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], Kathleen M. Pryer[1]

[1]Department of Biology, Duke University, Durham 27708, North Carolina, USA

[2]Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

[3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames 50011, Iowa, USA

[4]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca 14850, New York, USA

*Corresponding author: E-mail: blake.fauskee@duke.edu

ORCID's

Blake D. Fauskee: 0000-0001-5251-0959

Li-Yaung Kuo: 0000-0002-3388-3757

Tracy A. Heath: 0000-0002-0087-2541

Pei-Jun Xie: 0000-0001-8578-5772

Kathleen M. Pryer: 0000-0002-9776-6736

Word Counts:

Introduction: 1303 words

Materials and Methods: 2699 words

Results: 1149 words

Discussion: 2101 words

Figures, Tables and Supporting Information: 9 Figures (all color), 0 tables, 2 supporting tables

Summary

- The role of RNA editing in plant organellar genomes is a longstanding evolutionary puzzle. It is widely regarded as a neutral corrective mechanism, yet it persists as a complex process requiring numerous nuclear-encoded PPR proteins. The evolutionary forces maintaining the system remain unclear.
- Ferns are the most diverse lineage of land plants that uniquely retain both C-to-U and U-to-C RNA editing in the plastome, providing a rare opportunity to investigate the evolution of both RNA editing types.
- We investigated two distantly related fern lineages—Hymenophyllaceae and Vittarioideae (Pteridaceae)—each containing sister sublineages with contrasting evolutionary rates (slow vs. fast). We combine genomic and transcriptomic data together with rigorous phylogenetic methods to explore the evolution of the plastid editomes in both lineages.
- Our analyses reveal consistent evolutionary patterns across both fern groups, including widespread neutral loss of nonsynonymous edits, but also the selective retention of potentially functionally significant sites.
- C-to-U edits at start codons and U-to-C edits at internal stop codons are highly conserved and show signatures of selective regulation, suggesting they function as molecular checkpoints. These findings provide the strongest evidence to date that RNA editing in plants may serve an adaptive role in plastid gene regulation.

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

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 et al. 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, in 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 post-transcriptionally.

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 adenosine-to-inosine (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* exceeds 3000 editing sites (Oldenkott et al. 2014) (Fig. 1). Ferns, the sister lineage to seed plants, range from having 100's 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 (Aubourg 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 and 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 (Lukeś 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 and Gray (1993) proposed a neutral framework, later formalized by Stoltzfus (1999) as the "Constructive Neutral Evolution", or CNE, hypothesis. CNE 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 and 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 backmutate 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 and 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 is 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 and Schuettpelz 2014) and the Hymenophyllaceae (Schuettpelz and 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 and Schuettpelz 2014). A similar pattern is exhibited in Hymenophyllaceae, where the sister subfamily Trichomanoideae displays a much higher substitution rate than the Hymenophylloideae (Schuettpelz and Pryer 2006). By expanding the available data for these two lineages, we report here on the plastid RNA editing sites for ten *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 withing a comparative phylogenetic context, we aim 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 ten *Adiantum* species, eight vittarioids, seven species of Trichomanoideae, and four from Hymenophylloideae (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 the same as for our transcriptomic data, see below). Namely, the plastome assemblies of *Antrophyum semicostatum, Scoliosorus ensiformis, Vaginularia junghuhnii,* and *Vittaria graminifolia* 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 (Omega Bio-Tek, Norcross, GA, USA; D5511).

DNA was quantified on the Qubit 2 Fluorometer (Thermo Fisher Scientific Inc., Walden, MA, USA) using the Qubit dsDNA High Sensitivity Quantification Assay kit (Thermo Fisher Scientific Inc., Walden, MA, USA; Q32851). DNA libraries for Illumina sequencing were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biosciences, Ipswitch, MA, USA; E7645) 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 (New England Biosciences, Ipswitch, MA, USA; E6609). Libraries were then sequenced on the Illumina Novaseq X Plus by Novogene Co., Ltd (Beijing, China) using 150 base-pair, paired-end chemistry. De-multiplexing was also performed by Novogene Co., Ltd (Beijing, China).

