Resolving Deep Relationships and Revealing Ancient Whole-Genome Duplications in Pteridaceae using Transcriptomic Data

Authors: Song, Michael J., Rothfels, Carl J., Schuettpelz, Eric, Nitta, Joel, Huiet, Layne, et al.

Source: American Fern Journal, 113(3) : 191-210

Published By: The American Fern Society

URL: https://doi.org/10.1640/0002-8444-113.3.191

American Fern Journal 113(3):191–210 (2023)
Published on 29 September 2023

Resolving Deep Relationships and Revealing Ancient Whole-Genome Duplications in Pteridaceae using Transcriptomic Data

Michael J. Song*
Department of Integrative Biology and the University Herbarium, University of California, Berkeley, CA 94720, USA

Carl J. Rothfels
Current address: Ecology Center and Department of Biology, Utah State University, Logan, UT 84332, USA

Eric Schuettpelz
Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

Joel Nitta
College of Liberal Arts and Sciences, Graduate School of Global and Transdisciplinary Studies, Chiba University, Chiba, Japan

Layne Huiet
Duke University Herbarium, Department of Biology, Durham, NC 27708, USA

Fay-Wei Li
Boyce Thompson Institute and Plant Biology Section, Cornell University, Ithaca, NY 14853, USA

Keir M. Wefferling*
Department of Natural and Applied Sciences and Gary A. Fewless Herbarium, University of Wisconsin, Green Bay, WI 54311, USA

Abstract.—Relationships among the major subclades in the fern family Pteridaceae have proven difficult to resolve. Here, we examine the backbone of this large and heterogeneous lineage using both phylotranscriptomic methods and a more focused, curated approach. We find that Pteridoideae and Parkerioideae are together sister to the rest of Pteridaceae and that Cryptogrammoideae is sister to Vittarioideae plus Cheilanthoideae. We find independent support from our phylotranscriptomic analyses, published cytological data, and genomic distributions of substitutions per site for several whole-genome duplication (WGD) events within Pteridaceae, mainly in Vittarioideae and Cheilanthoideae. However, the various inference methods gave differing approximations for the placement of WGD events within each clade. This study demonstrates that phylotranscriptomic analyses, which employ large datasets at the cost of requiring simpler models and potentially a greater risk of systematic error, can be used in concert with more curated approaches to resolve deep phylogenetic relationships. It also provides an example of the difficulty of confidently inferring ancient WGD event placement, even when using multiple methods.

Key Words.—curated phylogenomics, paleopolyploidy, polyploidy, phylotranscriptomics, phylogenetics, ancient whole-genome duplications

*Corresponding authors: songm@smccd.edu; wefferlk@uwgb.edu

With more than 1,000 species and accounting for nearly 10% of known fern diversity (Smith *et al.*, 2006; PPG I, 2016), Pteridaceae is among the largest and most heterogeneous families of leptosporangiate ferns (Schuettpelz *et al.*, 2007, Schuettpelz and Pryer, 2009; Testo and Sundue, 2016; Nitta *et al.*, 2022). The family has been consistently supported as monophyletic in molecular phylogenetic studies (Pryer, Smith, and Skog, 1995; Hasebe *et al.*, 1995; Gastony and Johnson, 2001; Schneider *et al.*, 2004; Schuettpelz *et al.*, 2007; Schuettpelz and Pryer, 2007; Lehtonen, 2011; Rothfels *et al.*, 2015) and is characterized by linear marginal sori, false indusia, and a chromosome base number of $x = 29$ or 30 (Windham, 1993; Kramer, Green, and Götz, 2013). However, Pteridaceae exhibits extreme variability in both form and habit, having adapted to environments as different as cloud forests, deserts, freshwater ponds, and tropical mangrove communities.

