12-2019

Exploring the Physiology and Evolution of Hornworts

Tanner A. Robison
Utah State University

EXPLORING THE PHYSIOLOGY AND EVOLUTION OF HORNWORTS

by

Tanner A. Robison

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Biology

Approved:

_____	_____
Paul Wolf, Ph.D.	Keith Mott, Ph.D.
Major Professor	Committee Member
_____	_____
Bruce Bugbee, Ph.D.	Richard S. Inouye, Ph.D.
Committee Member	Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

ABSTRACT

Exploring the Evolution and Physiology of Hornworts

by

Tanner Robison, Master of Science

Utah State University, 2019

Major Professor: Dr. Paul G. Wolf
Department: Biology

To improve the ease, accuracy, and speed of organellar genome annotation for hornworts and ferns, we developed a bioinformatic tool which takes into account the process of RNA editing when examining annotations. This software works by checking the coding sequences of annotated genes for internal stop codons and for overlooked start codons which might be the result of RNA editing. If these codons are determined to be the result of RNA editing, existing annotations are edited and the appropriate additional annotations are made. This work directly aided in the processing of a large number of plastid genomes in the family Pteridaceae. The annotation and analysis of the plastomes in Pteridaceae revealed a suite of highly mobile genetic elements, composed of 3 open reading frames. These mobile genetic elements have only limited homology to a select

few algal sequences, but are not found in any plants outside of Polypodiophyta. Additionally, these mobile elements appear to be driving structural evolution in the plastomes of the family of Pteridaceae. Further analysis of these mobile elements revealed that they can be found across the diversity of Polypodiophyta, including some of its earliest diverging clades, suggesting that they have been driving these changes across the clade for much of its history.

(82 pages)

PUBLIC ABSTRACT

Exploring the Evolution and Physiology of Hornworts

Tanner A. Robison

Plants contain organelles called chloroplasts, which is where photosynthesis takes place. Chloroplasts also contain their own DNA, which is separate from the DNA in the nucleus. This DNA does not change much over evolutionary time, so it can be used to investigate relationships between organisms. Here we created a tool that makes it easier to analyze this chloroplast DNA as well making it easier to share complete chloroplast genomes on public databases. In addition, we also found a mobile element in the chloroplast DNA of a group of ferns, which appears to be driving structural changes in their genomes.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Chloroplasts, the organelles in which photosynthesis takes place, are also host to a small, circular, and highly gene dense genome, called the plastome. This gene density, along with the critical nature of the few genes that are encoded (<100), means that plastomes experience low rates of nucleotide substitution in comparison to other genomes (Wolfe et al. 1987). In many respects, the plastomes of land plants and algae actually have more in common with cyanobacteria than they do with the genomes of eukaryotes in which they reside. This, along with a similar observation in the mitochondrial genome has led ride to endosymbiotic theory -- the notion that the organelles that distinguish eukaryotes from all other life are, in fact, of prokaryotic origin. While plastomes do resemble the genomes of bacteria, they are highly reduced in size -- roughly one twentieth the size (Nakayama & Archibald 2012). This reduction in size is likely due to the fact that after the endosymbiotic event, the host began to simultaneously take up genes from its symbiont, in a process known as endosymbiotic gene transfer, while also importing protein products into the symbiont, thus relaxing selection on the maintenance on those genes (Nakayama & Archibald 2012; McFadden 2001). This process is still occuring (Ayliffe & Timmis 1992; Huang et al. 2003; Shahmuradov et al. 2003; Stegemann et al. 2003; Matsuo et al. 2005) and it is estimated that endosymbiotic gene transfer was so widespread that ~18% of genes encoded by the nucleus are of proto-plastid origin (Martin et al. 2002; Martin & Herrmann 1998).

While there are many examples of gene transfer occurring from chloroplast to nucleus, and chloroplasts to mitochondria, there are very few examples of genes transferring to the plastome (Timmis et al. 2004). The chloroplast appears to be highly resistant to accepting forgein DNA, with the gene content of all land plants being nearly identical (Mower & Vickrey 2018). In addition to this conserved gene content, gene order in plastomes is also highly conserved among land plants, so much so that the plastomes of land plants, spanning hundreds of millions of years of evolution, are nearly collinear (Mower & Vickrey 2018). This, along with their above noted low substitution rates, makes plastomes extremely powerful tools for phylogenetic analysis, and has resulted in profound insights into the evolution of plants (Taberlet et al. 1991; Chaw et al. 2004; Yoon et al. 2004; Pryer et al. 2004; Shaw et al. 2005; Leliaert et al. 2012; Gitzendanner et al. 2018).

Despite this general pattern of conserved gene order and content, there are a few lineages that show profound reconfiguration in their plastomes (Knox 2014; Chumley et al. 2006; Guisinger et al. 2011; Cai et al. 2008; Haberle et al. 2008). Ferns in particular have undergone relatively more genomic rearrangements than most lineages (Labiak & Karol 2017; Zhu et al. 2016; Stein et al. 1992; Wolf et al. 2015), but there has been little work to try to pinpoint why these changes occur. In addition, there also exist gaps in the taxonomic sampling of ferns, with most studies focusing on either deeply divergent groups of ferns (Gao et al. 2011; Wei et al. 2017), or very closely related taxa (Labiak & Karol 2017). There has yet to be a family scale investigation into the plastomes of ferns,

thus not only leaving gaps in our phylogenetic understanding of ferns, but also in our understanding of how their plastomes change on a finer scale.

I worked to fill both those knowledge gaps in a study of the fern family Pteridaceae. Pteridaceae is among the most diverse group of ferns, accounting for more than 10% of species in ferns (PPG I 2016). One particular group within Pteridaceae, called 'vittariod' ferns (Rothfels & Schuettpelz 2014; Grusz et al. 2016), has shown particularly striking changes in molecular evolutionary rates based multi gene analysis. Additionally, the enigmatic nature of Vittarioideae, makes a sampling and analysis of Pteridaceae a perfect case in which to investigate plastome evolution. In our study of Pteridaceae, not only do we improve the phylogenetic resolution of the family, but we also uncover the presence of a mobile element in the plastomes of Pteridaceae, which seems to not only be driving genomic rearrangements in Pteridaceae, but also among all ferns. These mobile elements also appear to be moving among the various genomic compartments of ferns (nucleus, plastome, mitochondria), and are driving the transfer of other genes as well.

RNA editing is a well-documented, but still incompletely understood process whereby the transcribed mRNA sequence differs from that predicted by the encoded genomic DNA sequence. There are several types of RNA editing, but in plants there are two main types U -> C RNA editing and C -> U. In most land plants, these edits occur in the chloroplast genome (plastome) at relatively low rates (~35 editing sites per plastome).

In ferns and hornworts, however, these editing sites can be extremely numerous (Wolf et al. 2004; Kugita et al. 2003).

This presents significant problems to researchers who wish to publish the plastomes of plants with high levels of editing on public databases, namely, many of these databases will not accept the genomic sequences of such organisms because it appears as though they have invalid protein annotations, containing issues such as internal stops or missing start codons. To make these sequences acceptable for publication, researchers must manually add annotations for each RNA editing site, which is extremely time consuming. With the advent of next generation sequencing, the cost of sequencing has reduced and the scale at which sequencing can be done has increased, resulting in an explosion in the number of genomes being sequenced. Thus, the need for a way for researchers to rapidly annotate and assess RNA editing sites in a particular genome is increasingly important.

In addition to the technical problems that RNA editing presents, it also poses interesting evolutionary problems. RNA editing occurs in organisms across the tree of life, from humans, to viruses, to plants (Su and Randau 2011; Steward et al. 1993; Li et al. 2009; Takenaka et al. 2013). In some cases, such as in humans, the purpose of such processes is clear: it allows for the efficient regulation of genes. The process of RNA editing allows for humans to terminate the translation of a gene early in certain cell types, while continuing the process of translation in others (Grohmann et al. 2010). This results in two different proteins being produced by one gene, which is in theory more

metabolically efficient. Where the value of such a process breaks down is in organisms which have RNA editing in extremely high levels, like ferns and hornworts, where the process seems to happen more or less randomly within a given gene and genome. In the case of the chloroplast genome, we know that RNA editing does not serve the same regulatory function that it does in human cells, but it is not clear whether the process is an adaptation of some kind or whether it is the result of relaxed selection caused by RNA editing factors (Takenaka et al. 2013). The latter is the predominant theory regarding the evolution of PPR mediated RNA editing, where RNA editing arises in plants through a process known as 'constructive neutral evolution' (Takenaka et al. 2013; Stoltzfus 1999). In the constructive neutral evolution theory, the process of RNA editing evolved first, resulting in relaxed selection on the specific nucleotide sites that it can edit, which then drifted to different nucleotides, resulting in the requirement of RNA editing (Takenaka et al. 2013). In a sense, RNA editing is solving a problem that it created. While this theory makes sense, it does little to explain why some lineages have near pathogenic levels of RNA editing while others have low or even no RNA editing. Both ferns and hornworts would be useful study systems for exploring the evolution of RNA editing in plants. However, such studies have not been conducted for three possible reasons: (1) until very recently, there simply has not been enough phylogenetic depth or breadth covering either ferns or hornworts, (2) while tools exist to predict RNA editing, there remains some level of uncertainty in those predicted sites, even when supplemented by RNA seq data, and (3) as cited earlier in this section, there is not a rapid, consistent method for annotating

such sites, making exploring their variation extremely time consuming. Thus, developing software to annotate these sites presents a potential avenue to not only improve the speed with which researchers can publish future genomes, but also to explore the evolution of RNA editing in lineages that have high levels of editing. In chapter 2, I describe software which developed to do just this. The software, ReFernment, surveys nonsense mutations in fern and hornwort organelles and determines whether those mutations could be the result of RNA editing and, if so, annotates those sites appropriately.

REFERENCES

Ayliffe MA, Timmis JN. 1992. Plastid DNA sequence homologies in the tobacco nuclear genome. Mol. Gen. Genet. 236:105–112.

Cai Z et al. 2008. Extensive reorganization of the plastid genome of Trifolium subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J. Mol. Evol. 67:696–704.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the Monocot–Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. J. Mol. Evol. 58:424–441.

Chumley TW et al. 2006. The complete chloroplast genome sequence of Pelargonium x hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. 23:2175–2190.

Gao L, Zhou Y, Wang Z-W, Su Y-J, Wang T. 2011. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. BMC Plant Biol. 11:64.

Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am. J. Bot. doi: 10.1002/ajb2.1048.

Grusz AL, Rothfels CJ, Schuettpelz E. 2016. Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. BMC Genomics. 17:692.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28:583–600.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. J. Mol. Evol. 66:350–361.

Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature. 422:72–76.

Knox EB. 2014. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. Proc. Natl. Acad. Sci. U. S. A. 111:11097–11102.

Labiak PH, Karol KG. 2017. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. Am. J. Bot. doi: 10.3732/ajb.1700135.

Leliaert F et al. 2012. Phylogeny and molecular evolution of the green algae. CRC Crit. Rev. Plant Sci. 31:1–46.

Martin W et al. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U. S. A. 99:12246–12251.

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol. 118:9–17.

Matsuo M, Ito Y, Yamauchi R, Obokata J. 2005. The Rice Nuclear Genome Continuously Integrates, Shuffles, and Eliminates the Chloroplast Genome to Cause Chloroplast–Nuclear DNA Flux. Plant Cell. 17:665–675.

McFadden GI. 2001. Chloroplast origin and integration. Plant Physiol. 125:50–53.

Mower JP, Vickrey TL. 2018. Chapter Nine - Structural Diversity Among Plastid Genomes of Land Plants. In: Advances in Botanical Research. Chaw, S-M & Jansen, RK, editors. Vol. 85 Academic Press pp. 263–292.

Nakayama T, Archibald JM. 2012. Evolving a photosynthetic organelle. BMC Biol. 10:35.

PPG I. 2016. A community-derived classification for extant lycophytes and ferns: PPG I. Jnl of Sytematics Evolution. 54:563–603.

Pryer KM et al. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. 91:1582–1598.

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

Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA. 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and Arabidopsis. Plant Mol. Biol. 52:923–934.

Shaw J et al. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92:142–166.

Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U. S. A. 100:8828–8833.

Stein DB et al. 1992. Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proc. Natl. Acad. Sci. U. S. A. 89:1856–1860.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123–135.

