonsynonymous U-to-C, C-to-U edits restoring start codons, and U-to-C edits correcting internal stop codons. While nonsynonymous edits were largely not conserved across species, start and stop codon edits stood out as notably conserved, particularly in Vittarioideae (Fig. 7). This pattern suggests that these subsets of edits may serve functional roles, contrasting with the expectation that RNA editing sites evolve neutrally. Together with our PGLS results (Fig. 6), these findings indicate that nonsynonymous edits are rapidly lost, while start and stop codon edits follow distinct evolutionary trajectories and are more conserved. However, the simultaneous presence of specific RNA editing sites in several species does not necessarily indicate strict ancestral retention – some gains may have occurred recently and independently in closely related species. To clarify whether conserved edits reflect long-term maintenance or recurrent gains, we estimated site-specific substitution rates as well as rates of gain and loss for RNA editing sites in a comparative phylogenetic context, providing direct insights into their evolutionary dynamics.

Our phylogenetic analyses reveal distinct evolutionary histories for the different classes of RNA editing sites. Nonsynonymous C-to-U and U-to-C RNA editing sites predominantly occupy high-rate categories relative to other sites in the same gene and are lost at a higher rate than gained (Figs 8, 9), aligning with expectations under CNE. By contrast, edits that restore start codons (C-to-U) and correct internal stop codons (U-to-C) display evolutionary patterns inconsistent with neutral loss. Edits restoring start codons exhibited a mixed evolutionary signal. In Hymenophyllaceae, only three start codon edits were found in genes with among-site rate heterogeneity, and all fell within the highest substitution rate category (Fig. 8). However,