High morphological and ecological disparity in Pteridaceae has resulted in numerous disagreements among historical classifications. For example, species now placed in Pteridaceae have been ascribed to over 20 families (Rothfels, 2008), with some lineages often recognized as their own distinct families (*e.g.*, Adiantaceae, Bommeriaceae, Ceratopteridaceae, Cheilanthaceae, Coniogrammaceae, and Vittariaceae). Five major groups have been consistently recovered as monophyletic in molecular phylogenetic analyses (*e.g.,* Schuettpelz *et al.,* 2007), and are treated as subfamilies in the Pteridophyte Phylogeny Group classification (PPG I, 2016): Parkerioideae, Cryptogrammoideae, Vittarioideae (comprising the reciprocally monophyletic *Adiantum* and vittarioid clades), Pteridoideae, and Cheilanthoideae. These subfamilies vary greatly in their relative species richness, with two genera and nine species in Parkeroideae, three genera and 31 species in Cryptogrammoideae, 13 genera and approximately 400 species in Pteridoideae, 12 genera and 345 species in Vittarioideae, and 23 genera and 426 species in Cheilanthoideae (species number estimates from PPG I, 2016). The relationships among these five clades have been contentious, with inconsistent inferences among studies (Schuettpelz and Pryer, 2007; Schuettpelz *et al.*, 2007; Kuo *et al.*, 2011; Rothfels *et al.*, 2015; Testo and Sundue, 2016; Qi *et al.*, 2018; Shen *et al.*, 2018; Wolf *et al.*, 2018; Nitta *et al.*, 2022; Pelosi *et al.*, 2022); no one hypothesis has emerged as the well-supported backbone across independent datasets.

Pteridaceae is also known for having high rates of polyploidy, closely associated with extensive hybridization and apomixis (Windham and Yatskievych, 2003; Rothfels, 2008; Grusz, Windham, and Pryer, 2009; Beck, Windham, and Pryer, 2011; Sigel *et al.*, 2011; Schuettpelz *et al.*, 2015; Kao *et al.*, 2019; Adjie *et al.*, 2007; Chao *et al.*, 2012a; Chao *et al.*, 2012b; Jaruwattanaphan, Matsumoto, and Watano, 2013). However, the poorly resolved phylogenetic backbone of the family has precluded an understanding of older whole-genome duplication (WGD) events in the history of the lineage. While some ancient WGDs have been inferred in the family from the One Thousand Plant Transcriptomes Initiative (OneKP, 2019), the *Ceratopteris* genome (Marchant *et al.*, 2022), and Pelosi *et al.* (2022), at least two studies have searched for and found no evidence of paleopolyploidy in Pteridaceae (Huang *et al.*, 2020; Fang *et al.*, 2022).

Over the last few years, advances in phylotranscriptomics have allowed for the integration of multiple approaches in the inference of both recent and relatively ancient WGD events (Yang and Smith, 2014; Yang *et al.*, 2015, 2018; McKain *et al.*, 2018; Li *et al.*, 2015; Li *et al.*, 2018). A "classic" approach largely relies on *Ks* plots, examining distributions of synonymous substitution distances between pairs of paralogs for departures from the expected exponential distribution. This method has known limitations especially when compared to the "gold standard"—synteny—but can be robust under some scenarios (Tiley, Barker, and Burleigh, 2018). More recently developed tree-based approaches allow for establishment of orthologous and paralogous relationships among gene copies (Yang and Smith, 2014; Gardner *et al.*, 2016; McKain *et al.*, 2018; Yang *et al.*, 2018; Morales-Briones *et al.*, 2021), and provide an alternative approach for the inference of WGD events (Tiley, Barker, and Burleigh, 2018; Li and Barker, 2020; Li *et al.*, 2015; Li *et al.*, 2018).

Phylotranscriptomic studies often allow researchers to resolve the placement of recalcitrant taxa or overcome other phylogenetic challenges using large datasets, as such datasets effectively remove stochastic variation in the substitution process as a source of estimation error (McKain *et al.*, 2018). At the same time, there is more potential for undetected systematic error with larger datasets, wherein data cannot be manually examined or curated—for example, to ensure that the alignments are accurate estimates of homology—and more biological realistic complex models are not computationally feasible (Phillips, Delsuc, and Penny, 2004; Philippe *et al.*, 2011; Rothfels *et al.* 2012). Here, we explore using a curated-data approach with a subset of our phylotranscriptomic dataset as a complementary method to phylotranscriptomic inference. We recognize that our approach (i.e., lacking a factorial design with each analysis/dataset) does not allow for a full comparison of phylotranscriptomic methods or dataset curation approaches; however, the study demonstrates a number of tools typically used in phylotranscriptomic studies and shows instances of concordance and discordance in the resulting inferred relationships and WGD events. Our study aims to leverage the availability of transcriptomes and recently accessible pipelines to 1) test previous phylogenetic hypotheses in Pteridaceae; 2) compare tree topologies and support between phylotranscriptomic and more curated data approaches; and 3) assess support for putative ancient WGDs in Pteridaceae.