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 version 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 version 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 version 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 version 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 (Omega Bio-Tek, Norcross, GA, USA; R6827) 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 (Millipore Sigma, Darmstadt, Germany**;** 69182). The resulting RNA extractions were then quantified on the Qubit 2 Fluorometer (Thermo Fisher Scientific Inc., Walden, MA, USA) using the Qubit RNA High Sensitivity Quantification Assay kit (Thermo Fisher Scientific Inc., Walden, MA, USA; Q32852).

RNA (cDNA) libraries were then constructed using one of two library preparation kits: the Zymo-Seq RiboFree Total RNA Library Kit (Zymo Research, Irvine, CA, USA; R3000) or the NEBNext Ultra II Library Prep (New England Biosciences, Ipswitch, MA, USA; E7775) 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 manufacturer's 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 number: 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 and 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 version 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 non-uniform (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 version 2.2.4 (Langmead and Salzberg 2012). Samtools version 1.14 (Li et al. 2009) was used to construct a BAM file that was used as input to bam-readcount version 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 versus 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 version 7.505 (Katoh and Standley 2013). Here only the coding sequences were aligned; flanking regions and introns were excluded. Alignment position was then used to compare 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 lacks *psaM* and *ycf66* that are present in Hymenophyllaceae. Based on previously published phylogenetic topologies across all ferns (Testo and Sundue 2016; Nitta et al. 2022), one outgroup for each dataset was also included: *Osmundastrum cinnamomeum* for Hymenophyllaceae and *Myriopteris*

lanosa for Vittarioideae. Gene alignments were carried out again to include outgroups using MAFFT version 7.505 (Katoh and Standley 2013) and alignments were then concatenated using AMAS (Boroweic 2016) and partitioned by gene. ML tree estimation was carried out in IQ-TREE2 version 2.2.2.7 (Minh et al. 2020) with 1,000 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 et al. 2004), 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 version 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 et al. 2004). 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 et al. 2004). 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 version 1.2.3 (Höhna et al. 2016).

RNA editing data were modeled using a continuous-time Markov chain (CTMC) 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 log-normal (UCLN) relaxed clock (Drummond et al. 2006). The mean and standard deviation of the log-normal distribution were assigned exponential priors (dnExponential(2.0) for the mean, dnExponential(3.0) for the standard deviation), 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 standard deviation of the log-normal 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 to 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 maximum-likelihood tree topology was fixed, and only node ages were estimated using a time-scaling move. We ran two independent Markov chain Monte Carlo (MCMC) chains, each running for 2,000,000 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 (ESS) of parameters and visual inspection of trace plots in Tracer version 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 1,000,000 generations, sampling parameters every 100 generations. Convergence was assessed by visual inspection of trace plots in Tracer version 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 (Xie et al. 2011; Fan et al. 2011) and path sampling (Lartillot 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 1,000 MCMC generations per power posterior after an initial burn-in of 2,000 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 greater than 1 suggests that the FreeK model provides a better fit to the data, while a Bayes factor much less than 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 maximum likelihood 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 155-124 (*H. fujisanense–H. holochilum*), whereas the faster Trichomanoideae

(TRI) show a lower number of C-to-U edits, ranging from 115–44 (*A. obscurum–D. tahitense*). The lower number of U-to-C edits across the Hymenophyllaceae overlap, ranging from 69–48 in the Hymenophylloideae (*H. pallidum–H. holochilum*) and from 59-32 in the Tricomanoideae (*A. obscurum–D. tahitense*). In Vittarioideae, the number of C-to-U edits is higher in the slower-evolving *Adiantum* (ADI), ranging from 539–299 (*A. shastense–A. davidii*), whereas the faster vittarioids (VIT) consistently show a lower number of C-to-U edits, ranging from 221–182 (*H. yakushimensis–V. junghuhnii*). The lower number of U-to-C edits across the Vittarioideae are very similar, ranging from 43–31 in *Adiantum* (*A. caudatum–A. tenerum*) and from 39–29 in the vittarioids (*H. yakushimensis–V. lineata*).