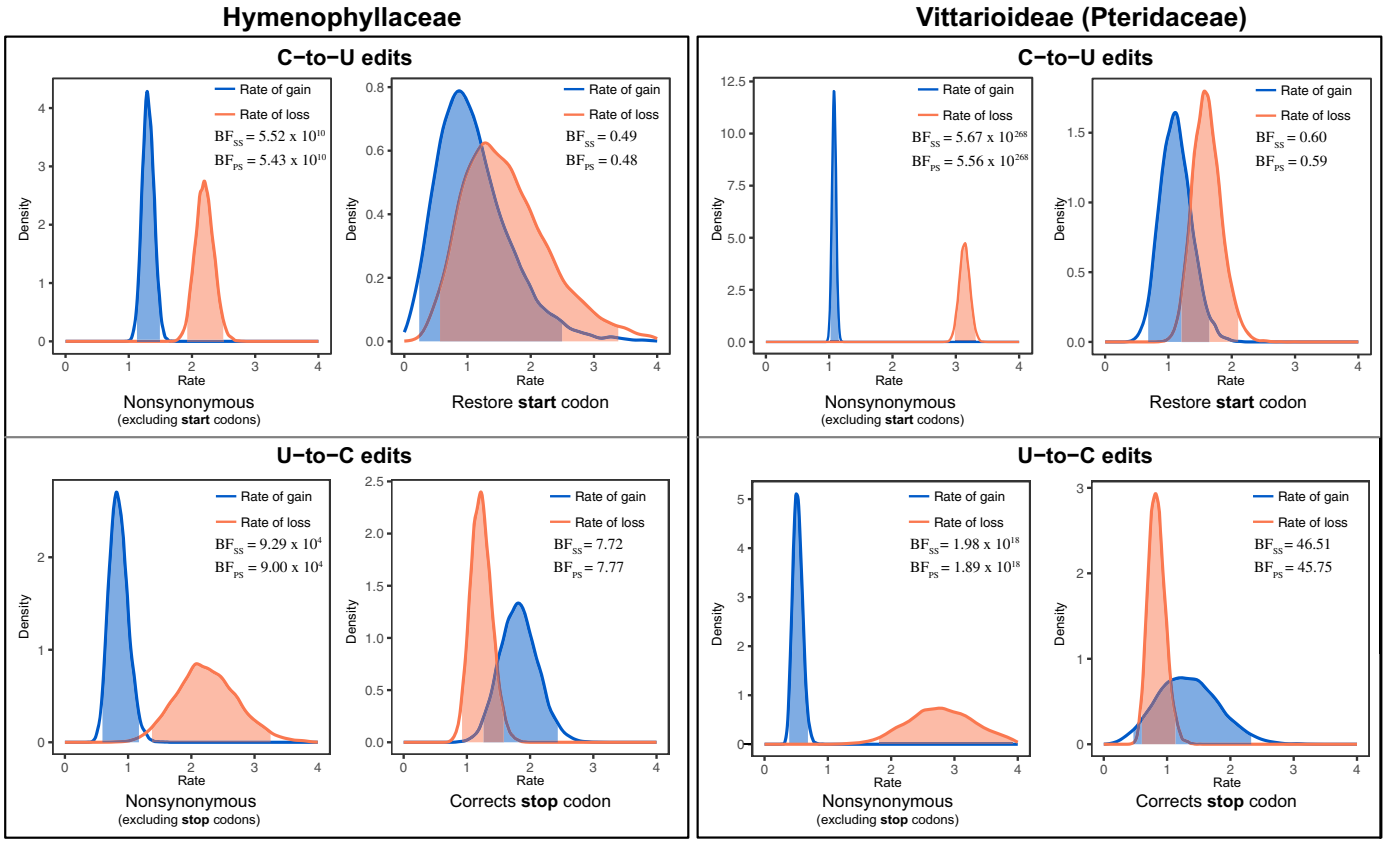

Fig. 9 Marginal densities of rate estimations for gain and loss of RNA editing sites for Hymenophyllaceae and Vittarioideae (Pteridaceae). Cytidine-to-uridine (C-to-U) edits are shown in the top panel and U-to-C edits in the bottom panel. Rate of gain is shown in blue, and rate of loss is shown in orange. Bayes factor (BF) values indicate support for a two-rate model over a one-rate model, with BF_{ss} and BF_{ps} representing values obtained from sampling marginal likelihoods with stepping-stone and path sampling methods, respectively.

in Vittarioideae, a greater proportion of start codon edits occupied low-rate categories than other nonsynonymous C-to-U edits (Fig. 8), suggesting some degree of conservation. Unlike nonsynonymous C-to-U edits, start codon edits follow an evolutionary model where gains and losses occur at equal rates (Fig. 9), indicating that they are not progressively lost over time. Strikingly, U-to-C edits that correct internal stop codons exhibit strong evolutionary conservation. These edits are disproportionately found in low-rate categories (Fig. 8), reinforcing their selective retention. Moreover, internal stop codon edits are gained at a higher rate than they are lost (Fig. 9), meaning even edits found at high-rate sites are progressively gained rather than lost – a pattern that starkly contrasts with nonsynonymous edits. This suggests that edits correcting internal stop codons are not only selectively retained over time, but are also continually introduced into the plastomes, further supporting their functional significance.

Nonsynonymous edits evolve via CNE while start and stop codon edits represent adaptive co-options

Overall, our findings reveal a duality in the evolutionary dynamics of RNA editing in plant chloroplast genomes. Nonsynonymous edits (both C-to-U and U-to-C), which constitute most RNA editing events, align with CNE, evolving neutrally and being progressively lost over time (Figs 6–9). The observed progressive loss is a signature of RNA editing evolving under CNE since losing an RNA editing site via back mutation is far more likely than gaining a novel RNA editing site. By contrast, RNA edits that restore start codons and correct internal stop codons evolve in patterns inconsistent with expectations under CNE. Edits correcting start codons have equal rates of gain and loss and, in Vittarioideae, more often occupy low substitution rate categories (Figs 8, 9). Edits correcting internal stop codons disproportionately occupy low-rate categories across both datasets, indicating strong selective retention. Additionally, internal stop codon edits exhibit a pattern of progressive gain rather than loss, further distinguishing them from nonsynonymous edits and reinforcing their functional significance. Their lower and more variable editing efficiencies, combined with their tendency to occur near the 5′-end of transcripts, suggest that these edits are not simply compensatory but are instead regulated, likely acting as molecular checkpoints for plastid gene expression. As schematized in Fig. 10, editing at internal stop codons may function as gene expression 'checkpoints', while start codon edits may act as 'switches' that control the initiation of translation.