Materials and Methods

Our approach to phylotranscriptomic and WGD inference uses several tools to estimate a phylogenetic backbone for Pteridaceae with 33 ingroup and 10 outgroup taxa, and to seek support for WGDs within that phylogeny. Two tree-based orthology inference approaches were used in order to compare resulting datasets and inferences of whole genome duplications: 1) the "1to1" ortholog determination approach of Yang and Smith (2014), implemented using the R package *baitfindR* (Nitta, 2020) on messenger RNA sequence data; and 2) the OrthoFinder pipeline (Emms and Kelly, 2019), designed to identify orthogroups, then orthologs, and recognize gene duplication events through reconciliation of

gene trees and species trees using translated peptides. Curated subsets of orthologs inferred by the Yang and Smith (2014) pipeline were, in turn, employed in various phylogenetic inference analyses: 1) maximum likelihood using RAxML (Stamatakis, 2014); 2) gene tree-species tree reconciliation using ASTRAL-III (Zhang *et al.*, 2018); and 3) Bayesian tree inference in RevBayes (Höhna et al., 2016). Finally, three approaches to inference of WGD or chromosome number evolution were used: 1) OrthoFinder (Emms and Kelly, 2019), as mentioned above; 2) ChromoEvol (Glick and Mayrose 2014) for chromosome number evolution implemented in RevBayes 1.2.0 (Höhna *et al.*, 2016); and 3) Ks plot examination in *wgd* (Zwaenepoel and Van de Peer, 2019).

Taxonomic sampling.—Taxa were selected to encompass the breadth of extant Pteridaceae, with an effort to sample diversity proportionally across the family (Appendix). In 2018, we gathered all transcriptome data that were publicly available or shared (Chien-Hsun Huang and Hong Ma provided assembled transcriptomes from Bioproject PRJNA422112 before the raw data were publicly available with the publication of Qi *et al.*, 2018) and then sequenced select taxa to fill in major gaps. Our ingroup sampling comprises 33 species, including representatives of each of the five Pteridaceae subfamilies: two species (two genera) from Parkeroideae; two species (two genera) from Cryptogrammoideae; six species (five genera) from Pteridoideae; five species (five genera) from Vittarioideae; and 11 species (10 genera) from Cheilanthoideae (Table 1). We also included a total of 10 outgroup species (Appendix).

Transcriptome sequencing and assembly.—For newly generated transcriptomes, RNA was extracted from approximately 20 mg of frozen leaf tissue with the Sigma-Aldrich Spectrum Plant Total RNA extraction kit (MilliporeSigma, Burlington, Massachusetts, USA) using their protocol A. Extracted RNA was stabilized and shipped in GenTegra-RNA matrix (NBS Scientific, Canonsburg, Pennsylvania, USA) to the Duke Center for Genomic and Computational Biology (Durham, North Carolina, USA) for sequencing. Libraries were prepared using the KAPA (Roche; MilliporeSigma, Burlington, Massachusetts, USA) stranded mRNA-seq kit (for samples included in Bioproject PRJNA716637; Appendix) or the TruSeq Ribo-zero (Illumina, San Diego, California, USA) library prep kit (for samples included in Bioproject PRJNA821853; Appendix); all libraries were sequenced on a HiSeq 4000 (Illumina, San Diego, California, USA), producing 150bp paired-end reads. Trimmomatic v. 0.36 (Bolger, Lohse, and Usadel, 2014) was used to trim adapters from the raw HiSeq data and Trinity v. 2.5.1 (Grabherr *et al.*, 2011) was used to assemble transcriptomes de novo. Forward and reverse reads were filtered to remove reads with average *phred* quality scores lower than 5.0 within a 4-bp sliding window at each end, and those reads that were shorter than 26 bp *(i.e.*, using default settings for Trimmomatic; Bolger, Lohse, and Usadel, 2014). Note that our interpretation and discussion of the transcriptome data below refers to "genes" or "orthologs" rather than the more strictly accurate "transcript clusters" (as described in the transcriptome assembly output for Trinity; Grabherr *et al.*, 2011).