Wei R et al. 2017. Plastid Phylogenomics Resolve Deep Relationships among Eupolypod II Ferns with Rapid Radiation and Rate Heterogeneity. Genome Biol. Evol. 9:1646–1657.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U. S. A. 84:9054–9058.

Wolf PG et al. 2015. An Exploration into Fern Genome Space. Genome Biol. Evol. 7:2533–2544.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21:809–818.

Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New

Phytol. 209:1747–1756.

CHAPTER 1

MOBILE ELEMENTS SHAPE PLASTOME EVOLUTION IN FERNS

Mobile elements shape plastome evolution in ferns

Tanner A. Robison[1*], Amanda L. Grusz[2,3], Paul G. Wolf[l], Jeffrey P. Mower[4], Blake D. Fauskee[2], Karla Sosa[5], and Eric Schuettpelz[3]

[1]Department of Biology, Utah State University, Logan UT, USA

[2]Department of Biology, University of Minnesota Duluth MN, USA

[3]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington DC, USA

[4]Center for Plant Science Innovation, University of Nebraska, Lincoln NE, USA

[5]Department of Biology, Duke University, Durham, NC, USA

*Corresponding author: Tanner Robison, Email: robison.tanner@gmail.com, Telephone number : 435-881-9600

Abstract

Plastid genomes display remarkable organizational stability over evolutionary time. From green algae to angiosperms, most plastid genomes are largely collinear, with only a few cases of inversion, gene loss, or, in extremely rare cases, gene addition. These plastome insertions are mostly clade-specific and are typically of nuclear or mitochondrial origin. Here, we expand on these findings and present the first family-level survey of plastome evolution in ferns, revealing a novel suite of dynamic mobile elements. Comparative plastome analyses of the Pteridaceae expose several mobile open reading frames that vary in sequence length, insertion site, and configuration among sampled taxa. Even between close relatives, the presence and location of these elements is widely variable when viewed in a phylogenetic context. We characterize these elements and refer to them collectively as Mobile Open Reading Frames in Fern Organelles (MORFFO). We further note that the presence of MORFFO is not restricted to Pteridaceae, but is found across ferns and other plant clades. MORFFO elements are regularly associated with inversions, intergenic expansions, and changes to the inverted repeats. They likewise appear to be present in mitochondrial and nuclear genomes of ferns, indicating that they can move between genomic compartments with relative ease. The origins and functions of these mobile elements are unknown, but MORFFO appears to be a major driver of structural genome evolution in the plastomes of ferns, and possibly other groups of plants.

Introduction

Plastid genomes (plastomes) are a rich source of molecular sequence data and have proven to be especially useful in explorations of plant evolutionary history. From single-gene analyses to full plastome phylogenomics, important evolutionary insights can be gleaned from these relatively small, highly conserved, and minimally repetitive chromosomes (Taberlet et al. 1991; Chaw et al. 2004; Yoon et al. 2004; Pryer et al. 2004; Shaw et al. 2005; Leliaert et al. 2012; Ruhfel et al. 2014; Givnish et al. 2010; Moore et al. 2010; Gitzendanner et al. 2018). Plastomes contain high proportions of protein coding genes compared to plant nuclear genomes, with many of these genes being essential to photosynthesis (Wicke et al. 2011). Consequently, plastomes experience relatively low nucleotide substitution rates, especially in the inverted repeats, making them extremely stable over evolutionary time (Wolfe et al. 1987; Li et al. 2016; Zhu et al. 2016).

The plastomes of land plants seem to be especially resistant to changes in gene content, which, along with gene order, generally varies little between distantly related lineages—even after hundreds of millions of years (Palmer 1985). Relatively few plastome genes have been lost, except in heterotrophic lineages in which photosynthetic genes are typically not required for survival (Bungard 2004). Even rarer is the acquisition of new genes (Timmis et al. 2004). Since the gain of *ycf1* and *ycf2* in the algal ancestors, very few new genes have been incorporated into land plant plastomes (Timmis et al. 2004; de Vries et al. 2015). However, some groups do show exceptional variability in

plastome structure, even among closely related taxa (Chumley et al. 2006; Guisinger et al. 2011; Cai et al. 2008; Haberle et al. 2008; Hirao et al. 2008). Notable among these exceptional lineages is Campanulaceae, within which a prolific group of inserted ORFs (Open Reading Frames) appear to have driven over 125 large inversions across the family (Knox 2014).

Overall stability in plastome structure across land plants contrasts, strikingly, with punctuated and/or persistent genomic rearrangements that are apparent in certain lineages (Chumley et al. 2006; Guisinger et al. 2011; Cai et al. 2008; Haberle et al. 2008; Hirao et al. 2008; Mower & Vickrey 2018). Ferns are among these, showing evidence of multiple genomic inversions since their initial diversification (Labiak & Karol 2017; Zhu et al. 2016; Stein et al. 1992; Wolf et al. 2015). An increasing number of fern plastome sequences are beginning to reveal a dynamic organellar genome, shaped in large part by structural inversions and/or shifts in gene content from single copy regions into the inverted repeat (Wolf et al. 2010). Genomic inversions like these have been associated with shifts in molecular evolutionary rate (Li et al. 2016; Zhu et al. 2016; Blazier et al. 2016) and may be moderated by selective constraints related to gene synteny and gene expression (Wicke et al. 2011; Cui et al. 2006).

Despite the great strides that have been made in our understanding of plastome evolution in ferns over the last decade (Mower & Vickrey 2018), dense taxonomic sampling is almost always lacking. Instead, studies of plastome evolution in ferns have focused on disparate groups of deeply divergent taxa (Gao et al. 2011; Wei et al. 2017),

or on comparisons of only a few closely related species (Labiak & Karol 2017). Here, we aim to bridge this gap with the first family-scale comparative analysis of plastome structure and content in ferns.

Our focus is on Pteridaceae, an early-diverging family of polypod ferns that comprises roughly 1,200 species and accounts for well over 10% of extant fern diversity (PPG I 2016). The family is cosmopolitan in distribution and occupies a wide array of niches, from shaded forests to xeric and even aquatic habitats (Tryon 1990). Members exhibit a range of reproductive modes and some groups are noteworthy for undergoing frequent whole genome duplication. Among the most striking evolutionary patterns in the family is a dramatic shift in molecular evolutionary rate that has been documented across plastid and nuclear genomes of the so-called "vittarioid" ferns (Rothfels & Schuettpelz 2014; Grusz et al. 2016). In this study, we leverage genome skimming data to assemble and analyze 27 new plastomes from across the Pteridaceae in an effort to: (1) examine plastome variation in Pteridaceae; (2) gain insight into genomic shifts within members of the Vittarioideae; and (3) reevaluate the phylogenetic relationships among major lineages comprising the family. Our data expose a massive plastome inversion and a group of mobile elements—newly characterized here—that appear to be a particularly dynamic component of fern plastomes, as evidenced from within Pteridaceae and beyond.

Materials and Methods

Taxon Sampling

Taxonomic sampling included 29 ingroup species, representing all major clades within the Pteridaceae (Schuettpelz et al. 2007), as well as three outgroup taxa (Table 4). Increased sampling from within subfamily Vittarioideae was undertaken in an effort to better understand the molecular evolutionary rate heterogeneity between the two main subclades therein, *Adiantum* and the vittarioid ferns.

DNA Extraction, Library Prep, and Sequencing

Whole genomic DNA for all newly sampled ingroup taxa (27 total) was extracted from silica-dried leaf tissue using the Qiagen DNeasy Plant Mini Kit (Germantown, Maryland) following the manufacturer's protocol. Whole genomic DNA for 26 samples (all except *Vittaria appalachiana*) was sent to the Duke University Center for Genomic and Computational Biology for in-house library preparation and sequencing. There, individual genomic libraries (~300 bp) were prepared using the Kappa Hyper Prep Kit (Wilmington, Massachusetts). In total, 32 samples (26 included in this study) were multiplexed and pooled over one complete flowcell (8 lanes) on the Illumina HiSeq 2000/2500 platform for 125 bp paired-end sequencing. The *V. appalachiana* DNA was sent to BGI (Shenzhen, China) and sequenced on the Illumina HiSeq 2000 platform, generating 5 Gb of 100 bp paired-end reads from an ~800 bp library.

Genome Assembly and Annotation

Raw sequence reads from *V. appalachiana* were assembled using Velvet version

1.2.03 (Zerbino & Birney 2008) according to previously described procedures (Guo et al. 2014; Sigmon et al. 2017). Genome assembly for all other ingroup taxa was performed using NOVOPlasty (Dierckxsens et al. 2017). NOVOPlasty implements a seed-based, *de novo* genome assembly, which can lessen structural assembly biases that may otherwise mask inferred rearrangements. NOVOPlasty employs the seed to retrieve a given sequence from the target genome, which is then extended and circularized (Dierckxsens et al. 2017).

In most cases, the *rbcL* gene from *Adiantum capillus-veneris* (NC_004766; Table 4) was used as a seed sequence, but in select cases, if *rbcL* had inadequate coverage, the entire plastome of *A. capillus-veneris* was used instead. Raw, unfiltered Illumina reads were subsampled to ≤ 30 million reads to reduce memory requirements. The default k-mer of 39 was used unless there was low organellar genome coverage (< 1%), in which case the k-mer was reduced to 23–30. In cases involving long repetitive regions, higher k-mers of 45–55 were used. For *Jamesonia brasiliensis* and *Cheilanthes bolborrhiza*, complete assembly of plastomes was not possible, but we were able to get contigs of considerable size, which have been included in this study.

Following assembly, genomes were annotated in Geneious 11.1 (Kearse et al. 2012)**,** using the gene sequences of *A. capillus-veneris* as a reference. Putative RNA editing sites were annotated to retain conserved open reading frames (Wolf et al. 2003, 2004). Intergenic sequences that differed dramatically from *A. capillus-veneris* were queried against the NCBI Nucleotide database using BLAST (Altschul et al. 1990) to

ensure that they did not result from the false assembly of mitochondrial or nuclear sequences. Assembly errors were further assessed by mapping raw reads to the newly assembled genome using Bowtie2 (Langmead & Salzberg 2012), looking for dips in read depth. Additionally, overall plastome assembly quality was assessed for each sample using Pilon (Walker et al. 2014). In all cases, changes proposed by Pilon were relatively minor (< 10 nucleotides) suggesting that the quality and accuracy from NOVOPlasty assembly was high.

Plastome phylogenomic analyses

In total, 32 plastomes were included in our phylogenomic analyses; 26 were new to this study and the remaining 6 were obtained from Genbank (Table 4). Each annotated plastome was opened in Geneious and all CDS/gene regions were extracted in FASTA format. Using these taxon-specific FASTA files, containing all CDS/gene regions, we then compiled a FASTA file for each locus. Sequences for each region were aligned using MAFFT 7.394 (Kuraku et al. 2013; Katoh et al. 2017) and alignments of all loci concatenated using Sequence Matrix (Vaidya et al. 2011). The resulting concatenated matrix, comprising 76 loci and 68,047 nucleotide sites, partitioned by gene, was processed through PartitionFinder2 (Lanfear et al. 2017, 2012) on the CIPRES Science Gateway version 3.3 (Miller et al. 2010) using the following settings: branchlengths=linked, models=GTR, GTR+G, GTR+I+G, and model_selection=AICc. PartitionFinder2 identified 36 unique model partitions spanning the concatenated matrix.

Phylogenomic analyses of the concatenated, partitioned dataset were implemented using maximum likelihood and Bayesian optimality criteria on the CIPRES Science Gateway version 3.3 (Miller et al. 2010). Maximum likelihood searches were conducted in RAxML V.8 using multiparametric bootstrapping (-b; 1000 replicates) and our previously described partitioned model (-q). Bayesian inference was performed using MrBayes 3.2.6 and comprised four independent runs, each with four chains (one cold, three heated) and otherwise default (i.e., flat) priors, with the exception that rates of evolution were allowed to vary among loci (ratepr = variable). Chains were run for 10 million generations and trees were sampled from the cold chain every 1,000 generations. To determine at which point the analysis had reached stationarity, the standard deviation of split frequencies among the independent runs (as calculated by MrBayes) was examined and the output parameter estimates were plotted using Tracer v1.6 (Rambaut et al. 2015). Based on convergence diagnostics, the first 2.5 million generations were excluded before obtaining a consensus phylogeny and clade posterior probabilities with the "sumt" command (contype = allcompat).