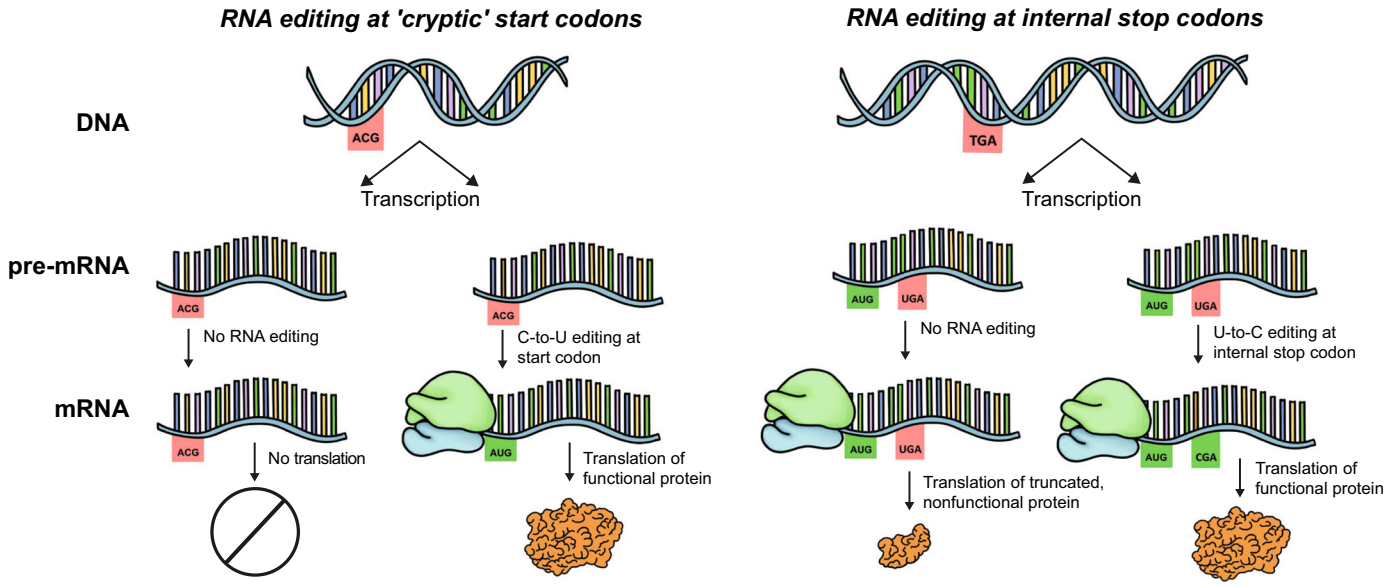

Fig. 10 Schematic of how RNA editing may regulate plastid gene expression. Red codons represent un-edited start or stop codons. Green codons represent edited start or stop codons. Illustrations by Tessa Patton.

These findings suggest that while most RNA editing sites arise and persist as neutral byproducts of CNE, a subset – specifically edits at internal stop codons, and to a lesser extent, start codons – may have been co-opted for regulatory functions. The progressive accumulation and selective retention of internal stop codon edits, coupled with their lower and more variable editing efficiencies, align with a model in which these edits play a role in modulating gene expression rather than simply restoring conserved protein sequences. Their positional bias near the 5′-end of transcripts further supports this interpretation, as edits occurring earlier in translation would provide a more efficient mechanism for controlling protein synthesis and reducing the metabolic cost of producing defective polypeptides. This is especially relevant in plastids, which lack nonsense-mediated decay (NMD) – a eukaryotic nuclear mechanism that degrades transcripts with premature stop codons (Baker & Parker, 2004). Without NMD, plastids would otherwise translate truncated proteins, making stop codon edits near the start of transcripts a potentially efficient means of minimizing wasteful translation. As illustrated in Fig. 10, such edits may serve as regulatory 'checkpoints' that fine-tune translation in the absence of transcript-level quality control. Notably, the regulation of these edits may enable the nucleus to exert fine-scale control over plastid gene expression, ensuring proper coordination between nuclear- and plastid-encoded components. While these patterns strongly suggest a functional role for start and stop codon edits, further experimental work – including plastid transformation and targeted knockouts of specific editing factors – is necessary to directly test their regulatory significance. Similarly, while the evolutionary trajectory of start codon edits is less conclusive, their distinct substitution rate patterns and lower editing efficiencies suggest that they, too, may be subject to regulatory control. Together, these patterns challenge the notion that RNA editing in plants is strictly a neutral process and instead point to a more dynamic evolutionary landscape where neutral and adaptive forces interact to shape RNA editing site evolution.

