Letter

Unparalleled Variation in RNA Editing among *Selaginella* Plastomes[1][OPEN]

David Roy Smith[2,3]

Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada

ORCID ID: 0000-0001-9560-5210 (D.R.S.).

Dear Editor,

One of the most extreme documented examples of chloroplast RNA editing comes from the seedless vascular plant *Selaginella uncinata* (Lycopodiophyta), for which an astonishing 3494 cytosine-to-uracil editing events have been discovered (Oldenkott et al., 2014). Is posttranscriptional chloroplast editing as rampant in other *Selaginella* species? Here, I examine plastome-wide RNA editing profiles for *Selaginella kraussiana* and *Selaginella lepidophylla* and report that the number and position of edited sites can be extremely variable among *Selaginella* plastomes, to a degree that is currently unparalleled in any other photosynthetic genus.

RNA editing sites were identified by mapping publicly available Illumina RNA sequencing (RNA-seq) reads from *S. kraussiana* (GenBank accessions SRR2045379–82) and *S. lepidophylla* (SRR6345606–15) onto the respective chloroplast genome sequences of these two lycophytes (Supplemental Materials and Methods; Mower et al., 2019). For each species, the RNA and plastome sequencing data came from the same cultivar (and laboratory; Ge et al., 2016; VanBuren et al., 2018), greatly reducing the potential of mistaking polymorphisms between specimens as editing events. Mapping of the RNA-seq reads gave near-complete coverage (98%) of the reference chloroplast genomes, including all genes. Mean coverage of the plastomes exceeded 500×, providing robust alignments for identifying edited sites, which were only characterized in regions with ≥ 5× coverage and ≥ 25% read support (Supplemental Materials and Methods); thus, keep in mind that sites with low editing efficiency (< 25%) were not recorded in this study.

A total of 1353 and 720 C-to-U changes, respectively, were identified in the *S. kraussiana* and *S. lepidophylla* chloroplast transcriptomes (Table 1; Supplemental Tables S1 and S2), making them the most heavily RNA-edited plastomes from the Viridiplantae (Ichinose and Sugita, 2016), outdone only by that of *Selaginella uncinata*. Approximately 80% of the observed edits from the two plastomes occurred in protein-coding regions and included synonymous and nonsynonymous changes as well as many instances in which start and/or stop codons were restored (Table 1; Supplemental Tables S3 and S4). The remainder of the edits (~20%) were restricted to intergenic and intronic segments (Table 1), meaning not a single change was recorded in ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs), and no U-to-C changes were found, which parallels the editing data from *S. uncinata* (Oldenkott et al., 2014).

The huge amount of RNA editing in *Selaginella* plastomes is striking, but equally remarkable is the variation in the number of edited sites among species. Indeed, The *S. uncinata* plastome has ~2150 and ~2775 more C-to-U alterations than its *S. kraussiana* and *S. lepidophylla* counterparts. In other words, there is a 2- to 5-fold difference in plastome editing across these three taxa—and that is likely an underestimate, as only 1139 nucleotides (nt) of intergenic chloroplast RNA from *S. uncinata* have been surveyed for editing (Oldenkott et al., 2014). To the best of my knowledge, this is the largest reported difference in chloroplast editing for any genus studied to date—but see Klinger et al. (2018) for other extreme examples.

The variation in chloroplast editing is also reflected in the location of C-to-U changes within the *Selaginella* plastomes, as well as in which protein-coding transcripts are (or are not) edited and in the relative number of editing sites in those genes (Supplemental Tables S1–S4). Alignments of the protein-coding chloroplast DNA and RNA from the three *Selaginella* species (Supplemental Materials and Methods) showed that more than 40% of the edits identified in *S. kraussiana* and *S. lepidophylla* are unique—i.e. the C-to-U change was found in only one of the species. Thus, the total pool of uniquely edited sites currently identified within the *Selaginella* genus easily exceeds 2500, not including edits within the intergenic regions, which could not be aligned.