Characterizing MORFFO elements

To search for MORFFO like sequences in GenBank, all MORFFO insertions found in Pteridaceae were aligned using Geneious, and a consensus sequence was generated. All BLAST (Altschul et al. 1990) (Madden 2013)queries for MORFFO were performed using this consensus sequence. This consensus sequence was also used when

querying both VecScreen (NCBI Resource Coordinators 2017) and RepeatMasker (Smit et al. 2013). Three main search strategies were employed when using NCBI Nucleotide BLAST. Initial queries for MORFFO like sequences were performed with BLASTN using the default parameters. Then, BLASTX and TBLASTN searches were performed using the default parameters and a word size of 3. Additional searches using specific MORFFO sequences rather than a consensus returned equivalent results.

To evaluate the level of selective constraint on the three MORFFO genes, estimates of d_N/d_S were calculated for five species that contained all three MORFFO genes (*Bommeria hispida*, *Hemionitis subcordata*, *Notholaena standleyi*, *Tryonia myriophylla*, and *Vaginularia trichoidea*). First, codon-based alignments were generated using the ClustalW-Codons option in MEGA version 7.0.18 (Kumar et al. 2016). Alignments were trimmed using Gblocks version 0.91b (Castresana 2000) in codon mode with a relaxed set of parameters (t=c, b2=3, b5=half). For each trimmed gene alignment, branchwise estimates of d_N/d_S were calculated for each species using the GA-branch model (Kosakovsky Pond & Frost 2005), which uses a genetic algorithm to optimize the number of d_N/d_S rate classes across the tree and maximum likelihood to optimize branch length and substitution rates, as implemented on the Datamonkey web server (Delport et al. 2010). For the analysis, the HKY rate matrix was chosen as the substitution rate model based on the Datamonkey model selection tool. To evaluate the influence of the tree topology on d_N/d_S estimates, the analysis was run using either a NJ tree or a user-defined tree that matched organismal relationships shown in Figure 1.

Phylogenetic relationships among MORFFO sequences were estimated using maximum likelihood best tree and bootstrap searches, implemented in RAxML V.8 (Stamatakis 2014) with multiparametric bootstrapping (-b; 1000 replicates).

Results

Genome Assembly and Annotation

We assembled and annotated 25 complete plastomes from previously unsampled species, representing all major clades within Pteridaceae (Schuettpelz et al. 2007), plus two partial plastome sequences for *Jamesonia brasiliensis* and *Cheilanthes bolborrhiza* (139,531 bp and 39,380 bp, respectively). The average length of complete plastid genome sequences was 153,153 bp (range 145,327 bp to 165,631 bp) with an average GC content of 41.46% (range 36.7–45.3%; Table 1). Gene content remained largely stable across samples, with no losses of protein coding genes relative to *Adiantum capillus-veneris.* We did, however, detect a loss of *trnT* in all vittarioid ferns sampled (Fig 1), as well as a loss of *trnV* in *Onychium japonicum, Ceratopteris cornuta*, plus all vittarioids with the exception of *Haplopteris elongata*. Across all samples, there were 82 protein coding genes, 33–35 tRNA genes and 4 rRNA genes. Gene order was unchanged across the family—with the exception of a 7,000 bp genomic inversion within the inverted repeats of all vittarioid species except *Vaginularia trichoidea.*

Several plastid DNA insertions were recovered from multiple clades within Pteridaceae. The most prominent of these comprised a suite of genomic insertions, here referred to as Mobile Open Reading Frames in Fern Organelles (MORFFO), that were

detected in most of the plastomes sampled. These MORFFO clusters are characterized by three large and distinct open reading frames (ORFs) that are variably absent, or present in a number of different arrangements (Fig 1). One ~1,300 bp ORF (*morffo1*) is flanked by inverted repeats of ~40 bp that are often in the motif TGT CGA TAG, repeated 3–5 times. The amino acid sequences of *morffo1* do not bear similarity to any characterized proteins in GenBank, but do bear similarity to a hypothetical protein found in the early-diverging fern *Mankuya chejuensis* and the green alga *Roya anglica* (Table 2). A larger ORF (*morffo2)* of ~1,700 bp has domains similar to primases associated with mobile elements in cyanobacteria and archaea (DN_5 superfamily) when queried using BLASTX (Altschul et al. 1990); Table 2). A smaller ORF (*morffo3*) of ~630 bp has no significant similarity to any known genes or proteins; it is found less frequently than the two larger ORFs, but is still prevalent. Often, but not always, *morffo1* is found inserted in frame with *morffo2* to form a larger ORF of ~3,500 bp. Importantly, *morffo1, morffo2,* and *morffo3* can be found in a variety of different arrangements, but when present they are always found immediately adjacent to one another (Fig 1).

The location of MORFFO elements varied across the genomes sampled, being included in the Large Single Copy (LSC), the Inverted Repeat (IR), or the Small Single Copy regions (SSC; Fig 2). As a whole, MORFFO sequences (*morffo1*, *morffo2*, *morffo3*) were similar among species, ranging from 92% to 45% sequence identity (Table 3). For those species with a full set of three MORFFO sequences, the genes appear to encode functional proteins. The coding sequences are intact (no internal stop codons or

frameshifting indels) and exhibit nonsynonymous (d_N) to synonymous (d_S) substitution rate ratios that are consistent with selective constraint ($d_N/d_S << 1$, ranging from 0.17 to 0.43), with the exception of *morffo1* from *Tryonia myriophylla,* which has at least two frameshifting indels and a d_N/d_S approaching 1 (S1 Table). A weak association was observed between the genomic location of MORFFO elements and phylogenetic position among species sampled (Table 3).

Chromosome-wide read depth analyses revealed no shifts in coverage spanning MORFFO insertions or insertion boundaries, indicating that these inserts are not an artifact of genomic library preparation or genome misassembly. Furthermore, MORFFO insertions were detected in *Vittaria appalachiana,* which was sequenced and assembled in a separate lab, using an alternative assembly protocol. We also examined each member of the MORFFO cluster using VecScreen (NCBI Resource Coordinators 2017) and RepeatMasker (Smit et al. 2013), neither of which yielded matches to any known vectors or transposable elements. We searched for MORFFO sequences against current draft assemblies of nuclear genomes of the ferns *Azolla* and *Salvinia* on FernBase (Li et al. 2018) as well as scaffolds for the draft genomes of *Ceratopteris,* and while they were not detected *Azolla or Salvinia*, we did observe the presence of *morffo1* in *Ceratopteris* scaffolds. In addition, we searched for the presence of MORFFO in available transcriptomes from members of Pteridaceae in the 1 kp project (Wickett et al. 2014; Matasci et al. 2014; Xie et al. 2014; Johnson et al. 2012), and found no evidence of transcription of MORFFO. , s.

To test whether MORFFO elements could be of mitochondrial origin, we filtered plastid reads using the mitochondrial option in NOVOPlasty (Dierckxsens et al. 2017), and then assembled the remaining reads using *morffo1* as a seed. This did not generate an assembly of any known mitochondrial sequence. Instead, a seemingly circular 2,139 bp contig was inferred in *Adiantum tricholepis*, containing *morffo1* and *morffo2*, but no known mitochondrial sequences. Furthermore, this contig lacked the flanking inverted repeats normally associated with *morffo1*. A control assembly using the mitochondrial genes *atp1* and *cox1* as seed sequences was also generated using non-plastid raw reads from this species. The assemblies based on mitochondrial genes had markedly lower average coverage depth (86.5) than that of the *morffo1* based assembly (556) and the plastome (359), suggesting that the MORFFO cluster in *Adiantum tricholepis* exists as an independent mobile element that is not an integrated component of the mitochondrial genome. Where this element resides within the cell is unclear.

Relationships among MORFFO sequences were estimated using a maximum likelihood optimality criterion and 1000 bootstrap replicates. Each MORFFO element—*morffo1, morffo2, morffo3*—comprises a monophyletic clade, within which some sequences are united with moderate to high bootstrap support (Supplementary Figure 1). However, relationships among sequences within each morffo clade were not congruent with the accepted species tree.

Plastome phylogenomic analyses

Our final, concatenated plastome alignment included 68,047 sites spanning 76 plastid loci for 31 taxa, including three outgroups. PartitionFinder2 returned a most favorable partition model with 36 subsets (AICc: 1211937.74576), from which partition blocks were assigned in RAxML (Stamatakis 2014) and MrBayes (Ronquist et al. 2012). Trees inferred using maximum likelihood (ML) and Bayesian optimality criteria were in full topological agreement with maximum support on all branches, with the exception of the branch subtending *Onychium japonicum*, *Tryonia myriophylla*, *Jamesonia brasiliensis*, *Gastionella chaerophylla*, and *Pityrogramma trifoliata*, which was supported by an ML bootstrap of 95 and a posterior probability of 1.0 (Fig 1).

Discussion

Comparative analyses of plastomes over the past two decades have dramatically improved our understanding of their evolution across land plants. Early data painted a picture of structural and organizational stability among deeply divergent embryophyte plastomes, punctuated by relatively few large-scale inversions (Hoot & Palmer 1994; Ogihara et al. 1988; Wolf et al. 2010). Recent evidence, however, has begun to expose the plastome as a dynamic molecule that in some lineages undergoes frequent changes in DNA content and structure (Knox 2014; Guisinger et al. 2011; Cremen et al. 2018; Lin et al. 2012). As more information has come to light, many highly rearranged plastomes have also been found to host sizeable insertions, occasionally including open reading frames (ORF) of unknown homology (Knox 2014; Cremen et al. 2018). In some cases, these

inserted ORFs appear to encode functional proteins, whereas in others they resemble conserved domains that have undergone extensive rearrangements and/or pseudogenization, comparable to what has been observed within some plastid genes (e.g., *ndhK*, *clpP*, and *ycf2*; (Haberle et al. 2008; Smith 2014; Sun et al. 2016; Lin et al. 2012). Several studies have determined that similar, undescribed plastid ORFs are the result of horizontal transfer from mitochondria to plastids (Rabah et al. 2017; Goremykin et al. 2009; Iorizzo et al. 2012; Ma et al. 2015; Burke et al. 2016).

Plastome ORF insertions like these—with no known sequence homology—have not been characterized in ferns, although previous authors have reported large intergenic expansions and insertions in some taxa (Logacheva et al. 2017; Gao et al. 2011). It was speculated that some of these intergenic expansions originated via intracellular transfer from the mitochondrion (Logacheva et al. 2017), but until now, limited sampling in previous studies has obscured the highly mobile nature of these peculiar sequences. Here, we take a focused phylogenetic approach, targeting the fern family Pteridaceae, to reveal a suite of highly mobile ORFs (MORFFO) within a broad sampling of plastomes from across the family. Preliminary analyses indicate that MORFFO elements, which are frequently associated with extensive genomic rearrangements, may be present in lineages well-removed from ferns.

Characterization of MORFFO elements

Logacheva et al. (2017) established that 'hypervariable' sequences of significant length, are found in the inverted repeat (IR) of *Woodwardia unigemmata* as well as the LSC of *Plagiogyria*. Our results are consistent with their findings, and further expose the dynamic nature of these sequences (MORFFO) among a collection of closely-related fern plastomes. Searches for MORFFO-like sequences outside of Pteridaceae returned similar, putatively homologous regions in many ferns (Table 2,3), but not in seed plants. Significantly, an 8 Kbp region in the plastome of the fern *Mankyua chejuensis* (Ophioglossaceae) contains an expanded complement of the MORFFO cluster. Additional searches for MORFFO-like sequences outside of vascular plants revealed similar conserved domains in several cyanobacteria plastomes (Table 2), as well as domains in the plastomes of the green algae *Prasiola crispa*, *Roya obtusa,* and *Roya anglica* (Table 2).

Within ferns, we note that MORFFO elements are frequently found adjacent to inferred sites of genomic inversion. For example: (1) *morffo1* is found adjacent to the border of one of two hypothesized inversions in the region spanning *rpoB-psbZ* which occurred in a common ancestor of the core leptosporangiates (Fig 3; (PPG I 2016)); (2) *morffo3* is found within a 9.7 Kbp inversion that characterizes leptosporangiate ferns (Kim et al. 2014); (3) *morffo1* and *morffo2* are found inserted adjacent to the 7 Kbp inversion seen in the plastomes of vittarioid ferns; (4) and *morffo1 and morffo2* also appear adjacent to an inversion described in filmy ferns (Fig 3; (Wolf et al. 2011; Kim et al. 2014; Gao et al. 2011). Although MORFFO insertions are frequently associated with

inversions, we are unsure why. One possibility is that MORFFO may target nucleotide sites that are prone to inversion. Conversely, the insertion of MORFFO could be directly influencing inversion events. In several taxa, we observe a proliferation of the inverted repeats flanking *morffo1*, possibly caused by replication slippage, or possibly by the repeated insertion and excision of *morffo1*. In other groups, plastome reorganization has been similarly associated with the presence of small dispersed repeats like these (Wicke et al. 2011). Likewise, the relationship between MORFFO insertion sites and inversions is not unlike the insertions seen in other dynamic embryophyte plastomes (Knox 2014).