In both Hymenophyllaceae and Vittarioideae, C-to-U edits primarily impacted 2nd-codon positions, while U-to-C edits mostly occurred at 1st-codon positions, and a small proportion of both editing types acted on 3rd-codon positions (Fig. 3). Most C-to-U edits resulted in nonsynonymous amino acid changes, with a small proportion of these edits resulting in 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 to both nonsynonymous and synonymous edits, whereas no significant difference in transcript position was observed between nonsynonymous and synonymous edits (Fig. 4).

462

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

473

474 *Comparative phylogenetic analyses of RNA editing sites*
475 A phylogenetic generalized least squares analysis was used to test the
476 relationship between evolutionary rate (as the root-to-tip distance on the ML
477 plastid phylogeny) and the number of RNA edits of each type (C-to-U and U-
478 to-C). A significant negative relationship between the root-to-tip distance
479 and the proportion of C-to-U editing was detected in both fern lineages (Fig.
480 6; $p=0.001$ for Vittarioideae and $p=0.002$ for Hymenophyllaceae). A negative
481 relationship was also observed between the proportion of U-to-C editing and
482 root-to-tip distance in Hymenophyllaceae ($p=0.043$); however, a similar
483 correlation was not supported for Vittarioideae ($p=0.42$; Fig. 6).

484

485 Sequence alignments for each gene were used to identify shared RNA editing
486 sites across species within the independent datasets. To systematically
487 compare the presence or absence of RNA editing site across taxa, each RNA
488 editing site was assigned a unique identifier by combining the gene name
489 with the alignment position where an edit occurs. This unique tag allowed for
490 a direct comparison between RNA editing site presence or absence across

491 species. The distribution of shared editing sites was visualized using UpSet
492 plots, providing a detailed overview of editing site conservation within each
493 dataset. The analyses revealed that nonsynonymous C-to-U RNA editing
494 sites were predominantly taxon-specific, with few edits shared across
495 multiple species in either dataset (Fig. 7). Nonsynonymous U-to-C edits
496 showed slightly higher conservation, particularly in Vittarioideae, but the
497 majority were not shared by many taxa (Fig. 7). Notably, C-to-U edits that
498 restore start codons and U-to-C edits that correct internal stop codons were
499 widely shared across many species within each dataset (Fig. 7).
500
501 To investigate the rate at which RNA editing sites evolve, site-specific
502 substitution rates were calculated for each site in those genes where the
503 best fitting substitution model (according to BIC) contained an among-site
504 rate heterogeneity parameter. Out of 85 genes for Vittarioideae, 73 had best-
505 fitting models accommodating an among-site rate-heterogeneity parameter,
506 while 68 out of 87 genes for Hymenophyllaceae qualified. For applicable
507 genes, the rate category for each unique edit site was plotted. Rate
508 categories were used to enable comparisons across genes within datasets.
509 These analyses reveal that most RNA editing sites fall into high-rate
510 categories (Fig. 8). However, a large proportion of U-to-C edits, both
511 nonsynonymous edits and edits on internal stop codons, fall into low-rate
512 categories (Fig.8). Additionally, a significant proportion of C-to-U edits at
513 internal stop codons occupy low-rate categories for Vittarioideae, while all
514 three unique start codon edits (within genes with suitable models for this
515 analysis) for Hymenophyllaceae occupy the highest rate category (Fig. 8).
516
517 Transition rates for RNA editing sites were estimated using comparative
518 phylogenetic models in RevBayes (Höhna et al. 2016). Both an equal-rate
519 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). In 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 are gained (Fig. 9). Start codon edits, however, were gained and lost at the same rate. In 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 post-transcriptionally, 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 Constructive Neutral Evolution (CNE) hypothesis (Covello and 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.