Acknowledgements

BDF was supported by a Graduate Research Fellowship from the National Science Foundation (DGE 2139754). Additional funding for this research was provided by Grants-in-Aid from the Duke University Biology Program, a Graduate Student Research Award from the Society of Systematic Biologists, and a Research Grant from the International Association of Plant Taxonomists (all awarded to BDF). We thank New England Biolabs for providing RNA library preparation kits as part of a beta test agreement, and particularly Deyra Rodriguez for technical assistance. We also appreciate the support of Dr Holly Forbes and the UC Botanic Garden, as well as the Dr Cecilia Koo Botanical Conservation Center, for providing plant samples. We are especially grateful to Dr Ben Redelings, Dr David Swofford, Dr Joe Bielawski, Dr Jeffrey Mower, Dr Fay-Wei Li, Dr María Virginia Sanchez-Puerta, Dr Anne Yoder, and Dr Paul Lewis for their insightful discussions on experimental design and data interpretation. Additionally, we sincerely thank Dr Layne Huiet for invaluable mentorship and assistance with molecular work. We also thank Tessa Patton for her illustrations in Fig. 10. Finally, we are extremely grateful to Bruno do Rosario Petrucci and Sigournie Brock for their hospitality during a research visit to Iowa State University by BDF. This paper is part of a doctoral dissertation completed at Duke University by BDF under the direction of KMP.

Competing interests

None declared.

Author contributions

BDF designed the research, isolated the nucleic acids, prepared the NGS libraries, assembled the organellar genomes, performed the editome and phylogenetic analyses, and wrote the manuscript. L-YK assisted with research design, assembled and annotated the organellar genomes, isolated the nucleic acids, prepared the NGS libraries, advised on the editome analyses, and edited the manuscript. TAH advised and assisted on the phylogenetic analyses and edited the manuscript. P-JX isolated the nucleic acids and prepared the NGS libraries. KMP advised and assisted with the research design and data interpretation and edited the manuscript.

ORCID

Blake D. Fauskee https://orcid.org/0000-0001-5251-0959
Tracy A. Heath https://orcid.org/0000-0002-0087-2541
Li-Yaung Kuo https://orcid.org/0000-0002-3388-3757
Kathleen M. Pryer https://orcid.org/0000-0002-9776-6736
Pei-Jun Xie https://orcid.org/0000-0001-8578-5772

Data availability

Newly generated raw genomic and transcriptomic data are available from the Sequence Read Archive under BioProject no. PRJNA1216602. Specific BioSample and SRA accession numbers are available in Table S1. Chloroplast genome assemblies are available from GenBank with accession nos. PV121143–PV121190. Table S1 lists which accession numbers accompany which species. Annotated chloroplast genome assemblies are additionally available at https://github.com/bfauskee/fauskee-fern-rna-editing-scripts along with all scripts and editing matrices.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Auborg S, Boudet N, Kreism M, Lecharny A. 2000. In *Arabidopsis thaliana*, 1% of the genome encodes for a novel protein family unique to plants. *Plant Molecular Biology* **42**: 603–613.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29**: 2157–2167.

Baker KE, Parker R. 2004. Nonsense-mediated decay: terminating erroneous gene expression. *Current Opinion in Cell Biology* **16**: 293–299.

Barkan A, Small I. 2014. Pentatrichopeptide repeat proteins in plants. *Annual Review of Plant Biology* **65**: 415–442.

Benne R, Den Ven Berg J, Brakenhoff JP, Sloor P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**: 819–826.

Bentolila S, Oh J, Hanson MR, Bukowski R. 2013. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genetics* **9**: e1003584.

Berget SM, Sharp PA. 1977. A spliced sequence at the 5′-terminus of adenovirus late mRNA. *Brookhaven Symposia in Biology* **29**: 332–344.

Birk MA, Liscovitch-Brauer N, Dominguez MJ, McNeme S, Yue Y, Hoff JD, Twersky I, Verhey JK, Sutton RB, Eisenberg E *et al.* 2023. Temperature-dependent RNA editing in octopus extensively recodes the neural proteome. *Cell* **186**: 2544–2555.

Bolger AM, Lohse M, Usadel B. 2014. TRIMMOMATIC: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**: e1660.

Chen H, Deng L, Jiang Y, Lu P, Yu J. 2011. RNA editing sites exist in protein-coding genes in the chloroplast genome of Cycas taitungensis. *Journal of Integrative Plant Biology* **53**: 961–970.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**: 1–8.