TABLE 1. Taxon names and chromosome counts used in this study.

Taxon	Haploid Chromosome Count	Source
Acrostichum aureum L.	30	Marcon, Barros, and Guerra (2003)
Actiniopteris semiflabellata Pic. Serm.	?	
Adiantum caudatum L.	30	Srivastava (1985)
Adiantum cf-davidii Franch.	?	
Adiantum hispidulum S. W.	?	
Adiantum jordanii Müll. Hal.	30	Smith (1974)
Adiantum macrophyllum Sw.	30	Jermy and Walker (1985)
Aleuritopteris chrysophylla (Hook.) Ching	30	Löve, Löve, and Pichi-Sermoli (1977)
Antrophyum callifolium Blume	57	Kato (1999)
Antrophyum semicostatum Blume	60	Cave (1959)
Argyrochosma nivea (Poir.) Windham	54	Sigel *et al.* (2011)
Aspidotis carlotta-halliae (W. H. Wagner & E. F. Gilbert) Lellinger	60	Windham and Yatskievych (2003)
Bommeria hispida (Mett. Ex Kuhn) Underw.	30	Windham and Yatskievych (2003)
Ceratopteris richardii Brongn.	39	Löve, Löve, and Pichi-Sermoli (1977)
Cheilanthes chusana Hook.	?	
Cheilanthes nitidula Wall. ex Hook.	29	Knobloch *et al.* (1975)
Coniogramme fraxinea (D. Don) Diels	60	Cave (1964)
Cryptogramma acrostichoides R. Br.	60	Pajaron, Pangua, and Garcia Alvarez (1999)
Gaga angustifolia (Kunth) Fay W. Li & Windham	?	
Haplopteris amboinensis (Fée) X. C. Zhang	90	Ammal and Bhavanandan (1992)
Haplopteris elongata (Sw.) E. H. Crane	90	Ammal and Bhavanandan (1992)
Haplopteris heterophylla C. W. Chen, Y. H. Chang & Yea C. Liu	?	
Myriopteris rufa Fée	90	Windham and Yatskievych (2003)
Notholaena montieliae Yatsk. & Arbeláez	30	Kao *et al.* (2019)
Onychium japonicum (Thunb.) Kunze	58	Kato *et al.* (1992)
Parahemionitis cordata (Roxb. ex Hook. & Grev.) Fraser-Jenk.	60	efloras.org
Pentagramma triangularis (Kaulf.) Yatsk., Windham & Wollenw.	30	Windham and Yatskievych (2003)
Pityrogramma trifoliata (L.) R. M. Tryon	58	efloras.org
Pteris ensiformis Burm.	58	Kato (1999)
Pteris vittata L.	58	Srivastava (1985)
Taenitis blechnoides (Willd.) Sw.	55	Darnaedi (1992)
Vaginularia trichoidea Fée	30	Löve, Löve, and Pichi-Sermoli (1977)
Vittaria lineata (L.) J. E. Sm.	120	Kato *et al.* (1992)

Identification of orthologs.—We used the R package *baitfindR* (Nitta, 2020) to infer one-to-one orthologs among the ingroup taxa. The *baitfindR* pipeline is an implementation of the Yang and Smith (2014) workflow, a set of tools for inferring orthology among loci in order to find candidate genes or loci for phylogenetic analyses and/or inference of whole genome duplications. We used the "1to1" (*i.e.*, one-to-one) method of Yang and Smith (2014) since it is the most conservative,

TABLE 2. Datasets used in phylogeny inference.