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* (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 (Knie et al. 2016; Guo et al. 2015). 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 2nd-codon positions (Fig. 3), producing mostly nonsynonymous amino acid changes, though a subset restores cryptic ACG start codons to AUG (Fig. 4). In contrast, U-to-C edits occur almost exclusively at 1st-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

606 For both C-to-U and U-to-C editing, most nonsynonymous edits occurred
607 with relatively high efficiency, whereas synonymous edits—those that
608 produce no amino acid change—took place with significantly lower
609 efficiency (Fig. 5). Notably, U-to-C edits that correct internal stop codons and
610 C-to-U edits that restore start codons occurred with lower efficiencies than
611 nonsynonymous edits of the same type, though the difference in restoring
612 start codon edits for Hymenophyllaceae was not significant (Fig. 5). This
613 unanticipated pattern challenges the notion that RNA editing functions solely
614 as a compensatory corrective mechanism. If editing served purely to restore
615 function, one would expect the sites with the most profound functional
616 impact, such as edits to restore start codons and to correct premature stop
617 codons, to occur with the highest efficiency. Instead, the lower and more
618 variable efficiencies of these edits, compared to nonsynonymous edits,
619 suggest they are regulated, potentially acting as a gene regulatory “switch”,
620 as has been previously proposed for start codon edits (Miyata and Sugita
621 2004; Bentolila et al. 2013; Li et al. 2018). Our findings extend this concept to
622 edits correcting internal stop codons. That these edits are significantly
623 positioned toward the 5’-end of the transcript further supports our
624 hypothesis that they may act as gene-regulatory checkpoints (Fig. 4). These
625 types of edits (restoring start codons and correcting stop codons) may
626 enable the nucleus to exert fine control over plastid gene expression,
627 ensuring proper coordination between nuclear and plastid components—an
628 essential aspect of cellular homeostasis given the interdependence of these
629 systems in plant cells (Figs. 4, 5).
630
631 *Start and stop codon edits are evolutionarily conserved whereas*
632 *nonsynonymous edits are progressively lost*
633 Across both datasets, the number of C-to-U edits was strongly and inversely
634 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). In 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 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. In 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, in Vittarioideae, a greater proportion of start codon edits occupied low-rate categories compared to 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

693 nonsynonymous edits. This suggests that edits correcting internal stop
694 codons are not only selectively retained over time, but are also continutally
695 introduced into the plastomes, further supporting their functional
696 significance.
697
698 *Nonsynonymous edits evolve via Constructive Neutral Evolution while start*
699 *and stop codon edits represent adaptive co-options*
700 Overall, our findings reveal a duality in the evolutionary dynamics of RNA
701 editing in plant chloroplast genomes. Nonsynonymous edits (both C-to-U
702 and U-to-C), which constitute most RNA editing events, align with
703 Constructive Neutral Evolution (CNE), evolving neutrally and being
704 progressively lost over time (Figs. 6–9). The observed progressive loss is a
705 signature of RNA editing evolving under CNE since losing an RNA editing site
706 via backmutation is far more likely than gaining a novel RNA editing site. In
707 contrast, RNA edits that restore start codons and correct internal stop
708 codons evolve in patterns inconsistent with expectations under CNE. Edits
709 correcting start codons have equal rates of gain and loss and, in
710 Vittarioideae, more often occupy low substitution-rate categories (Figs. 8–9).
711 Edits correcting internal stop codons disproportionately occupy low-rate
712 categories across both datasets, indicating strong selective retention.
713 Additionally, internal stop codon edits exhibit a pattern of progressive gain
714 rather than loss, further distinguishing them from nonsynonymous edits and
715 reinforcing their functional significance. Their lower and more variable editing
716 efficiencies, combined with their tendency to occur near the 5’ end of
717 transcripts, suggest that these edits are not simply compensatory but are
718 instead regulated, likely acting as molecular checkpoints for plastid gene
719 expression.
720
721 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 and 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. 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 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. 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

The authors declare no competing interests.