Covello PS, Gray MW. 1993. On the evolution of RNA editing. *Trends in Genetics* **9**: 265–268.

Crick FH. 1958. On protein synthesis. *Symposium of the Society for Experimental Biology* **12**: 138–163.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPLASTY: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45**: e18.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Edera AA, Sanchez-Puerta MV. 2021. Computational detection of plant RNA editing events. *Methods in Molecular Biology* **2181**: 13–34.

Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. 2011. Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution* **28**: 523–532.

Fauskee BD, Sigel EM, Pryer KM, Grusz AL. 2021. Variation in frequency of plastid RNA editing within Adiantum (Pteridaceae) implies rapid evolution in fern plastomes. *American Journal of Botany* **108**: 2558–2571.

Gernhard T. 2008. The conditioned reconstruction process. *Journal of Theoretical Biology* **253**: 769–778.

Gilbey L. 2022. *GGWAFFLE: waffle charts*. R package v.0.2.5. [WWW document] URL https://liamgilbey.github.io/ggwaffle/.

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.

Gutmann B, Royan S, Schallenberg-Rüdinger M, Lenz H, Castleden IR, McDowell R, Vacher MA, Tonti-Filippini J, Bond CS, Knoop V *et al.* 2020. The expansion and diversification of pentatricopeptide repeat RNA-editing factors in plants. *Molecular Plant* **13**: 215–230.

Guzowska-Nowowiejska M, Fiedorowicz E, Plader W. 2009. Cucumis sativus, melon, pumpkin, and squash: are rules of editing in flowering plants' chloroplast genes so well known indeed? *Gene* **434**: 1–8.

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Hulsenbeck JP, Ronquist F. 2016. REVBAYES: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* **65**: 726–736.

Huang YY, Matzke AJ, Matzke M. 2013. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS ONE* **8**: e74736.

Huiet L, Li FW, Kao TT, Prado J, Smith AR, Schuettpelz E, Pryer KM. 2018. A worldwide phylogeny of Adiantum (Pteridaceae) reveals remarkable convergent evolution in leaf blade architecture. *Taxon* **67**: 488–502.

Ishibashi K, Small I, Shikanai T. 2019. Evolutionary model of plastidial RNA editing in angiosperms presumed from genome-wide analysis of *Amborella trichopoda*. *Plant Cell Physiology* **60**: 2141–2151.

Jeffreys H. 1961. Some tests of significance, treated by the theory of probability. In: *Theory of probability, 3rd edn.* Oxford, UK: Oxford University Press.

Jiang Y, Fan SL, Song MZ, Yu JN, Yu SX. 2012. Identification of RNA editing sites in *Gossypium hirsutum* chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genetics and Molecular Research* **11**: 987–1001.

Jombart T, Balloux F, Dray S. 2010. ADEPHYLO: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**: 1907–1909.

Kahlau S, Aspinall S, Gray JC, Bock R. 2006. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *Journal of Molecular Evolution* **63**: 194–207.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kendall DG. 1948. On the generalized "birth-and-death" process. *Annals of Mathematical Statistics* **19**: 1–15.

Khanna A, Larson DE, Srivatsan SN, Mosoir M, Abbott TE, Kiwala S, Ley TJ, Duncavage EJ, Walter MJ, Walker JR *et al.* 2022. Bam-readcount—rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software* **7**: 3722.

Knie N, Grewe F, Fischer S, 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**: 134.

Knoop V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. *Cellular and Molecular Life Sciences* **68**: 567–586.

Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K. 2003. RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Research* **31**: 2417–2423.

Kuo LY, Qi X, Ma H, Li FW. 2018. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. *American Journal of Botany* **105**: 1545–1555.

Langmead B, Salzburg SL. 2012. Fast gapped-read alignment with BOWTIE2. *Nature Methods* **9**: 357–359.

Lartillot N, Philippe H. 2006. Computing bayes factors using thermodynamic integration. *Systematic Biology* **55**: 195–207.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete character data. *Systematic Biology* **50**: 913–925.

Li FW, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux PM, Eily A, Koppers N, Kuo LY, Li Z *et al.* 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**: 460–472.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMTOOLS. *Bioinformatics* **25**: 2078–2079.

Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. 2011. How a neutral ratchet can build cellular complexity. *IUBMB Life* **63**: 528–537.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Mirzaei S, Mansouri M, Mohammadi-Nejad G, Sablok G. 2018. Comparative assessment of chloroplast transcriptional responses highlights conserved and unique patterns across Triticeae members under salt stress. *Photosynthesis Research* **136**: 357–369.

Miyata Y, Sugita M. 2004. Tissue- and stage-specific RNA editing of rps14 transcripts in moss (*Physcomitrella patens*) chloroplasts. *Journal of Plant Physiology* **161**: 113–115.

Mower JP. 2008. Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. *Molecular Biology and Evolution* **25**: 52–61.

Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **344**: 305–311.

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. 2022. An open and continuously updated fern tree of life. *Frontiers in Plant Science* **13**: 909768.

Oldenkott B, Yamaguchi K, Tsuji-Tsukinoki S, Knie N, Knoop V. 2014. Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. *RNA* **20**: 1499–1506.

O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Molecular Biology and Evolution* **25**: 1120–1128.

Paradis E, Schliep K. 2019. APE 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Pelosi JA, Davenport R, Barbazuk WB, Sessa EB, Kuo LY. 2024. An efficient and effective RNA extraction protocol for ferns. *Applications in Plant Sciences* **12**: e11617.

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, Fitzjohn RG, Alfaro ME, Harmon LK. 2014. GEIGER v.2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**: 2216–2218.

Pinheiro J, Bates D, R Core Team. 2024. NLME: linear and nonlinear mixed effects models. R package v.31.166 [WWW document] URL https://CRAN.R-project.org/package=nlme.

Rambaut A, Drummond AJ, Xie D, Beale G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using TRACER 1.7. *Systematic Biology* **67**: 901–904.

Revell LJ. 2012. PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Robison TA, Grusz AL, Wolf PG, Mower JP, Fauskee BD, Sosa K, Schuettpelz E. 2018. Mobile elements shape plastome evolution in ferns. *Genome Biology and Evolution* **10**: 2558–2571.

Rothfels CJ, Schuettpelz E. 2014. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. *Systematic Biology* **63**: 31–54.

Rüdinger M, Polsakiewicz M, Knoop V. 2008. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Molecular Biology and Evolution* **25**: 1405–1414.

Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB. 2013. Arabidopsis chloroplast quantitative editotype. *FEBS Letters* **587**: 1429–1433.

Schafran P, Hauser DA, Nelson JM, Xu X, Mueller LA, Kulshrestha S, Smalley I, de Vries S, Irisarri I, de Vries J *et al.* 2025. Pan-phylum genomes of hornworts reveal conserved autosomes but dynamic accessory and sex chromosomes. *Nature Plants* **11**: 49–62.

Schuettpelz E, Chen CW, Kessler M, Pinson JB, Johnson G, Davila A, Cochran AT, Huiet L, Pryer KM. 2016. A revised generic classification of vittarioid ferns (Pteridaceae) based on molecular, micromorphological, and geographic data. *Taxon* **65**: 708–722.

Schuettpelz E, Pryer KM. 2006. Reconciling extreme branch length differences: decoupling time and rate through the evolutionary history of filmy ferns. *Systematic Biology* **55**: 485–502.

Sidow A, Nguyen T, Speed TP. 1992. Estimating the fraction of invariable codons with a capture–recapture method. *Journal of Molecular Evolution* **35**: 253–260.

Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. 2020. Plant organellar RNA editing: what 30 years of research has revealed. *The Plant Journal* **101**: 1040–1056.

Steel M, Huson D, Lockhart PJ. 2000. Invariable sites models and their use in phylogenetic reconstruction. *Systematic Biology* **49**: 225–232.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. *Journal of Molecular Evolution* **49**: 169–181.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**: 57–86.

Testo W, Sundue M. 2016. A 4000-species dataset provides new insight into the evolution of ferns. *Molecular Phylogenetics and Evolution* **105**: 200–211.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. PILON: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: e112963.

Wang W, Zhang W, Wu Y, Maliga P, Messing J. 2015. RNA editing in chloroplasts of *Spirodela polyrhiza*, an aquatic monocotyledonous species. *PLoS ONE* **10**: e0140285.

Wolf PG, Rowe CA, Hasebe M. 2004. High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus*-veneris. *Gene* **339**: 89–97.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**: 150–160.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306–314.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Table S1 Voucher information for newly generated sequences presented in this study.

Table S2 Bayes factor test results.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.