Some similarities in the editing patterns were also observed (Table 1; Supplemental Tables S3 and S4). For example, in *S. uncinata*, *S. kraussiana*, and *S. lepidophylla*, the gene encoding the D1 protein of PSII (*psbA*), which

[1]This work was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant to D.R.S.).

[2]Author for contact: dsmit242@uwo.ca

[3]Senior author.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: David Roy Smith (dsmit242@uwo.ca).

[OPEN]Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.19.00904

Table 1. *Cytosine-to-uracil RNA editing in the plastomes of* Selaginella lepidophylla *(*Sl*),* S. kraussiana *(*Sk*), and* S. uncinata *(*Su*)*

Genomic Feature	*Sl*	*Sk*	*Su*[b]
Total no. of editing sites[a]	720	1353	>3494
Editing sites in protein-coding regions	581	1,104	3,415
Nonsynonymous substitutions	530	972	2987
Synonymous substitutions	51	132	428
Start codon restoration	22	18	52
Stop codon restoration	9	12	31
Most edited gene (no. sites)	*rpoB* (65)	*ccsA* (66)	*rpoB* (214)
Editing sites in intergenic regions	128	236	>5[c]
Editing sites in intronic regions	11	13	74[c]
Plastome size (kb)	114.69	129.97	144.17
Plastome GC content	51.9	52.3	54.8

[a]RNA editing sites in the large inverted (or direct) repeat region were counted only once. [b]Data from Oldenkott et al. (2014). [c]Only 1139 nt of intergenic chloroplast RNA and four introns were surveyed for editing.

is over 1000 nt long, has no detectable editing sites. Likewise, for both *S. uncinata* and *S. lepidophylla*, the chloroplast gene for the beta subunit of RNA polymerase (*rpoB*) gene has the largest number of editing sites (but not so for *S. kraussiana*). For all three species, many of the editing events are clustered close together (Supplemental Tables S1 and S2; Oldenkott et al., 2014).

Together, these data suggest that throughout the evolution and diversification of *Selaginella*, there has been the gain and/or loss of thousands of chloroplast RNA-editing sites and that this process is still ongoing. What's more, the same is probably true for the mitochondrial genome of this genus, which undergoes equal (or even greater) amounts of C-to-U RNA editing than the plastome (Smith, 2009; Hecht et al., 2011; Oldenkott et al., 2014). I did try to mine mitochondrial transcripts from the *S. kraussiana* and *S. lepidophylla* RNA-seq datasets but was unsuccessful.

As more *Selaginella* species are investigated, the breadth of the variation in RNA editing is sure to grow. Preliminary analyses of the *Selaginella moellendorffii* chloroplast genome suggest that it has at least 1800 C-to-U modifications (Oldenkott et al., 2014), more than those of *S. kraussiana* and *S. lepidophylla*. Not surprisingly, there appears to be a positive relationship between *Selaginella* plastome GC content, which is among the highest of any lineage, and the number of C-to-U editing sites (Table 1; Smith, 2009). Thus, to capture the complete range of chloroplast RNA editing, it might be useful to target species that are predicted to have very high plastome GC contents, such as *Selaginella fragilis*, as well as those with much lower predicted GC compositions, like *S. sinensis* (Smith, 2009).

No matter how large the variation turns out to be, the question remains: why do *Selaginella* plastomes (and mitogenomes) undergo such extensive C-to-U editing? The evolutionary origins of RNA editing in organelle systems can be eloquently explained by the concept of constructive neutral evolution, which "posits that the biochemical elements of an RNA editing system must be in place before there is an actual need for editing" (Gray, 2012). Among the key players in plant organelle RNA editing are nuclear-encoded, aspartic acid-tyrosine-tryptophan (DYW)-domain-containing pentatricopeptide repeat (PPR) proteins, some of which are known to be site recognition factors for editing events (Ichinose and Sugita, 2018). In land plants, the size and diversity of DYW-type PPR gene families appears to be positively associated with the abundance of organelle RNA editing (Rüdinger et al., 2012). And, as one might expect, the *S. moellendorffii* genome (Banks et al., 2011) encodes an expanded DYW-type PPR protein family: ~312 members (Cheng et al., 2016). As more data become available, it will be particularly interesting to compare variation in the number of PPR proteins from *Selaginella* species with chloroplast RNA editing abundance and to take advantage of bioinformatics programs that use the PPR-RNA binding code to predict binding events (Harrison et al., 2016; Yan et al., 2019). I anticipate that species with the largest number of RNA editing sites will have the most expanded PPR protein gene families and predicted PPR binding events, and vice versa.