Author Contributions

BDF designed the research, isolated nucleic acids, prepared NGS libraries, assembled organellar genomes, performed editome and phylogenetic analyses, and wrote the manuscript. LYK assisted with research design, assembled and annotated organellar genomes, isolated nucleic acids and

prepared NGS libraries, advised on editome analyses, and edited the manuscript. TAH advised and assisted on phylogenetic analyses and edited the manuscript. PJX isolated nucleic acids and prepared NGS libraries. KMP advised and assisted with research design and data interpretation and edited the manuscript.

Data Availability

Newly generated raw genomic and transcriptomic data are available from the Sequence Read Archive under BioProject number: PRJNA1216602. Specific BioSample and SRA accession numbers are available in Table S1. Chloroplast genome assemblies are available from Genbank with accession numbers PV121143– PV121190. Table S1 lists which accession numbers go with 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.

Supporting Information

Table S1: Sampling information. Includes collection numbers for each specimen, voucher locations, Genbank accession numbers for sequences used for chloroplast genome assemblies, Genbank accession numbers used for plastome annotation references, as well as SRA BioSample numbers, and SRA accession for DNA and RNA data.

Table S2: Marginal likelihoods and Bayes factor values. Bayes factors were used to test whether the two-rate model (FreeK) was preferred over the one-rate model for each RNA editing subset. SS: marginal likelihood approximations from stepping-stone sampling approach. PS: marginal likelihood approximations from path-sampling approach

Figure Legends

Figure 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 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 ca. 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 et al. 2004, s) Rüdinger et al. 2008.

Figure 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 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 and Pryer 2006) and Vittarioideae (Rothfels and Schuettpelz 2014).

Figure 3: RNA editing by codon position in Hymenophyllaceae (left) and Vittarioideae (right). Colored bars depict the number of edits for each

837 species at each codon position. Species names are abbreviated at the
838 bottom. C-to-U edits are shown on the top and U-to-C edits on the bottom.
839
840 Figure 4: Proportion of RNA editing sites resulting in nonsynonymous (light
841 blue) and synonymous (gray) codon changes, as well as cryptic start codon
842 restoration (green, C-to-U only) and internal stop codon correction (orange,
843 U-to-C only). Hymenophyllaceae species are shown on the left and
844 Vittarioideae on the right. The top panels (A–B) show the proportions of C-to-
845 U edits by codon change type, while the middle panels (C–D) show the
846 proportions of U-to-C edits. The bottom panels (E–F) display the relative
847 genic locations of U-to-C edits, grouped by codon change type. Significant
848 differences in relative genic locations between codon change classes are
849 marked with asterisks denoting increasing p-value thresholds, and non-
850 significant p-values are reported. Species abbreviations match those in
851 Figure 3.
852
853 Figure 5: Summary of RNA editing efficiencies across Hymenophyllaceae
854 (left, lilac) and Vittarioideae (right, yellow). Efficiencies are grouped by codon
855 change type. C-to-U edits are shown on the top and U-to-C on the bottom.
856 Significant differences in efficiencies between codon change classes are
857 marked with asterisks denoting increasing p-alue thresholds. Non-significant
858 p-values are reported.
859
860 Figure 6: Relationship between the number of RNA edits and the nucleotide
861 substitution rate expressed as the root-to-tip branch length for each species.
862 Vittarioids are shown in orange/yellow and Hymenophyllaceae in dark
863 purple/light purple. Lighter colors represent U-to-C edits and darker colors
864 represent C-to-U edits. Trendlines are taken from a phylogenetic generalized