The variable presence, location, and configuration of MORFFO observed in a phylogenetic context suggests that these ORFs are mobile elements. With a few notable exceptions, plastid genes are not frequently gained or lost, yet our results indicate that MORFFO moves into, out of, and across the plastome in relatively short evolutionary timescales. While MORFFO sequences have been observed in mitochondrial contigs (Logacheva et al. 2017), it is important to note that they are not found in either of the currently available complete mitochondrial genomes of ferns (Guo et al. 2017). Additionally, the location of these insertions in mitochondrial genomes seems to vary as much as in those of plastids, making it difficult to pinpoint a potential mitochondrial origin. Furthermore, we have also noted the presence of MORFFO in the nuclear genome of *Ceratopteris*. Thus, MORFFO appears to be moving across genomes as readily as within them. However, this does not explain the origin of the elements nor the mechanisms of their movement.

Plasmid-like sequences have been observed in the plastomes of diatoms, green algae, dinoflagellates, and red algae (Zhang et al. 1999; Cremen et al. 2018; Ruck et al. 2014; La Claire & Wang 2000; Lee et al. 2016). Although we are not aware of previous work describing chloroplast plasmids in land plants, this is a plausible mechanism to explain the variable presence and location of MORFFO elements. It would also explain the variability in order and direction of MORFFO insertions (Fig 1). As noted above (see results section), we were able to assemble a circular sequence containing *morffo1* and *morffo2* from *Adiantum tricholepis*, which did not have the MORFFO insert in its plastome. The coverage analysis for this sequence indicates that MORFFO is likely a high copy-number sequence that is independent of both the plastome and the mitochondrial genome. This, combined with the fact that *morffo2* has regions that share similarity to conserved domains (DN_5 superfamily) associated with primase genes found in mobile elements of cyanobacteria, strongly suggests that these sequences could be of plasmid origin, possibly from a plastid plasmid.

Alternatively, MORFFO elements could be of viral origin. In addition to being similar to plasmid primases, the conserved domain found in *morffo2* also resembles primase genes found in phages. Viral origins could explain why MORFFO is found frequently but irregularly in fern plastomes. Likewise, many of the above arguments in favor of a plasmid or plasmid-like origin for MORFFO sequences can also be attributed to viral origin.

The structural similarities that *morffo1* shares with bacterial insertion sequences is noteworthy, especially because insertion sequences are known to cause inversions (Darmon & Leach 2014). This, along with the clear mobility of these sequences strongly suggest that *morffo1* could be a previously undescribed insertion sequence. The case for *morffo1* being an insertion sequence is made stronger by the fact that of all the MORFFO sequences, it appears to display the most independence. In non-plastid DNA, it is almost always found without the other MORFFO sequences. It is also more frequently seen independent of additional MORFFO sequences in the plastomes of ferns outside of Pteridaceae. Furthermore, copies of *morffo1* were detected in the nuclear genome of *Ceratopteris* and the mitochondrial genome of *Asplenium nidus*, suggesting that it may be a particularly promiscuous mobile element. The relationship of *morffo2* and *morffo3* to *morffo1*, however, remains unclear. If *morffo1* is an independent insertion sequence, then how are *morffo2* and *morffo3* inserted?

Phylogenetic analysis of MORFFO elements reveals three strongly-supported, monophyletic groups: *morffo1, morffo2,* and *morffo3* elements. Relationships within each MORFFO clade do not reflect the accepted species phylogeny, but phylogenetic similarity across clades may reflect shared histories of degradation among MORFFO elements (Supplementary Figure 1.)

We also note that whereas MORFFO elements are pervasive in Pteridaceae, they appear to be less common in most other groups of ferns. In part, this may be an artifact of the historical reliance on reference-based assemblies, which can be biased towards

assembling genomes that appear more similar to their reference, thus reducing the likelihood of detecting significant rearrangements. Based on the few sequences available in GenBank, it would seem that MORFFO elements may be prevalent in *Plagiogyria* and Ophioglossaceae; however, it remains to be determined how widespread this cluster of ORFs is among other lineages of ferns. More studies at the family level are needed to understand the extent to which MORFFO sequences are moving throughout fern genome space. The current paucity of fern nuclear and mitochondrial genomes makes it difficult to determine the source of these inserts. At the time of writing there are only two fern mitochondrial sequences (Guo et al. 2017) available in GenBank, and no nuclear genomes, although several are in preparation. As more genomes are published in the coming years, the reservoir from which MORFFO clusters are migrating should become clear.

Plastome variation across Pteridaceae

Pteridaceae is an ecologically and morphologically diverse family comprising more than 10% of extant fern species (Schuettpelz et al. 2007). Within this group, subfamily Vittarioideae, comprising the genus *Adiantum* and the so-called vittarioid ferns, is especially noteworthy. High levels of molecular substitution rate heterogeneity has been detected between members of the genus *Adiantum* and the vittarioid ferns, in both plastid and nuclear DNA sequences (Rothfels & Schuettpelz 2014; Grusz et al. 2016). As noted above, we also find variation in plastome structure across the family,

with MORFFO elements (*morffo1*, *morffo2*, and *morffo3*) being repeatedly gained, lost, and/or rearranged, even among closely related taxa (Fig 1).

The physical position of MORFFO cluster insertions is relatively conserved within the 5 major clades comprising Pteridaceae (Fig 1), but in some cases the location and composition of these clusters varies widely, even between congeneric relatives (e.g., *Myriopteris lindheimeri* vs. *M. scabra* and *Vittaria appalachiana* vs. *V. graminifolia*; Fig 1). Based on our sampling, we find no evidence of MORFFO elements within Cryptogrammoideae or Parkerioideae. However, unique insertions of MORFFO sequences have taken place in some members of the Pteridoideae, including *Jamesonia brasiliensis* (between *trnN* and *ycf2*), *Tryonia myriophylla* (between *trnD* and *trnY*), and *Pteris vittata* (between *psbM* and *petN*). MORFFO elements were notably absent in the species of *Gastoniella*, *Pityrogramma*, and *Onychium* sampled (Fig 1). Compared to the Pteridoideae, subfamilies Vittarioideae and Cheilanthoideae exhibit relative stability in their MORFFO insertion sites (Fig 1). Altogether, we find *at least* nine unique MORFFO cluster insertions across the Pteridaceae (there are almost certainly more), not including a multitude of species-specific rearrangements, gains, and losses of individual MORFFO elements (*morffo1*, *morffo2*, and *morffo3*) following cluster insertions.

Within each independent MORFFO cluster insertion, the presence and position of *morffo1*, *morffo2*, and *morffo3* are highly variable among species sampled. For example, a variety of insertions, rearrangements, and losses of all three elements can be found in the MORFFO cluster found between *rpoB* and *trnD* within subfamily Vittarioideae (Fig

1). Likewise, the MORFFO cluster between *rps12* and *rrn16* in Cheilanthoideae shows insertions and losses of all three MORFFO elements, including a duplication of *morffo2* in *Hemionitis subcordata* (Fig 1).

Interestingly, vittarioid ferns do not appear to have experienced expansion or contraction of the IR, which have been associated with extensive genomic rearrangements, gene loss, and the proliferation of repetitive regions in other groups (Zhu et al. 2016). In addition to the variable presence of MORFFO elements, we find that their insertion into the ancestral vittarioid IR (Fig 1) may have also coincided with a loss of *trnT*. Most vittarioid ferns share an additional gene loss (*trnV*)—with the exception of *Haplopteris elongata*, in which *trnV* is found intact. Given that *Vaginularia trichoidea* and *H. elongata* are inferred to be successively sister to the remaining vittarioid ferns sampled (Fig 1; (Schuettpelz et al. 2016), this topology implies the gain of *trnV* into the plastome of *H. elongata*. Nevertheless, a shared ~7 Kbp inversion between *rrn16* and *rrn5* in all vittarioid ferns except *V. trichoidea* further supports our phylogenetic inferences based on DNA sequence data (Figs 1 and 3).

The vittarioid ferns are characterized by high levels of plastome rearrangement, elevated molecular substitution rates, a shift to epiphytism, morphological reduction, and shared ancestral whole genome duplication (Pryer et al. 2016). This array of shared traits leads one to ask which (if any) might have driven these changes in vittariod plastome structure and expression. Similarly, frequent rearrangements, insertions, and losses of the MORFFO suite within the plastomes of subfamily Cheilanthoideae coincide with

adaptations to extreme xeric environments, extensive whole genome duplications, hybridization, and apomixis—any of which may relate to the changes we detect in plastome structure across this subfamily.

Acknowledgements

Thanks to Jim Van Etten, Dan Sloan, Aaron Duffy, and Michael McKain for various fruitful discussions, Nico Devos (Duke University genome sequencing Center) for technical assistance, and Sally Chambers for providing the *Vittaria appalachiana* sample. Thanks to Blaine Marchant for generously allowing us to use the *Ceratopteris* nuclear genome in our search for MORFFO. Thanks also to the University of Utah Center for High-Performance Computing, particularly Anita Orendt, for providing computational resources for data analyses, and to the Minnesota Supercomputing Institute. We further acknowledge the Smithsonian Laboratories of Analytical Biology. Funding for this research was provided in part by the United States National Science Foundation (award DEB-1405181 to ES).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Ayliffe MA, Timmis JN. 1992. Plastid DNA sequence homologies in the tobacco nuclear genome. Mol. Gen. Genet. 236:105–112.

Blazier JC et al. 2016. Variable presence of the inverted repeat and plastome stability in Erodium. Ann. Bot. 117:1209–1220.

Bungard RA. 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. Bioessays. 26:235–247.

Burke SV et al. 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 16:140.

Cai Z et al. 2008. Extensive reorganization of the plastid genome of Trifolium subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J. Mol. Evol. 67:696–704.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the Monocot–Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. J. Mol. Evol. 58:424–441.

Chumley TW et al. 2006. The complete chloroplast genome sequence of Pelargonium x hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. 23:2175–2190.

Cremen MCM, Leliaert F, Marcelino VR, Verbruggen H. 2018. Large Diversity of Nonstandard Genes and Dynamic Evolution of Chloroplast Genomes in Siphonous Green Algae (Bryopsidales, Chlorophyta). Genome Biol. Evol. 10:1048–1061.

Cui L et al. 2006. Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. BMC Evol. Biol. 6:13.

Darmon E, Leach DRF. 2014. Bacterial genome instability. Microbiol. Mol. Biol. Rev. 78:1–39.

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 26:2455–2457.

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

Gao L, Zhou Y, Wang Z-W, Su Y-J, Wang T. 2011. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. BMC Plant Biol. 11:64.

Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am. J. Bot. doi: 10.1002/ajb2.1048.

Givnish TJ et al. 2010. Assembling the Tree of the Monocotyledons: Plastome Sequence Phylogeny and Evolution of Poales. Ann. Mo. Bot. Gard. 97:584–616.

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. Mol. Biol. Evol. 26:99–110.

Grusz AL, Rothfels CJ, Schuettpelz E. 2016. Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. BMC Genomics. 17:692.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28:583–600.

Guo W et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within Juniperus plants and have shifted multiple times during cupressophyte evolution. Genome Biol. Evol. 6:580–590.

Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns Ophioglossum californicum and Psilotum nudum are highly repetitive with the largest organellar introns. New Phytol. 213:391–403.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. J. Mol. Evol. 66:350–361.

Hein A, Polsakiewicz M, Knoop V. 2016. Frequent chloroplast RNA editing in early-branching flowering plants: pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. BMC Evol. Biol. 16:23.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the Cryptomeria japonica D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.

Hoot SB, Palmer JD. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of Anemone and related genera. J. Mol. Evol. 38:274–281.

Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature. 422:72–76.

Iorizzo M et al. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12:61.

Johnson MTJ et al. 2012. Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. PLoS One. 7:e50226.