Dataset	Method	Gene Alignments	Characters	Model
Large	Maximum Likelihood (RAxML)	371	908,103 bp	GTR + I + G
Medium	ASTRAL	104	275,927 bp	GTRCAT
Small	Bayesian inference (RevBayes)	10	34,841 bp	GTR + I + G

Outgroups: ***Alsophila podophylla*** Hook., ***Cystodium sorbifolium*** (Sm.) J. Sm., ***Cystopteris fragilis*** (L.) Bernh., ***Dennstaedtia hirsuta*** (Sw.) Mett. ex Miq., ***Deparia lobato-crenata*** (Tagawa) M. Kato, ***Deparia petersenii*** (Kunze) M. Kato, ***Gymnocarpium oyamense*** (Baker) Ching, ***Lindsaea linearis*** Sw., ***Polystichum acrostichoides*** (Michx.) Schott, ***Pteridium revolutum*** (Blume) Nakai, ***Orthiopteris campylura*** (Kunze) Copel., ***Struthiopteris spicant*** (L.) Weiss, and ***Tectaria nayarii*** Mazumdar.

only keeping strict orthologs. From this list of candidates, we removed any gene that was represented by fewer than four sequences, that did not BLAST to a combined *Arabidopsis/Azolla/Salvinia* genome set (*Arabidopsis* TAIR10 from ensemblgenomes.org, Lamesch *et al*, 2012; *Azolla* and *Salvinia* from fernbase.org, Li *et al.*, 2018), or which did not include at least one intron (determined by aligning the assembled transcriptomes to the combined genomes); these criteria were included in support of a parallel project of building a set of low- or single-copy nuclear gene baits for use in Pteridaceae phylogenomic and phylogeographic studies.

These criteria reduced the set of 9,682 candidate one-to-one orthologous genes to 3,306. From the set of 3,306 genes (ingroup only), *Gaga angustifolia* appeared in the most alignments (3,098) and was used in a BLAST search to query FASTA files containing sequences from all the other taxa (ingroup and outgroup); we retained the best hits, then built new alignments with all taxa. These new alignments were used in the subsequent locus selection and tree-building steps.

To allow a comparison with the WGD inference results from ChromEvol and *wgd* (see below), we additionally used OrthoFinder v2.3.3 (Emms and Kelly 2019) to infer a set of one-to-one orthologous genes from translated peptide sequences and place whole-genome duplication events on our phylogeny (described in "Inference of whole-genome duplication events" section below).

Datasets and alignments.—We used both phylotranscriptomic and curated-data approaches to yield three datasets: 1) a "large dataset" with the highest taxa representation for maximum likelihood (ML) analysis of a concatenated matrix in RAxML (Stamatakis, 2014); 2) a "medium dataset" of gene trees for species-tree inference in ASTRAL (Zhang *et al.*, 2018); and 3) a "curated dataset" for Bayesian inference (BI) in RevBayes (Höhna *et al.*, 2016) (Table 2). The large dataset included every gene alignment that contained at least 32 taxa; we selected this cut-off to accommodate computational limitations while at the same time maximizing taxon representation. The medium dataset consisted of 104 of these clusters that had the highest taxon representation, to accommodate computational limitations. For our curated dataset analyzed in RevBayes we chose the 10 best-sampled loci that start with a start codon; these were gene alignments that we were most confident in (in terms of homology and alignment) and we chose 10 of these

to accommodate our limited computational resources. All sequences were aligned with MAFFT Ver.7 (Katoh and Standley, 2013) using the linsi iterative refinement method employing WSP and consistency scores (–localpair –maxiterate 1000). The focus of this study is not to perform a systematic comparison between the three methods, but rather to see whether a large dataset phylotranscriptomic analysis would yield a similar result to a smaller more highly curated approach. We sought to perform these analyses like a regular user might be expected to. Thus, for the phylotranscriptomic analyses we used a large supermatrix, but then we also used a small dataset as a user would, should they have fewer data.

We compiled a dataset of chromosome counts from the Chromosome Counts Database (CCDB; Rice *et al.*, 2015) for 27 out of the 33 total ingroup taxa (summarized in Table 1); we used the CCDB as a starting point, and added additional/more reliable counts as available after doing a search through the literature. For species that had multiple reported cytotypes, we chose the lowest counts for our analysis not including aneuploidy.