865 least squares analysis. P-values are shown at top of figure and significant p-
866 values are indicated by asterisks.
867
868 Figure 7: Proportion of shared RNA editing sites for Hymenophyllaceae (left)
869 and Vittarioideae (Pteridaceae; right). Nonsynonymous C-to-U edits are
870 shown in blue (top panel), nonsynonymous U-to-C edits in yellow (2nd panel),
871 C-to-U edits at start codons in green (3rd panel), and U-to-C edits at internal
872 stop codons in dark orange (bottom panel). Horizontal bars to the left of
873 species abbreviations show the number of edits present in each species.
874 Vertical bars denote how many edits are shared by an exclusive group of
875 taxa, defined by the connected dots. For Vittarioideae only sets of taxa
876 sharing only 2 or fewer edits are omitted and sets of taxa sharing one or fewer
877 edits are omitted for Hymenophyllaceae. Species abbreviations are
878 according to legend in bottom panel of Fig. 3.
879
880 Figure 8: Distribution of RNA editing sites across five rate categories in
881 Hymenophyllaceae (left) and Vittarioideae (Pteridaceae; right). Rate
882 categories 1–4 were estimated using either a discretized gamma distribution,
883 or a free-heterogeneity model where rates are inferred independently without
884 an underlying parametric distribution. Among these, rate category 1
885 represents the slowest evolving sites, with rates increasing up to category 4
886 which represents the fastest evolving sites. Category 0 corresponds to
887 invariant sites. Each box represents a unique RNA editing site.
888 Nonsynonymous C-to-U edits are shown in blue, nonsynonymous U-to-C
889 edits in yellow, C-to-U edits at start codons in green, and U-to-C edits at
890 internal stop codons in dark orange. The bottom panel displays summary
891 treemaps, illustrating the overall proportion of rate categories occupied by
892 edited sites in each dataset.

Figure 9: Marginal densities of rate estimations for gain and loss of RNA editing sites for Hymenophyllaceae and Vittarioideae (Pteridaceae). 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 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.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

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. https://doi.org/10.1023/A:1006352315928

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. https://doi.org/10.1093/molbev/mss084

924 Baker KE, Parker R. 2004. Nonsense-mediated decay: terminating erroneous
925 gene expression. *Current Opinions in Cell Biology* 16:293–299.
926 https://doi.org/10.1016/j.ceb.2004.03.003
927
928 Barkan A, Small I. 2014. Pentatrichopeptide repeat proteins in plants. *Annual*
929 *Review of Plant Biology* 65:415–442. https://doi.org/10.1146/annurev-
930 arplant-050213-040159
931
932 Benne R, Ven Den Berg J, Brakenhoff JP, Sloor P, Van Boom JH, Tromp MC.
933 1986. Major transcript of the frameshifted *coxII* gene from
934 trypanosome mitochondria contains four nucleotides that are not
935 encoded in the DNA. *Cell* 46:819–826. https://doi.org/10.1016/0092-
936 8674(86)90063-2
937
938 Bentolila S, Oh J, Hanson MR, Bukowski R. 2013. Comprehensive high-
939 resolution analysis of the role of an *Arabidopsis* gene family in RNA
940 editing. *PLoS Genetics* 9:e1003584.
941 https://doi.org/10.1371/journal.pgen.1003584
942
943 Berget SM, Sharp PA. 1977. A spliced sequence at the 5'-terminus of
944 adenovirus late mRNA. *Brookhaven Symposium of Biology* 29:332–
945 344. https://doi.org/10.1073/pnas.74.8.3171
946
947 Birk MA, Liscovitch-Brauer N, Dominguez MJ, McNeme S, Yue Y, Hoff JD,
948 Twersky I, Verhey JK, Sutton RB, Eisenberg E, Rosenthal JJ. 2023.
949 Temperature-dependent RNA editing in octopus extensively recodes
950 the neural proteome. *Cell* 186:2544–2555.
951 https://doi.org/10.1016/j.cell.2023.05.004
952
953 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for
954 Illumina sequence data. *Bioinformatics* 30:2114–2120.
955 https://doi.org/10.1093/bioinformatics/btu170
956
957 Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and
958 computing of summary statistics. *PeerJ* 4:e1660.
959 https://doi.org/10.7717/peerj.1660
960
961 Chen H, Deng L, Jiang Y, Lu P, Yu J. 2011. RNA editing sites exist in protein-
962 coding genes in the chloroplast genome of *Cycas taitungensis*.
963 *Journal of Integrative Plant Biology.* 53:961–970.
964 https://doi.org/10.1111/j.1744-7909.2011.01082.x
965