Jung J, Kim JI, Jeong Y-S, Yi G. 2018. AGORA : Organellar genome annotation from the amino acid and nucleotide references. Bioinformatics. doi: 10.1093/bioinformatics/bty196.

Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. doi: 10.1093/bib/bbx108.

Kearse M et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma. Oxf. Engl. 28, 1647--1649.

Kim HT, Chung MG, Kim K-J. 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. Mol. Cells. 37:372–382.

Knox EB. 2014. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. Proc. Natl. Acad. Sci. U. S. A. 111:11097–11102.

Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22:1208–1222.

Krause K. 2008. From chloroplasts to ‘cryptic’ plastids: evolution of plastid genomes in parasitic plants. Curr. Genet. 54:111–121.

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 Res. 31:2417–2423.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33:1870–1874.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41:W22–8.

Labiak PH, Karol KG. 2017. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. Am. J. Bot. doi: 10.3732/ajb.1700135.

La Claire JW, Wang J. 2000. Localization of plasmidlike DNA in giant-celled marine green algae. Protoplasma. 213:157–164.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol. Biol. Evol. 34:772–773.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.

Lee J et al. 2016. Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. Sci. Rep. 6:23744.

Leliaert F et al. 2012. Phylogeny and molecular evolution of the green algae. CRC Crit. Rev. Plant Sci. 31:1–46.

Lenz H, Knoop V. 2013. PREPACT 2.0: Predicting C-to-U and U-to-C RNA Editing in Organelle Genome Sequences with Multiple References and Curated RNA Editing Annotation. Bioinform. Biol. Insights. 7:1–19.

Li F-W et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat Plants. 4:460–472.

Li F-W, Kuo L-Y, Pryer KM, Rothfels CJ. 2016. Genes Translocated into the Plastid

Inverted Repeat Show Decelerated Substitution Rates and Elevated GC Content. Genome Biol. Evol. 8:2452–2458.

Lin C-P, Wu C-S, Huang Y-Y, Chaw S-M. 2012. The complete chloroplast genome of Ginkgo biloba reveals the mechanism of inverted repeat contraction. Genome Biol. Evol. 4:374–381.

Liu C et al. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics. 13:715.

Logacheva MD et al. 2017. Comparative analysis of inverted repeats of polypod fern (Polypodiales) plastomes reveals two hypervariable regions. BMC Plant Biol. 17:255.

Madden T. 2013. *The BLAST Sequence Analysis Tool*. National Center for Biotechnology Information (US).

Ma P-F, Zhang Y-X, Guo Z-H, Li D-Z. 2015. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. Sci. Rep. 5:11608.

Martin W et al. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U. S. A. 99:12246–12251.

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol. 118:9–17.

Matasci N et al. 2014. Data access for the 1,000 Plants (1KP) project. Gigascience. 3:17.

Matsuo M, Ito Y, Yamauchi R, Obokata J. 2005. The Rice Nuclear Genome Continuously Integrates, Shuffles, and Eliminates the Chloroplast Genome to Cause Chloroplast–Nuclear DNA Flux. Plant Cell. 17:665–675.

McFadden GI. 2001. Chloroplast origin and integration. Plant Physiol. 125:50–53.

McKain MR, Hartsock RH, Wohl MM, Kellogg EA. 2017. Verdant: automated annotation, alignment and phylogenetic analysis of whole chloroplast genomes. Bioinformatics. 33:130–132.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). pp. 1–8.

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl. Acad. Sci. U. S. A. 107:4623–4628.

Mower JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res. 37:W253–9.

Mower JP, Vickrey TL. 2018. Chapter Nine - Structural Diversity Among Plastid Genomes of Land Plants. In: Advances in Botanical Research. Chaw, S-M & Jansen, RK, editors. Vol. 85 Academic Press pp. 263–292.

Nakayama T, Archibald JM. 2012. Evolving a photosynthetic organelle. BMC Biol. 10:35.

NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45:D12–D17.

Ogihara Y, Terachi T, Sasakuma T. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. U. S. A. 85:8573–8577.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19:325–354.

PPG I. 2016. A community-derived classification for extant lycophytes and ferns: PPG I. Jnl of Sytematics Evolution. 54:563–603.

Pryer KM et al. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. 91:1582–1598.

Pryer KM, Huiet L, Li F-W, Rothfels CJ, Schuettpelz E. 2016. Maidenhair Ferns, Adiantum, are Indeed Monophyletic and Sister to Shoestring Ferns, Vittarioids (Pteridaceae). Syst. Bot. 41:17–23.

Rabah SO et al. 2017. Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew. Plant Genome. 10. doi: 10.3835/plantgenome2017.03.0020.

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2015. Tracer v1. 6. 2014.

Ronquist F et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

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

Ruck EC, Nakov T, Jansen RK, Theriot EC, Alverson AJ. 2014. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. Genome Biol. Evol. 6:644–654.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms--inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14:23.

Schuettpelz E et al. 2016. A revised generic classification of vittarioid ferns (Pteridaceae) based on molecular, micromorphological, and geographic data. Taxon. 65:708–722.

Schuettpelz E, Schneider H, Huiet L, Windham MD, Pryer KM. 2007. A molecular phylogeny of the fern family Pteridaceae: assessing overall relationships and the affinities of previously unsampled genera. Mol. Phylogenet. Evol. 44:1172–1185.

Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA. 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and Arabidopsis. Plant Mol. Biol. 52:923–934.

Shaw J et al. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92:142–166.

Sigmon BA, Adams RP, Mower JP. 2017. Complete chloroplast genome sequencing of vetiver grass (Chrysopogon zizanioides) identifies markers that distinguish the non-fertile 'Sunshine'cultivar from other …. Ind. Crops Prod. https://www.sciencedirect.com/science/article/pii/S0926669017304922.

Smit A, Hubley R, Green P. 2013. 2013–2015. RepeatMasker Open-4.0.

Smith DR. 2014. Mitochondrion-to-plastid DNA transfer: it happens. New Phytol. 202:736–738.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U. S. A. 100:8828–8833.

Stein DB et al. 1992. Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proc. Natl. Acad. Sci. U. S. A. 89:1856–1860.

Sun Y et al. 2016. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Mol. Phylogenet. Evol. 96:93–101.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. 2013. RNA editing in plants and its evolution. Annu. Rev. Genet. 47:335–352.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123–135.

Tonti-Filippini J, Nevill PG, Dixon K, Small I. 2017. What can we do with 1000 plastid genomes? Plant J. 90:808–818.

Tryon RM. 1990. Pteridaceae. In: Pteridophytes and Gymnosperms. Kramer, KU & Green, PS, editors. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 230–256.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27:171–180.

de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. 2015. YCF1: A Green TIC? Plant Cell. 27:1827–1833.

Walker BJ et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963.

Wei R et al. 2017. Plastid Phylogenomics Resolve Deep Relationships among Eupolypod II Ferns with Rapid Radiation and Rate Heterogeneity. Genome Biol. Evol. 9:1646–1657.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76:273–297.

Wickett NJ et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U. S. A. 111:E4859–68.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U. S. A. 84:9054–9058.

Wolf PG et al. 2015. An Exploration into Fern Genome Space. Genome Biol. Evol. 7:2533–2544.

Wolf PG et al. 2011. The evolution of chloroplast genes and genomes in ferns. Plant Mol. Biol. 76:251–261.

Wolf PG, Roper JM, Duffy AM. 2010. The evolution of chloroplast genome structure in ferns. Genome. 53:731–738.

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.

Wolf PG, Rowe CA, Sinclair RB, Hasebe M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, Adiantum capillus-veneris L. DNA Res. 10:59–65.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252–3255.

Xie Y et al. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 30:1660–1666.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21:809–818.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. Nature. 400:155–159.

Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209:1747–1756.

Kuo, L.-Y., X Qi, H. Ma, F.-W. Li. 2018. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. American Journal of Botany. *in press*.

CHAPTER 2

REFERNMENT: AN R PACKAGE FOR ANNOTATING RNA EDITING IN PLASTID GENOMES

ReFernment: An R package for annotating RNA editing in plastid genomes[1]

Tanner A. Robison [2,3] and Paul G. Wolf [2]

[2]Department of Biology, Utah State University, Logan, UT, USA

Email addresses: TR: robison.tanner@gmail.com

PW: paul.wolf@usu.edu

Number of words: yyyy

[1] Manuscript received _____; revision accepted _____.

[3] Author for correspondence: robison.tanner@gmail.com

ABSTRACT

- *Premise of the study:* In the absence of cDNA, the annotation of RNA editing in plastomes must be done manually, representing a significant time cost to those studying the organellar genomes of ferns and hornworts.
- *Methods and Results:* We developed an R package to automatically annotate apparent nonsense mutations in plastid genomes. The software successfully annotates such sites and results in no false positives for data with no sequencing or assembly errors.
- *Conclusions:* Compared to manual annotation, ReFernment offers greater speed and accuracy for annotating RNA editing sites. This software should be especially useful for researchers generating large numbers of plastome sequences for taxa with high levels of RNA editing.

Key words: Plastome; chloroplast; genome; annotation; RNA editing; Genbank; NCBI

INTRODUCTION

The development of Next Generation Sequencing has led to an explosion of available genome data, especially for plastid genomes (plastomes). These relatively small genomes are a major source of data for phylogenetic analyses. Currently (September, 2018), more than 2700 plastome sequences from green plants have been published (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid) in public databases, which has in turn aided in the resolution of deep phylogenetic relationships across plant diversity (Ruhfel et al., 2014; Tonti-Filippini et al., 2017; Gitzendanner et al., 2018). However, researchers assembling and annotating plastomes are often faced with the problem of RNA editing, whereby the sequence of the initial transcript is altered prior to translation. In some groups of plants RNA editing can be high: up to 78% of protein coding genes in plastomes of ferns (Wolf et al., 2004) and hornworts (Kugita et al., 2003). Many of these RNA editing sites will alter the sequences of start codons, stop codons, or result in stop codons within the genomic coding sequence. The most common forms for RNA editing in plastomes are U to C or C to U editing (Kugita et al., 2003; Wolf et al., 2004). Whereas many of the automated annotation tools presently available are generally good at annotating plastid genes, none of them account for RNA editing (Wyman et al., 2004; Liu et al., 2012; McKain et al., 2017; Jung et al., 2018). This results in annotated genes that appear to be missing start codons, stop codons, or containing numerous internal stops based on their nucleotide

translations. Reasonably, issues like these make it difficult to get some plastome sequences approved for public databases such as GenBank.

Although RNA editing appears to occur at a lower rate in angiosperms than in other clades, 138 RNA editing sites were detected in the plastome of Amborella (Hein et al., 2016). Thus, the need to annotate RNA editing sites may not be restricted to a few seed-free lineages. Tools are available to predict RNA editing sites, for example PREPACT (Lenz and Knoop, 2013) and PREP Suit (Mower, 2009). Among these, only PREPACT can adjust gene annotations, but only if cDNA sequences are provided. In many cases this is not technically or financially feasible, in which case researchers must manually add these annotations by examining each nonsense mutation, and determine whether RNA editing would likely restore this site. This process, while necessary for admission to public repositories, is tedious and time-consuming—especially considering these edits to nonsense mutations occur in a highly predictable manner. Here, we attempt to solve this problem by introducing ReFernment, a simple R package that automatically annotates nonsense codons in DNA translations to account for RNA editing and provides conceptual translations for coding sequences https://github.com/TARobison/ReFernment .

METHODS AND RESULTS

ReFernment operates by refining existing annotations. Thus, the software uses an annotation generated by programs such as DOGMA, CpGAVAS, Verdant, or AGORA (Wyman et al., 2004; Liu et al., 2012; McKain et al., 2017; Jung et al., 2018), and adjusts these annotations to account for RNA editing. ReFernment requires both a GFF3 (no sequence) file and a GenBank flat file (including nucleotide sequence), and its basic operation is extremely simple. First, ReFernment checks the starting and final codons of each gene. In both cases, ReFernment initially checks whether the codon is a valid start or stop. If the codon is not valid, it checks whether an RNA editing event would result in the restoration of the codon to a valid start or stop (e.g. ACG -> AUG). If the codon is not valid, even after checking for possible RNA editing, ReFernment checks whether nearby codons (within 5 codons) represent valid codons; if so, ReFernment changes the gene boundaries to start or stop at those valid sites. Next, ReFernment checks whether a gene has any internal stops, and if so, checks whether RNA editing would restore these nonsense mutations, adjusting the translation to account for this. ReFernment then edits the imputed GenBank flat file, adding conceptual translations and annotations indicating the sites where RNA editing occurred with ‘misc_feature’ flags, adding necessary RNA editing flags to the relevant genes, and providing a conceptual translation for each gene. Finally, ReFernment produces a five column feature table, formatted correctly for submission to GenBank, and a protein fasta file with the conceptual translations for coding sequences where RNA editing has occurred.