Phylogenetic analyses.—The ML tree inference on the large dataset was conducted under the GTR + I + G model in RAxML using the rapid bootstrap with majority rules extended "MRE-based" bootstrapping criterion (Stamatakis, 2015). Given the size and complexity of our large dataset, we used the most complex model readily available on CIPRES (Miller, Pfeiffer, and Schwartz, 2010), as we expect that the true biological process is going to be considerably more complex than our most complex models (e.g., Fabreti and Höhna, 2022). For our medium dataset, gene trees were inferred locally using RAxML (Stamatakis, 2014) under the GTRCAT model (Stamatakis, 2015) and these trees were used as input for species-tree inference in ASTRAL-III (Zhang *et al.*, 2018) using default parameters, again on CIPRES. For our curated dataset, BI trees were inferred using RevBayes (Höhna *et al.*, 2016) under a GTR + I + G model, with the data partitioned by gene and codon position. RevBayes was used because we were interested in whether we would infer the same topology as a ML tree inferred from a supermatrix of phylotranscriptomic data, but instead only using a small number of nuclear genes. Four MCMC runs were performed for 1,000,000 generations; each run converged after 100,000 generations. As the runs converged on the same topology, we report the maximum a posteriori (MAP) topology for the first run. As our main goal is resolving the backbone of Pteridaceae, we do not expect that the different models of evolution assumed by the three methods used in this study would bias the results (Kelchner and Thomas, 2007).

Inference of whole-genome duplication events.—As part of the *baitfindR* (Nitta, 2020) pipeline to select orthologs, assembled transcriptomes were translated into peptides using Transdecoder (Haas, 2021) and redundant sequences were removed using cd-hit (Li and Godzik, 2006). These translated transcriptomes were used as input for OrthoFinder v2.3.3 (Emms and Kelly, 2019). Diamond v2.0.6 (Buchfink, Xie, and Huson, 2015) and MAFFT v7.310 (Katoh and Standley, 2013) were used for multiple sequence alignment, and

FastTree v2.1.10 (Price, Dehal, and Arkin, 2010) for initial tree inference. For the OrthoFinder analyses, three outgroup accessions were included: *Cystopteris fragilis*, *Deparia lobato-crenata*, and *Struthiopteris spicant*. After preliminary unguided runs (which resulted in apparently erroneous topological relationships; *i.e.*, similar to, but not concordant with our results from RAxML, ASTRAL, and RevBayes analyses), a guide tree was input based on the congruent topology of the ML and BI trees derived from the large dataset (Fig. 1). We chose to use a guide tree for OrthoFinder, as we employed what we view as more robust phylogenetic inference approaches (RAxML, ASTRAL, and RevBayes) to those used by the OrthoFinder pipeline (i.e., STRIDE and STAG; Emms and Kelly 2017, 2018). We used a threshold of at least 500 gene duplications at a node as evidence for a WGD to be considered.

We used ChromoEvol (Glick and Mayrose 2014) implemented in RevBayes 1.2.0 (Höhna *et al.*, 2016) to model chromosome changes across the ML phylogeny, which was made into an ultrametric tree using penalized likelihood to estimate relative divergence times in the ape package (Paradis and Schliep, 2019) in R (R 4.1.3; R Core Team, 2013), with the *chronos* function with the following parameters: lambda = 1, model = "correlated". No fossil constraints were used as we were not attempting to infer absolute divergence times. The results were visualized using the revGadgets package (Tribble *et al.*, 2021). The ChromoEvol model used was the same as the one implemented in RevBayes on the "Chromosome Evolution Tutorial" (https://revbayes.github.io/tutorials/chromo/) with the additional inclusion of parameters for demi-polyploidy included. We assessed convergence of the ancestral state analysis by checking that the effective sample sizes (ESSs) for all parameters were greater than 100 using the program Tracer (Rambaut *et al.*, 2018).

For comparison with the above approaches, we used *wgd* (Zwaenepoel and Van de Peer, 2019) to infer WGDs from *Ks* plots using the default parameters. Putative WGDs were identified by visual inspection (Yang *et al.*, 2015) for peaks that deviate from the expected exponential distribution of distances among paralogs, which, while subjective, is more conservative than mixture-model-based approaches, which often overestimate WGD events (Tiley, Barker, and Burleigh, 2018). *Ks* plots can be found in the supplemental data.