966 Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence
967 arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*
968 12:1–8. https://doi.org/10.1016/0092-8674(77)90180-5
969
970 Covello PS, Gray MW. 1993. On the evolution of RNA editing. *Trends in*
971 *Genetics* 9:265–268. https://doi.org/10.1016/0168-9525(93)90011-6
972
973 Crick FH. 1958. On protein synthesis. *Symposium of the Society of*
974 *Experimental Biology* 12:138–163.
975
976 Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of
977 organelle genomes from whole genome data. *Nucleic Acids Research*
978 45:e18. https://doi.org/10.1093/nar/gkw955
979
980 Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics
981 and dating with confidence. *PLoS Biology* 4:e88.
982 https://doi.org/10.1371/journal.pbio.0040088
983
984 Edera AA, Sanchez-Puerta MV. 2021. Computational detection of plant RNA
985 editing events. *Methods in Molecular Biology* 2181:13–34.
986 https://doi.org/10.1007/978-1-0716-0787-9_2
987
988 Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. 2011. Choosing among partition
989 models in Bayesian phylogenetics. *Molecular Biology and Evolution*
990 28:523–532. https://doi.org/10.1093/molbev/msq224
991
992 Fauskee BD, Sigel EM, Pryer KM, Grusz AL. 2021. Variation in frequency of
993 plastid RNA editing within *Adiantum* (Pteridaceae) implies rapid
994 evolution in fern plastomes. *American Journal of Botany* 108:2558–
995 2571. https://doi.org/10.1002/ajb2.1649
996
997 Gernhard T. 2008. The conditioned reconstruction process. *Journal of*
998 *Theoretical Biology* 253:769–778.
999 https://doi.org/10.1016/j.jtbi.2008.04.005
1000
1001 Gilbey L. 2022. ggwaffle: Waffle Charts. *R package version 0.2.5*.
1002
1003 Guo W, Grewe F, Mower JP. 2015. Variable frequency of plastid RNA editing
1004 among ferns and repeated loss of uridine-to-cytidine editing from
1005 vascular plants. *PLoS One* 10:e0117075.
1006 https://doi.org/10.1371/journal.pone.0117075
1007
1008 Guzowska-Nowowiejska M, Fiedorowicz E, Plader W. 2009. *Cucumis sativus*,
1009 melon, pumpkin, and squash: are rules of editing in flowering plants'

chloroplast genes so well known indeed? *Gene* 434:1–8. https://doi.org/10.1016/j.gene.2008.12.017

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. https://doi.org/10.1093/sysbio/syw021

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. https://doi.org/10.1371/journal.pone.0074736

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. https://doi.org/10.12705/673.3

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. https://doi.org/10.1093/pcp/pcz111

Jeffreys H. 1961. Some tests of significance, treated by the theory of probability. *Theory of Probability (3rd ed.)* Oxford, U.K.: 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. http://dx.doi.org/10.4238/2012.April.19.4

Jombart T, Balloux F, Dray S. 2010. Adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26:1907–1909. https://doi.org/10.1093/bioinformatics/btq292

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. https://doi.org/10.1007/s00239-005-0254-5