ReFernment operates under the assumption that only U-to-C or C-to-U RNA editing is occurring in the plastome (Takenaka et al., 2013) Additionally, ReFernment assumes that all nonsense mutations are the result of RNA editing. Since most of the genes that reside within the plastome are vital to photosynthetic function, it is assumed that these genes will remain operational. There may be cases where internal stops, bad starts, or missing stops are actually the result of an uncorrected mutation, especially in parasitic lineages (Krause, 2008). When ReFernment was tested against plastomes with high levels of RNA editing, confirmed with cDNA data (AB086179 and AY178864.1), every nonsense mutation was correctly annotated, and there were no false positive annotations. A major limitation of ReFernment is that the annotations it produces are only as good as the annotations it is provided. If a gene annotation is frameshifted, if a pseudogene is annotated as a coding sequence, if there are assembly errors, or if an annotation has the incorrect start and stop sites, ReFernment might interpret this as RNA editing, rather than an error. In other words, ReFernment is not a substitute for manually checking gene annotations, nor is ReFernment a fix for sloppy annotation. In an attempt to mitigate these problems, if there are more than 5 detected internal stops in a gene, ReFernment will produce an error suggesting that the user manually check that gene. There are cases where genes have more than 5 RNA edited internal stops, but these are relatively rare, so users should use best judgement.

The utility of ReFernment is simple: it saves users time in the final stages of annotation. Manually accounting for RNA edits generally takes hours for a typical fern or hornwort plastid genome, but with ReFernment, this process takes less than a minute. There are currently efforts to publish some 1,000 additional fern plastomes in the coming years, and hopefully similar efforts are underway for hornworts, meaning many thousands of hours can be saved by the implementation of this simple program. ReFernment not only saves the researcher time, but also provides consistent methodology for the annotation of RNA editing. In many cases, RNA editing sites are not annotated in plastid sequences and only conceptual translations are provided. This not only results in confusion in how to annotate such sites consistently, but also make it difficult for researchers interested in the evolution of such sites to readily identify them.

CONCLUSIONS

ReFernment offers easy and rapid annotation of RNA edited sites and automatic conceptual translation of amino acid sequences, streamlining the process of GenBank submission and saving the user valuable time.

ACKNOWLEDGEMENTS

The authors thank Aaron Duffy, Carol Rowe, and two anonymous reviewers for improving this manuscript. The authors also thank William Pearse for providing advice on the software.

LITERATURE CITED

Gitzendanner, M.A., P.S. Soltis, G.K.-S. Wong, B.R. Ruhfel, and D.E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American journal of botany*. Available at: http://dx.doi.org/10.1002/ajb2.1048.

Hein, A., M. Polsakiewicz, and V. Knoop. 2016. Frequent chloroplast RNA editing in early-branching flowering plants: pilot studies on angiosperm-wide coexistence of editing sites and their nuclear specificity factors. *BMC evolutionary biology* 16: 23.

Jung, J., J.I. Kim, Y.-S. Jeong, and G. Yi. 2018. AGORA : Organellar genome annotation from the amino acid and nucleotide references. *Bioinformatics* . Available at: http://dx.doi.org/10.1093/bioinformatics/bty196.

Krause, K. 2008. From chloroplasts to "cryptic" plastids: evolution of plastid genomes in parasitic plants. *Current genetics* 54: 111–121.

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

Lenz, H., and V. Knoop. 2013. PREPACT 2.0: Predicting C-to-U and U-to-C RNA Editing in Organelle Genome Sequences with Multiple References and Curated RNA Editing Annotation. *Bioinformatics and biology insights* 7: 1–19.

Liu, C., L. Shi, Y. Zhu, H. Chen, J. Zhang, X. Lin, and X. Guan. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC genomics* 13: 715.

McKain, M.R., R.H. Hartsock, M.M. Wohl, and E.A. Kellogg. 2017. Verdant: automated annotation, alignment and phylogenetic analysis of whole chloroplast genomes. *Bioinformatics* 33: 130–132.

Mower, J.P. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic acids research* 37: W253–9.

Ruhfel, B.R., M.A. Gitzendanner, P.S. Soltis, D.E. Soltis, and J.G. Burleigh. 2014. From algae to angiosperms--inferring the phylogeny of green plants (Viridiplantae)

from 360 plastid genomes. *BMC evolutionary biology* 14: 23.

Takenaka, M., A. Zehrmann, D. Verbitskiy, B. Härtel, and A. Brennicke. 2013. RNA editing in plants and its evolution. *Annual review of genetics* 47: 335–352.

Tonti-Filippini, J., P.G. Nevill, K. Dixon, and I. Small. 2017. What can we do with 1000 plastid genomes? *The Plant journal: for cell and molecular biology* 90: 808–818.

Wolf, P.G., C.A. Rowe, and M. Hasebe. 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.

Wyman, S.K., R.K. Jansen, and J.L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

SUMMARY AND CONCLUSIONS

The results of the study in chapter two exposes the dynamic nature of MORFFO sequences in the fern family Pteridaceae. It is clear that MORFFO not only mediates genomic rearrangements in the plastid genomes of ferns, but also drives endosymbiotic gene transfer to other genomic compartments -- and may even be of nuclear genome origins. Furthermore, these results do significant work to challenge the notion that the plastome is a 'fossilized' genome that does not undergo significant changes in structure or content. As plastomes continue to be useful tools for phylogenetic and evolutionary insights, it is critical that we understand the processes which drive changes in their structure and gene content. In the case of ferns, it is important to recognize that structural differences between the plastomes of ferns can sometimes be the result of these MORFFO elements, and therefore that variations in gene order may not always be the best tool for determining phylogenetic relationships. Further work is needed to determine the extent to which MORFFO elements affect the plastomes of ferns. Additionally, experimental work to determine the nature and origins of MORFFO may provide valuable insights into our understanding on how the various genomes within a plant interact with one another.

In chapter 3 I describe a software application I wrote, which will aid researchers

in the annotation and identification of RNA editing sites within the plastomes and mitogenomes of ferns and hornworts. The use of ReFernment will not only have the practical benefit of saving significant amounts of time during the annotation process, but should allow them to make deeper insights into the evolution of RNA editing sites and to explore their function. This tool has already been used in the study in chapter 2 and proved to save significant amounts of time. In future iterations of the program we hope to implement more robust, molecularly informed statistical models to predict sites which are not associated with nonsense mutations. Such an implementation will provide more accurate amino acid translations, thus reducing the number of false positive differences in plastid protein alignments and ultimately improving the quality of phylogenetic inferences based on genomic sequences.

Table 1: Summary of basic genomic features of plastomes used in this study. IR = Inverted Repeat, LSC = Large Single Copy, SSC = Small Single Copy

Species	Inverted Repeat Size	LSC Size	Genome Size	SSC Size	%GC	*MORFFO*
Adiantum aleuticum	26,289	83,345	157,519	21,596	45	1,2
Adiantum capillus-veneris	23,448	82,282	150,568	21,390	41	Absent
Adiantum tricholepis	23,233	82,740	150,667	21,461	42	Absent
Antrophyum semicostatum	20,977	87,492	150,274	20,828	40	1,3
Bommeria hispida	23,142	82,491	156,749	27,974	43	1,2,3
Calciphilopteris ludens	26,585	82,423	157,068	21,475	43	1,2
Ceratopteris cornuta	22,287	83,623	149,424	21,227	37	Absent
Ceratopteris richardii	22,020	83,178	148,444	21,226	35	Absent
Cheilanthes bolborrhiza	~25,000	na	39,380*	na	44	1,2
Cheilanthes micropteris	23,306	88,393	157,567	22,562	41	1,3
Cryptogramma acrostichoides	22,652	83,690	150,162	21,168	42	Absent
Cystopteris chinensis	26,671	83429*	131808*	21708*	40	Absent
Dryopteris decipiens	23,456	82,462	150,978	21,604	42	Absent
Gastoniella chaerophylla	22,657	81,918	148,099	20,867	40	2
Haplopteris elongata	27,188	80,810	156,002	20,816	41	1,2
Hemionitis subcordata	30,921	82,607	165,631	21,182	43	1,2,3
Jamesonia brasiliensis	27,704	na	139,531*	20,941	41	1
Llavea cordifolia	23,208	81,944	149,387	21,027	42	Absent
Myriopteris covillei	25,567	83,093	155,548	21,321	42	2
Myriopteris lindheimeri	25,694	83,059	155,770	21,323	42	2
Myriopteris scabra	27,115	82,874	162,051	24,947	42	1
Notholaena standleyi	27,261	83,769	159,556	21,265	42	1,2,3
Onychium japonicum	23,419	82,289	150,156	21,029	41	Absent
Pellaea truncata	23,240	82,865	150,713	21,368	42	Absent
Pentagramma triangularis	23,378	85,675	153,445	21,014	42	1,3
Pityrogramma trifoliata	22,465	82,321	148,156	20,905	40	Absent
Pteridium aquilinum	23,384	84,335	152,362	21,259	41	2
Pteris vittata	25,275	82,604	154,108	20,954	42	2
Scoliosorus ensiformis	21,078	82,358	145,327	20,813	40	Absent
Tryonia myriophylla	24,141	87,238	156,327	20,807	40	1,2,3
Vaginularia trichoidea	21,618	84,026	147,192	19,930	39	1,2,3

Vittaria appalachiana	22,185	84,330	149,531	20,831	40	1
Vittaria graminifolia	22,066	86,058	151,035	20,845	40	1,2

Table 2: Summary of hits to MORFFO sequences in NCBI blast, using either BLASTX or TBLASTN. * = putative DN_5 superfamily conserved domain

Species	Organism group and genome compartment	search strategy	MORFFO match	Length of match (aa)	% identity	Accession number
Actinostachys pennula	Fern plastome	TBLASTN	*morffo2*	231	34	KU764518.1
Alsophila spinulosa	Fern plastome	TBLASTN	*morffo2*	78	44	FJ556581.1
Angiopteris angustifolia	Fern plastome	TBLASTN	*morffo2*	332	49	KP099647
Angiopteris evecta	Fern plastome	TBLASTN	*morffo2*	331	47	DQ821119.1
Asplenium nidusmitochondria	Fern mitochondrial genome	TBLASTN	*morffo1*	260	52	AM600641.1
Asplenium prolongatum	Fern plastome	TBLASTN	*morffo2*	110/59	77/64	KY427332.1
*Chondrocystis(plasmid)**	Cyanobacterium	TBLASTN	*morffo2*	233	27	AP018284.1
*Crocosphaera watsonii**	Cyanobacterium	BLASTX	*morffo2*	365	43	WP_007310072.1
Dryopteris fragrans	Fern plastome	TBLASTN	*morffo1/morffo2*	323/355	46/49	KX418656.2
Helminthostachys zeylanica	Fern plastome	TBLASTN	*morffo2*	226	70	KM817788.2
Huperzia lucidula	Lycopod plastome	TBLASTN	*morffo2*	114	35	AY660566.1
Huperzia serrata	Lycopod plastome	TBLASTN	*morffo2*	114	35	KX426071.1
Lepisorus clathratus	Fern plastome	TBLASTN	*morffo1/morffo2*	183/208	50/56	KY419704.1
Lygodium japonicum	Fern plastome	TBLASTN	*morffo2*	152	36	HM021803.1
*Myxosarcina spp.**	Green alga plastome	BLASTX	*morffo2*	497	27	WP_052055951.1
*Nostoc punctiforme**	Cyanobacterium	TBLASTN	*morffo2*	245	28	CP001037.1
Ophioglossum californicum	Fern plastome	TBLASTN	*morffo1/morffo2/morffo3*	120/169/65	56/41/42	KC117178.1
Polypodium glycyrrhiza	Fern plastome	TBLASTN	*morffo3*	*162*	53	KP136832
Prasiola crispa	Green alga plastome	TBLASTN	*morffo2*	416	25	KR017750.1
Roya anglica	Green alga plastome	TBLASTN	*morffo1*	202	30	NC_024168
Roya obtusa	Green alga plastome	TBLASTN	*morffo1*	202	30	KU646496.1
Volvox carteri	Green alga plastome	TBLASTN	*morffo1/morffo2*	200/303	28/25	EU755299.1
Woodwardia unigemmata	Fern plastome	TBLASTN	*morffo1/morffo2*	247/105	51/77	KT599101.1