Plant organelle RNA editing is a burgeoning field and was recently implicated in chloroplast-to-nucleus communication (Zhao et al., 2019), opening new research avenues. Given its status as a model lineage and its unrivaled number and diversity of chloroplast C-to-U editing sites, *Selaginella* is well positioned to become a leading system for studying posttranscriptional organelle editing. I look forward to seeing what future work will uncover.

Accession Numbers

All accession numbers used in this study are listed in the Supplemental Materials and Methods.

Supplemental Data

The following supplemental materials are available.

Supplemental Materials and Methods. Detailed information on experimental procedures.

Supplemental Table S1. Chloroplast RNA editing in the plastome of *S. kraussiana*.

Supplemental Table S2. Chloroplast RNA editing in the plastome of *S. lepidophylla*.

Supplemental Table S3. Chloroplast RNA editing in the protein-coding transcripts of *S. kraussiana*.

Supplemental Table S4. Chloroplast RNA editing in the protein-coding transcripts of *S. lepidophylla*.

Received July 23, 2019; accepted August 17, 2019; published September 3, 2019.

LITERATURE CITED

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, et al (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science **332:** 960–963

Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, Castleden I, Song Y, Song B, Huang J, et al (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J **85:** 532–547

Ge Y, Liu J, Zeng M, He J, Qin P, Huang H, Xu L (2016) Identification of WOX family genes in *Selaginella kraussiana* for studies on stem cells and regeneration in lycophytes. Front Plant Sci **7:** 93

Gray MW (2012) Evolutionary origin of RNA editing. Biochemistry **51:** 5235–5242

Harrison T, Ruiz J, Sloan DB, Ben-Hur A, Boucher C (2016) aPPRove: An HMM-based method for accurate prediction of RNA-pentatricopeptide repeat protein binding events. PLoS One **11:** e0160645

Hecht J, Grewe F, Knoop V (2011) Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: The root of frequent plant mtDNA recombination in early tracheophytes. Genome Biol Evol **3:** 344–358

Ichinose M, Sugita M (2016) RNA editing and its molecular mechanism in plant organelles. Genes (Basel) **8:** E5

Ichinose M, Sugita M (2018) The DYW domains of pentatricopeptide repeat RNA editing factors contribute to discriminate target and non-target editing sites. Plant Cell Physiol **59:** 1652–1659

Klinger CM, Paoli L, Newby RJ, Wang MY, Carroll HD, Leblond JD, Howe CJ, Dacks JB, Bowler C, Cahoon AB, et al (2018) Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. Genome Biol Evol **10:** 1019–1038

Mower JP, Ma PF, Grewe F, Taylor A, Michael TP, VanBuren R, Qiu YL (2019) Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. New Phytol **222:** 1061–1075

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

Rüdinger M, Volkmar U, Lenz H, Groth-Malonek M, Knoop V (2012) Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. J Mol Evol **74:** 37–51

Smith DR (2009) Unparalleled GC content in the plastid DNA of *Selaginella*. Plant Mol Biol **71:** 627–639

VanBuren R, Wai CM, Ou S, Pardo J, Bryant D, Jiang N, Mockler TC, Edger P, Michael TP (2018) Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. Nat Commun **9:** 13

Yan J, Yao Y, Hong S, Yang Y, Shen C, Zhang Q, Zhang D, Zou T, Yin P (2019) Delineation of pentatricopeptide repeat codes for target RNA prediction. Nucleic Acids Res **47:** 3728–3738

Zhao X, Huang J, Chory J (2019) GUN1 interacts with MORF2 to regulate plastid RNA editing during retrograde signaling. Proc Natl Acad Sci USA **116:** 10162–10167