Results

The ML concatenated-data phylogeny based on the large dataset, the species tree based on the medium dataset, and the BI phylogeny based on the curated dataset all inferred identical and well-supported topologies for ingroup relationships (Fig. 1; Fig. S1). This is noteworthy, as the curated dataset used an order of magnitude fewer "genes". According to all phylogenies, the Parkerioideae plus Pteridoideae clade is well supported as sister to the rest of Pteridaceae, and Cryptogrammoideae is sister to Vittarioideae plus Cheilanthoideae. Support values were generally high, with one exception: in the BI tree from the small dataset (Fig. 1), relationships among *Cheilanthes chusana*, *C. nitidula*, and *Aleuritopteris chrysophylla* are not strongly supported. Among the outgroups to

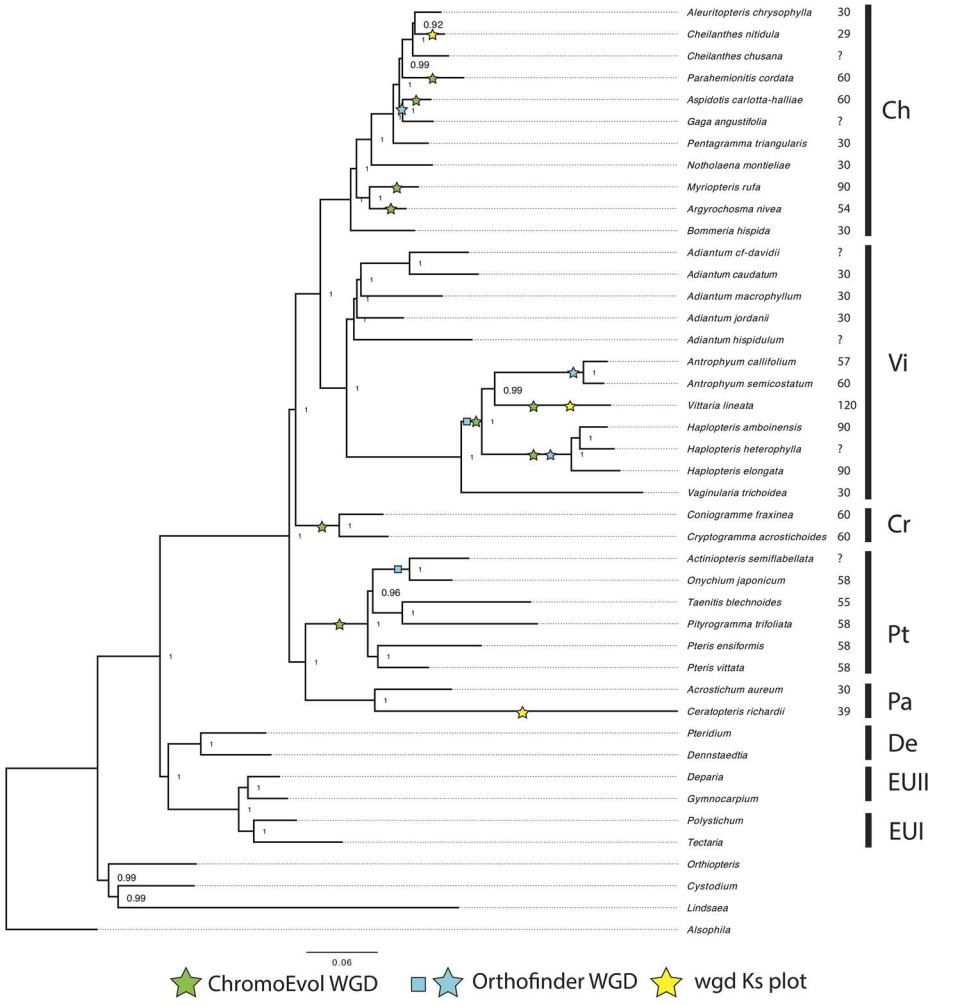

FIG. 1. Maximum a posteriori (MAP) tree based on 10 genes partitioned by gene and codon position ("small" dataset). Posterior probability values shown above branches. Branch lengths are in expected substitutions per site. Major clades in Pteridaceae labelled: Cheilanthoideae (Ch), Vittarioideae (Vi), Cryptogrammoideae (Cr), Pteridoideae (Pt), Parkerioideae (Pa). WGD events inferred by ChromoEvol with more than 75% PP are represented with a green star. WGD events inferred by OrthoFinder are represented by a blue star if there were a large number of genes (>1,000) supporting the event or a square if there were fewer genes (>500, <1,000) supporting the event (see Figure S1 and S2). WGD events inferred for outgroups are not shown as we did not include all the outgroups in WGD analyses. Chromosome numbers (Table 1) are shown for taxa for which we have counts available. De = Dennstaedtiaceae; EuI = Eupolypods I (Polypodiineae); EuII = Eupolypods II (Aspleniineae).