Table 3: Summary of matches for MORFFO sequences within ferns, using BLASTN. IR = Inverted Repeat, LSC = Large Single Copy, SSC = Small Single Copy

Species	MORFFO present	Length of match (bp)	% identity	Region	Accession number
Adiantum aleuticum	*morffo1/ morffo2*	1316/ 1851	88/67	IR	MH173079
Alsophila podophylla	*morffo1/morffo2/morffo3*	1250/1778/631	78/76/81	IR	MG262389
Antrophyum semicostatum	*morffo1/ morffo3*	1323/ 629	53/ 57	LSC	MH173087
Asplenium prolongatum	*morffo2*	265	81	IR	KY427332
Bommeria hispida	*morffo1/ morffo2/ morffo3*	1296/ 1708/ 1708	89/ 87/ 84	SSC	MH173074
Calciphilopteris ludens	*morffo1/ morffo2 (truncated)*	1324/ 928	78/ 51	IR	MH173084
Cheilanthes bolborrhiza	*morffo1/ morffo2*	1304/ 1859	92/ 90	IR	MH173073
Cheilanthes micropteris	*morffo1/ morffo3*	1292/ 630	50/ 50	LSC	MH173078
Cibotium barometz	*morffo1/ morffo2*	1296/1850	77/50	IR	NC_037893
Dicksonia squarrosa	*morffo1*	1321	50	IR	KJ569698
Diplopterygium glaucum	*morffo3*	639	55	LSC	KF225594
Drynaria roosii	*morffo1*	1314	51	LSC	KY075853
Haplopteris elongata	*morffo1/ morffo2*	1336/ 1859	56/ 55	IR	MH173086
Hemionitis subcordata	*morffo1/ morffo2/ morffo3*	1296/ 1863/ 665	91/ 75/ 92	IR	MH173072
Hymenasplenium unilaterale	*morffo1/morffo3*	252/ 643	65/61	IR/LSC	KY427350
Jamesonia brasiliensis	*morffo1*	1310	51	IR	MH173077
Mankyua chejuensis	*morffo1/ morffo2*	193/186	68/67	IR	KP205433
Myriopteris covillei	*morffo2*	1886	49	IR	MG953517
Myriopteris lindheimeri	*morffo2*	852	67	IR	HM778032
Myriopteris scabra	*morffo1*	1310	98	IR	MH173083
Notholaena standleyi	*morffo1/ morffo2/ morffo3*	1312/ 1866/ 640	89/ 86/ 91	IR	MH173067
Pentagramma triangularis	*morffo2/ morffo3*	1860/ 630	83/ 84	LSC	MH173070
Plagiogyria glauca	*morffo1*	1305	49	LSC	KP136831
Plagiogyria glauca	*morffo2*	1856	46	Mito	Wolf et al. (2014)
Plagiogyria japonica	*morffo2*	1295	50	LSC	HQ658099
Pteridium aquilinum	*morffo2*	280	65	LSC	HM535629.1
Pteris vittata	*morffo2*	1172	80	LSC	MH173068
Rhachidosorus consimilis	*morffo2*	430	70	IR	KY427356
Tryonia myriophylla	*morffo1/ morffo2/ morffo3*	1326/ 1854/ 641	89/ 84/ 86	LSC	MH173076
Vaginularia trichoidea	*morffo1/ morffo2/ morffo3*	1315/ 1836/ 630	53/ 52/ 62	LSC	MH173085

Vittaria appalachiana	*morffo1*	1316	68	LSC	MH173091

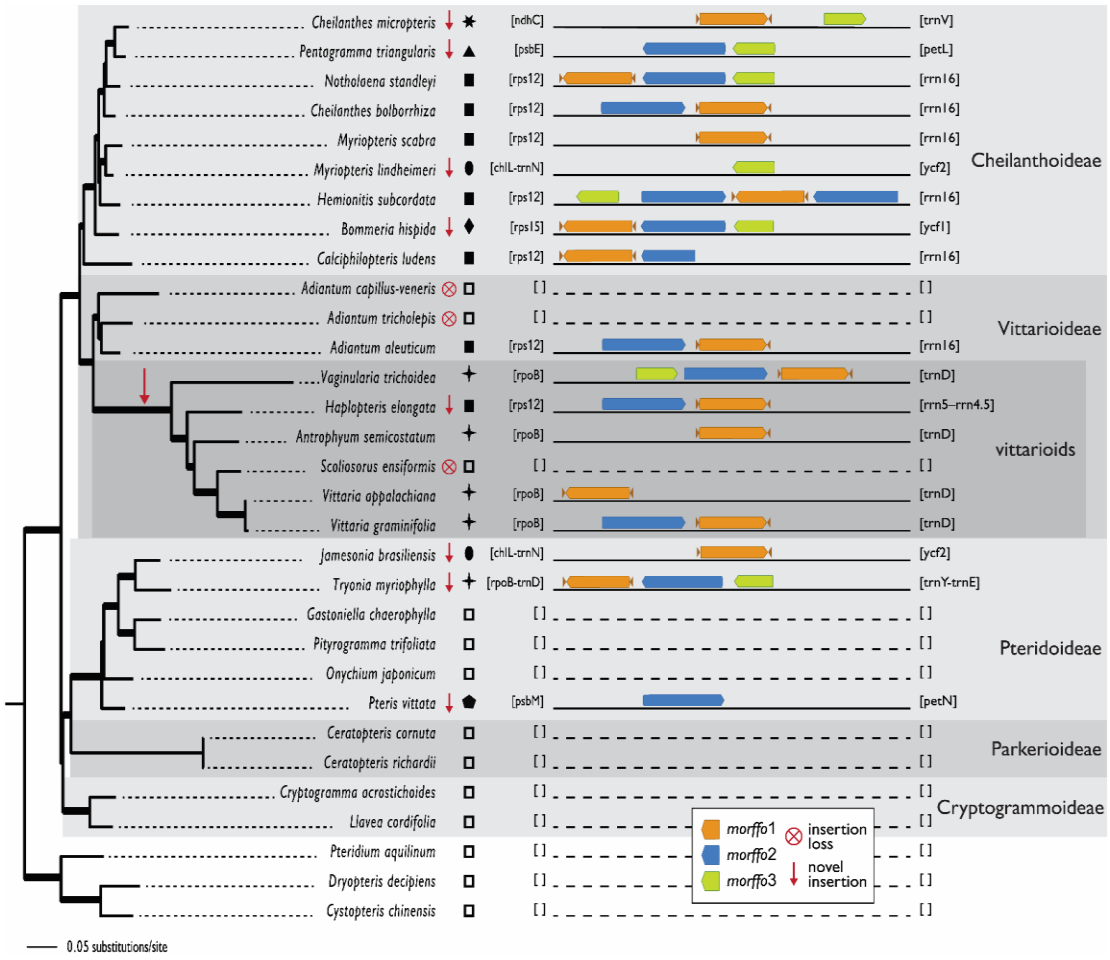
Table 4. Taxonomic sampling and voucher information for samples used in this study.

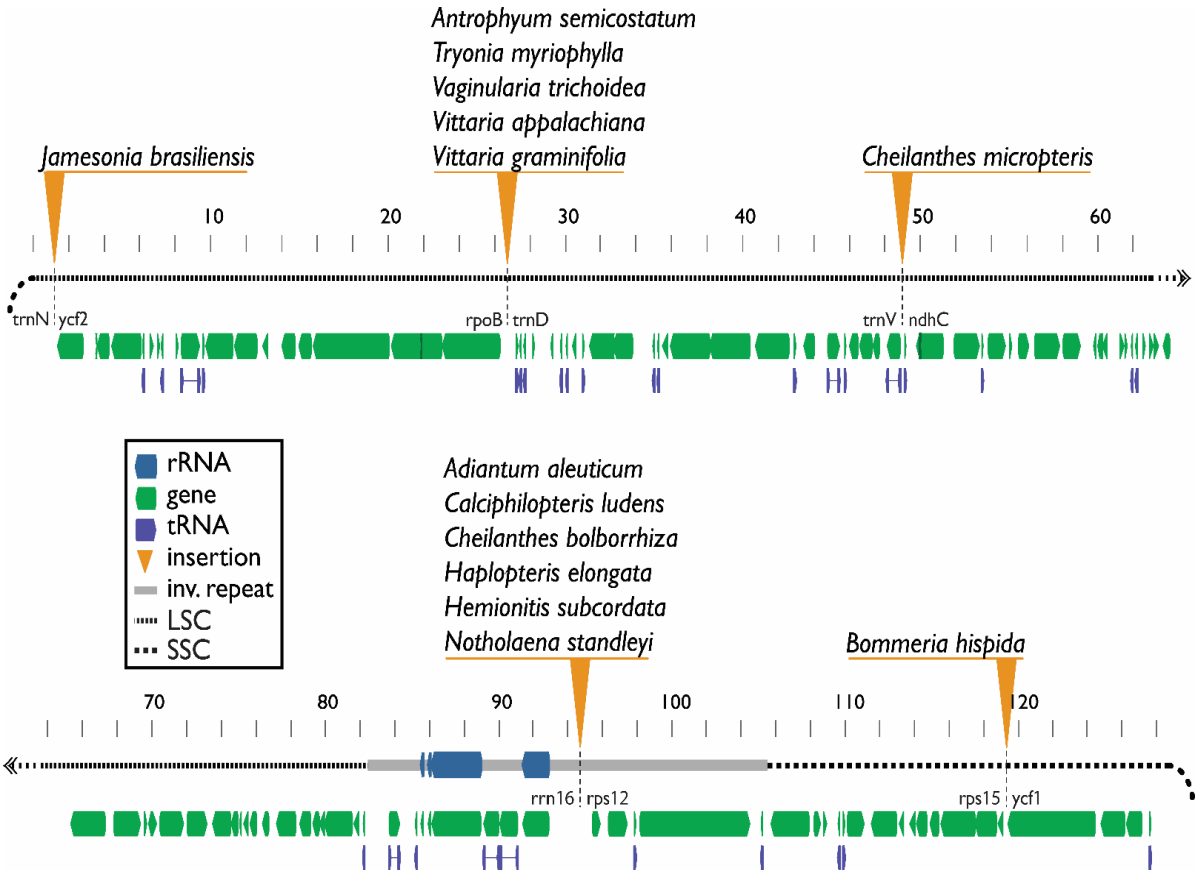
Taxon	Voucher or citation	Genbank Accession
Adiantum aleuticum(Rupr.) C. A. Paris	Rothfels 4097 (DUKE)	MH173079
Adiantum capillus-venerisL.	Wolf et al. 2004	NC004766
Adiantum tricholepisFée	Rothfels 08-094 (DUKE)	MH173071
Antrophyum semicostatumBlume	Schuettpelz 1561 (BO)	MH173087
Bommeria hispida(Mett. ex Kuhn) Underw.	Beck 1130 (DUKE)	MH173074
Calciphilopteris ludens(Wall. ex Hook.) Yesilyurt & H. Schneid.	Huiet s.n. (DUKE)	MH173084
Ceratopteris cornuta(P. Beauv.) Lepr.	Rothfels 4298 (DUKE)	MH173082
Ceratopteris richardiiBrongn.	Marchant et al. unpub.	KM052729
Cheilanthes bolborrhizaMickel & Beitel	Rothfels 3294 (DUKE)	MH173073
Cheilanthes micropterisSw.	Prado 2132 (DUKE)	MH173078
Cryptogramma acrostichoidesR. Br.	Rothfels 4195 (DUKE)	MH173081
Cystopteris chinensis(Ching) X. C. Zhang & R. Wei	Wei et al. 2017	KY427337
Dryopteris decipiens(Hook.) Kunze	Wei et al. 2017	KY427348
Gastoniella chaerophylla(Desv.) Li Bing Zhang & Liang Zhang	Prado 2178 (SP)	MH173080
Haplopteris elongata(Sw.) E. H. Crane	Schuettpelz 1559 (BO)	MH173086
Hemionitis subcordata(D. C. Eaton ex Davenp.) Mickel	Rothfels 3163 (DUKE)	MH173072
Jamesonia brasiliensisChrist	Schuettpelz 1444 (SP)	MH173077
Llavea cordifoliaLag.	Schuettpelz 1744 (US)	MH173088
Myriopteris covillei(Maxon) Á. Löve & D. Löve	Schuettpelz 443 (DUKE)	MG953517
Myriopteris lindheimeri(Hook.) J. Sm.	Schuettpelz 450	NC014592
Myriopteris scabra(C. Chr.) Grusz & Windham	Windham 3495 (DUKE)	MH173083
Notholaena standleyiMaxon	Schuettpelz 435 (DUKE)	MH173067
Onychium japonicum(Thunb.) Kunze	Schuettpelz 1057 (DUKE)	MH173069
Pellaea truncataGoodd.	Schuettpelz 430 (DUKE)	MH173066
Pentagramma triangularis(Kaulf.) Yatsk., Windham & E. Wollenw.	Schuettpelz 1332 (DUKE)	MH173070
Pityrogramma trifoliata(L.) R. M. Tryon	Rothfels 3658 (DUKE)	MH173075
Pteridium aquilinum(L.) Koon.	Der et al. unpub.	NC014348
Pteris vittataL.	Schuettpelz 893 (DUKE)	MH173068
Scoliosorus ensiformis(Hook.) T. Moore	Schuettpelz 1782 (US)	MH173090
Tryonia myriophylla(Sw.), Schuettp., J. Prado & A. T. Cochran	Schuettpelz 1434 (SP)	MH173076
Vaginularia trichoideaFée	Schuettpelz 1553 (BO)	MH173085
Vittaria appalachianaFarrar & Mickel	Stevens OH-p1-s11 (PUR)	MH173091