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular*

Biology and Evolution 30:772–780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kendall DG. 1948. On the generalized "birth-and-death" process. *Annals of Mathematical Statistics* 19:1–15. https://doi.org/10.1214/aoms/1177730285

Khanna A, Larson DE, Srivatsan SN, Mosoir M, Abbott TE, Kiwala S, Ley TJ, Duncavage EJ, Walter MJ, Walker JR, Griffith OL. 2022. Bam-readcount—rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Softwre* 7:3722 https://doi.org/10.21105/joss.03722

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. https://doi.org/10.1186/s12862-016-0707-z

Knoop V. 2011. When you can't trust the DNA: RNA editing changes transcript sequences. *Cellular and Molecular Life Sciences* 68:567–586. https://doi.org/10.1007/s00018-010-0538-9

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. https://doi.org/10.1093/nar/gkg327

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. https://doi.org/10.1002/ajb2.1152

Langmead B, Salzburg SL. 2012. Fast gapped-read alignment with Bowtie2. *Nature Methods* 9:357–359. https://doi.org/10.1038/nmeth.1923

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete character data. *Systematic Biology* 50:913–925. https://doi.org/10.1080/106351501753462876

Li FW, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux PM, Eily A, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4:460–472. https://doi.org/10.1038/s41477-018-0188-8

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

Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. 2011. How a neutral ratchet can build cellular complexity. *IUBMB Life* 63:528–537. https://doi.org/10.1002/iub.489

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. https://doi.org/10.1093/molbev/msaa015

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. https://doi.org/10.1007/s11120-017-0469-5

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. https://doi.org/10.1078/0176-1617-01220

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. https://doi.org/10.1098/rstb.1994.0068

Nitta JH, Schuettpelz E, Ramírez-Barahona S, Iwasaki W. 2022. An open and continuously updated fern tree of life. *Frontiers in Plant Sciences* 13:909768. https://doi.org/10.3389/fpls.2022.909768

O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. 2008. On the expansion of the pentatrichopeptide repeat

gene family in plants. *Molecular Biology and Evolution* 25:1120–1128. https://doi.org/10.1093/molbev/msn057

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. https://doi.org/10.1261/rna.045575.114

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. https://doi.org/10.1093/bioinformatics/bty633

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 https://doi.org/10.1002/aps3.11617

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, Fitzjohn RG, Alfaro ME, Harmon LK. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Pinheiro J, Bates D, R Core Team. 2024. nlme: Linear and Nonlinear Mixed Effects Models. *R package version 31.166*. 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. https://doi.org/10.1093/sysbio/syy032

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

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. https://doi.org/10.1093/gbe/evy189

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. https://doi.org/10.1093/sysbio/syt058

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. https://doi.org/10.1093/molbev/msn084

Schafran P, Hauser DA, Nelson JM, Xu X, Mueller LA, Kulshrestha S, Smalley, I, de Vries S, Irisarri I, de Vries J, Davies K, Villareal JCA, Li FW. Pan-phylum genomes of hornworts reveal conserved autosomes but dynamic accessory and sex chromosomes. *Nature Plants* 11:49–62. https://doi.org/10.1038/s41477-024-01883-w

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. https://doi.org/10.1080/10635150600755438

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. https://doi.org/10.12705/654.2

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. https://doi.org/10.1007/BF00178601

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. https://doi.org/10.1111/tpj.14578

Steel M, Huson D, Lockhart PJ. 2000. Invariable sites models and their use in phylogenetic reconstruction. *Systematic Biology* 49:225–232. https://doi.org/10.1093/sysbio/49.2.225

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. https://doi.org/10.1016/j.ympev.2016.09.003

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. https://doi.org/10.1371/journal.pone.0112963

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. https://doi.org/10.1371/journal.pone.0140285

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. https://doi.org/10.1016/j.gene.2004.06.018

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. https://doi.org/10.1093/sysbio/syq085

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 39:306–314. https://doi.org/10.1007/BF00160154