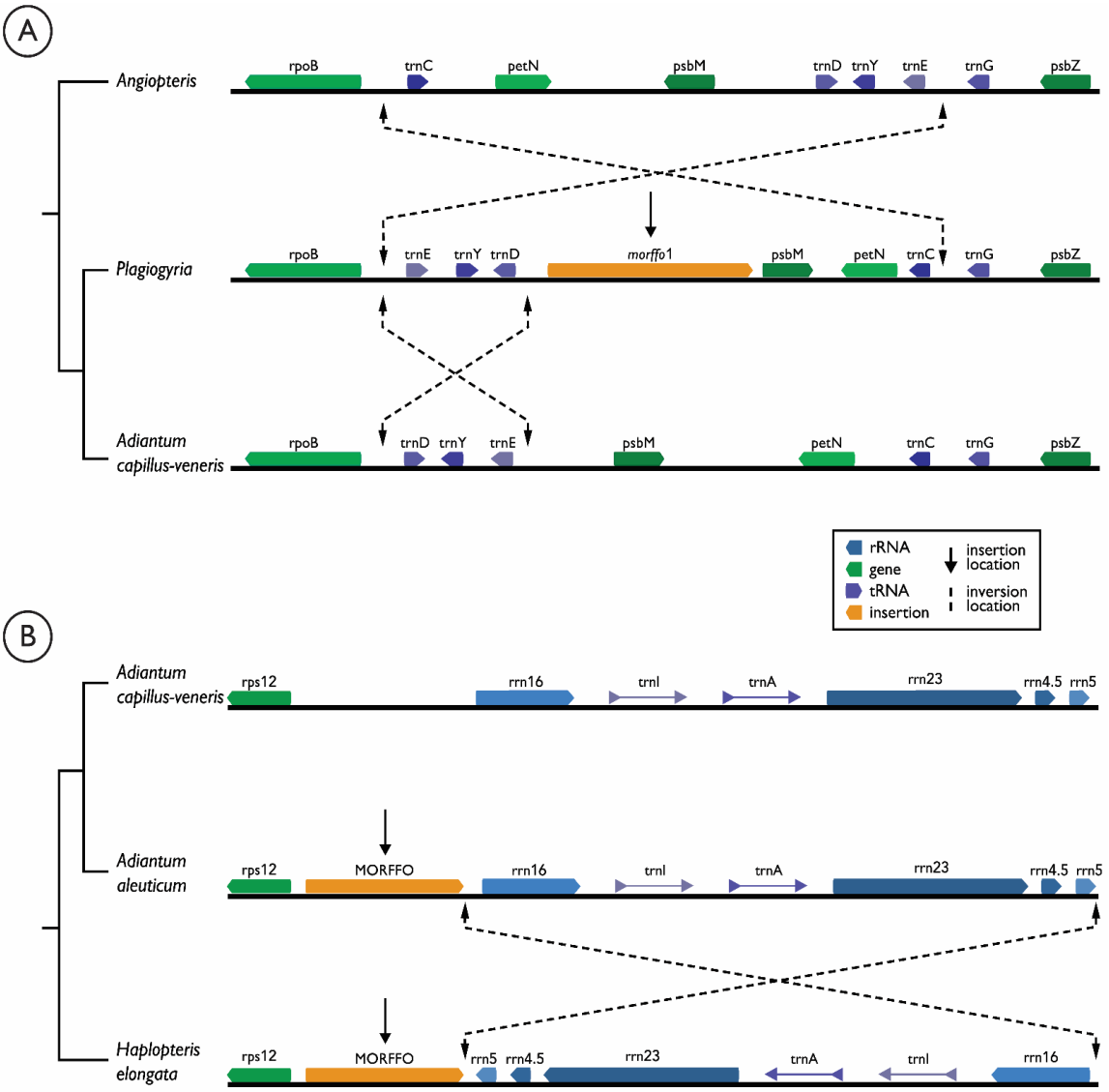
Figure 1. Distribution of MORFFO elements across the Pteridaceae phylogeny. Topology results from maximum likelihood analysis of plastome data (-lnL = 609991.403586); thickened branches indicate bootstrap/posterior probability support = 100/1.0. Symbols highlight shared insertion sites, with empty squares signifying evident lack of a MORFFO insertion. Short arrows flanking *morffo1* indicate short inverted repeats. Novel insertions and losses, as inferred by maximum parsimony, are depicted as arrows or crossed-out circles, respectively.

Figure 2. Detected insertion sites in plastomes of Pteridaceae, relative to *Adiantum capillus-veneris*. Light gray bar denotes inverted repeat region.

Figure 3. Major inversion events uncovered in fern plastomes. A) Depiction of the two inversion events necessary to explain gene order differences between *Angiopteris* and *Adiantum* and their relationship to *morffo1*. B) Depiction of the inversion events seen in early leptosporangiate and vittarioid ferns, highlighting relationship of *morffo1* to the event.


Cheilanthes micropteris
Pentagramma triangularis
Notholaena standleyi
Cheilanthes bolborrhiza
Myriopteris scabra
Myriopteris lindheimeri
Hemionitis subcordata
Bommeria hispida
Calciphilopteris ludens
Adiantum capillus-veneris
Adiantum tricholepis
Adiantum aleuticum
Vaginularia trichoidea
Haplopteris elongata
Antrophyum semicostatum
Scoliosorus ensiformis
Vittaria appalachiana
Vittaria graminifolia
Jamesonia brasiliensis
Tryonia myriophylla
Gastoniella chaerophylla
Pityrogramma trifoliata
Onychium japonicum
Pteris vittata
Ceratopteris cornuta
Ceratopteris richardii
Cryptogramma acrostichoides
Llavea cordifolia
Pteridium aquilinum
Dryopteris decipiens
Cystopteris chinensis
[ndhC]
[trnV]
[psbE]
[petL]
[rps12]
[rrn16]
[chlL-trnN]
[ycf2]
[rps15]
[ycf1]
[rpoB]
[trnD]
[rrn5–rrn4.5]
[rpoB-trnD]
[trnY-trnE]
[psbM]
[petN]
[ ]
Cheilanthoideae
Vittarioideae
vittarioids
Pteridoideae
Parkerioideae
Cryptogrammoideae
morffo1
morffo2
morffo3
insertion loss
novel insertion
0.05 substitutions/site


Antrophyum semicostatum
Tryonia myriophylla
Vaginularia trichoidea
Vittaria appalachiana
Vittaria graminifolia
Jamesonia brasiliensis
Cheilanthes micropteris
10
20
30
40
50
60
trnN
ycf2
rpoB
trnD
trnV
ndhC
rRNA
gene
tRNA
insertion
inv. repeat
LSC
SSC
Adiantum aleuticum
Calciphilopteris ludens
Cheilanthes bolborrhiza
Haplopteris elongata
Hemionitis subcordata
Notholaena standleyi
Bommeria hispida
70
80
90
100
110
120
rrn16
rps12
rps15
ycf1


A
Angiopteris
rpoB
trnC
petN
psbM
trnD
trnY
trnE
trnG
psbZ
Plagiogyria
rpoB
trnE
trnY
trnD
morffo1
psbM
petN
trnC
trnG
psbZ
Adiantum capillus-veneris
rpoB
trnD
trnY
trnE
psbM
petN
trnC
trnG
psbZ
rRNA
gene
tRNA
insertion
insertion location
inversion location
B
Adiantum capillus-veneris
rps12
rrn16
trnI
trnA
rrn23
rrn4.5
rrn5
Adiantum aleuticum
rps12
MORFFO
rrn16
trnI
trnA
rrn23
rrn4.5
rrn5
Haplopteris elongata
rps12
MORFFO
rrn5
rrn4.5
rrn23
trnA
trnI
rrn16

APPENDIX A

PERMISSION TO USE: GENOME BIOLOGY AND EVOLUTION

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Sep 08, 2019

This Agreement between Utah State Univsersity -- Tanner Robison ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4660801326105
License date	Sep 02, 2019
Licensed content publisher	Oxford University Press
Licensed content publication	Genome Biology and Evolution
Licensed content title	Mobile Elements Shape Plastome Evolution in Ferns
Licensed content author	Robison, Tanner A; Grusz, Amanda L
Licensed content date	Aug 25, 2018
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Exploring the evolution and physiology of hornworts
Publisher of your work	Utah State University
Expected publication date	Sep 2019
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Exploring the evolution and physiology of hornworts
Institution name	Utah State University
Expected presentation date	Sep 2019
Requestor Location	Utah State Univsersity 5305 Old Main Hill LOGAN, UT 84321 United States Attn: Utah State Univsersity
Publisher Tax ID	GB125506730
Total	0.00 USD
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it

apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

APPENDIX B

PERMISSION TO USE: COAUTHORS

Hi Jeff,

Hope all is going well for you. I am in the process of getting my thesis through all the various hoops and such at Utah State, and one of the things I realized is that I need permission from every co-author on the MORFFO paper before I can use it as part of my thesis. So, here I am now, asking you if I can do that (pretty please!).

Thanks so much,

Tanner

Hi Tanner,

Of course. Please feel free to include it in your thesis.

Best,

Jeff

Hi Amanda,

Hope all is going well for you. I am in the process of getting my thesis through all the various hoops and such at Utah State, and one of the things I realized is that I need permission from every co-author on the MORFFO paper before I can use it as part of my thesis. So, here I am now, asking you if I can do that (pretty please!).

Thanks so much,
Tanner

Hi Tanner,

Absolutely. You have my permission to use the MORFFO paper as part of your thesis. Let me know if you need a formal letter.

Cheers,

Amanda

Hi Karla,

Hope all is going well for you. I am in the process of getting my thesis through all the various hoops and such at Utah State, and one of the things I realized is that I need permission from every co-author on the MORFFO paper before I can use it as part of my thesis. So, here I am now, asking you if I can do that (pretty please!).

Thanks so much,

Tanner

Hi Tanner,

I hope you're doing well! And of course! Not a problem at all.

I hope the other hoops are easily jumpable, if still present. Take care,
Karla

Hi Eric,

Hope all is going well for you. I am in the process of getting my thesis through all the various hoops and such at Utah State, and one of the things I realized is that I need permission from every co-author on the MORFFO paper before I can use it as part of my thesis. So, here I am now, asking you if I can do that (pretty please!).

Thanks so much,
Tanner

Of course!

Eric Schuettpelz

Hey Blake,

Hope all is going well for you. I am in the process of getting my thesis through all the various hoops and such at Utah State, and one of the things I realized is that I need permission from every co-author on the MORFFO paper before I can use it as part of my thesis. So, here I am now, asking you if I can do that (pretty please!).

Thanks so much,
Tanner

Absolutely. I wouldn't even be a co-author if not for you.

Best wishes,
Blake