Pteridaceae, Dennstaedtiaceae (*Dennstaedtia*, *Pteridium*) is inferred as sister to the eupolypods (*Deparia*, *Gymnocarpium*, *Polystichum*, *Tectaria*) in the ASTRAL tree (Fig. S2) and the BI tree, but sister to Pteridaceae in the ML tree.

The OrthoFinder output (Fig. 1) displayed a large number of basal duplications (*i.e.*, along the stem of our tree and therefore not necessarily along the

stem of Pteridaceae and therefore not shown on figure; this included 2,902 gene duplications), plus additional duplications within Pteridaceae. Larger duplication events (more than 1,000 gene duplications) were inferred by OrthoFinder at the following nodes: *Aspidotis* + *Gaga*, the common ancestor of *Antrophyum*, and the common ancestor of *Haplopteris* (Fig. 1). Duplication events with fewer duplications (between 500 and 1,000 gene duplications) were inferred at the base of *Haplopteris* + *Vittaria* + *Antrophyum* and at *Onychium* + *Actiniopteris* (Fig. 1). Duplications less than 500 at a node were not mapped to the phylogeny nor considered further but are shown in Supplemental materials (Fig. S3). Our ChromoEvol analysis corroborated the OrthoFinder WGD in *Haplopteris* (Fig. 1), but also inferred a duplication shared by all the Pteridoideae, several events in Cheilanthoideae, and a weakly supported event at the base of Cryptogrammoideae (Fig. S4). Our *Ks* plot analysis inferred terminal WGDs in three lineages: *Cheilanthes nitidula*, *Vittaria lineata*, and *Ceratopteris richardii*.

Discussion

Our analyses support the five major clades of Pteridaceae inferred by Schuettpelz *et al.* (2007), which itself represented a major reassessment of the group and challenged earlier treatments. Both the phylotranscriptomic (large) and curated (medium and small) datasets and approaches support the same topology (Fig. 1). While we corroborate the sister relationships of Vittarioideae to Cheilanthoideae and Parkerioideae to Pteridoideae, we find Cryptogrammoideae is not the sister lineage to the rest of the Pteridaceae—as found in Schuettpelz *et al.* (2007), Kuo *et al.* (2011), and Schuettpelz and Pryer (2007)—but instead is sister to the Vittarioideae plus Cheilanthoideae.

Some previous studies with smaller numbers of genes have overlapping taxon sampling with the present study but reached different conclusions to varying degrees (Fig. 2). Our findings agree in part (*i.e.*, with the exception of the placement of Cryptogrammoideae) with the plastid-only inferences of Schuettpelz *et al.* (2007) and Schuettpelz and Pryer (2007). Rothfels *et al.* (2015) used a curated phylogenomics approach to infer relationships and found a congruent topology but lacked sampling of Parkerioideae and did not find strong support for the backbone. Wolf *et al.* (2018) used target sequence capture of nuclear-encoded genes (again lacking members of Parkerioideae) and inferred the same topology as Rothfels *et al.* (2015). Testo and Sundue (2016) found, based on six chloroplast regions, quite a different topology from our results, with Parkerioideae sister to remaining Pteridaceae; Cryptogrammoideae sister to Pteridoideae, Vittarioideae, and Cheilanthoideae; and Pteridoideae sister to Vittarioideae plus Cheilanthoideae. Their study, however, was fern-wide in scope (it included 4,000 species).

Most recent studies using transcriptome and/or plastome datasets have converged on a similar taxonomic understanding to the one put forward here. Shen *et al.* (2018) used transcriptome data and found relationships consistent with our

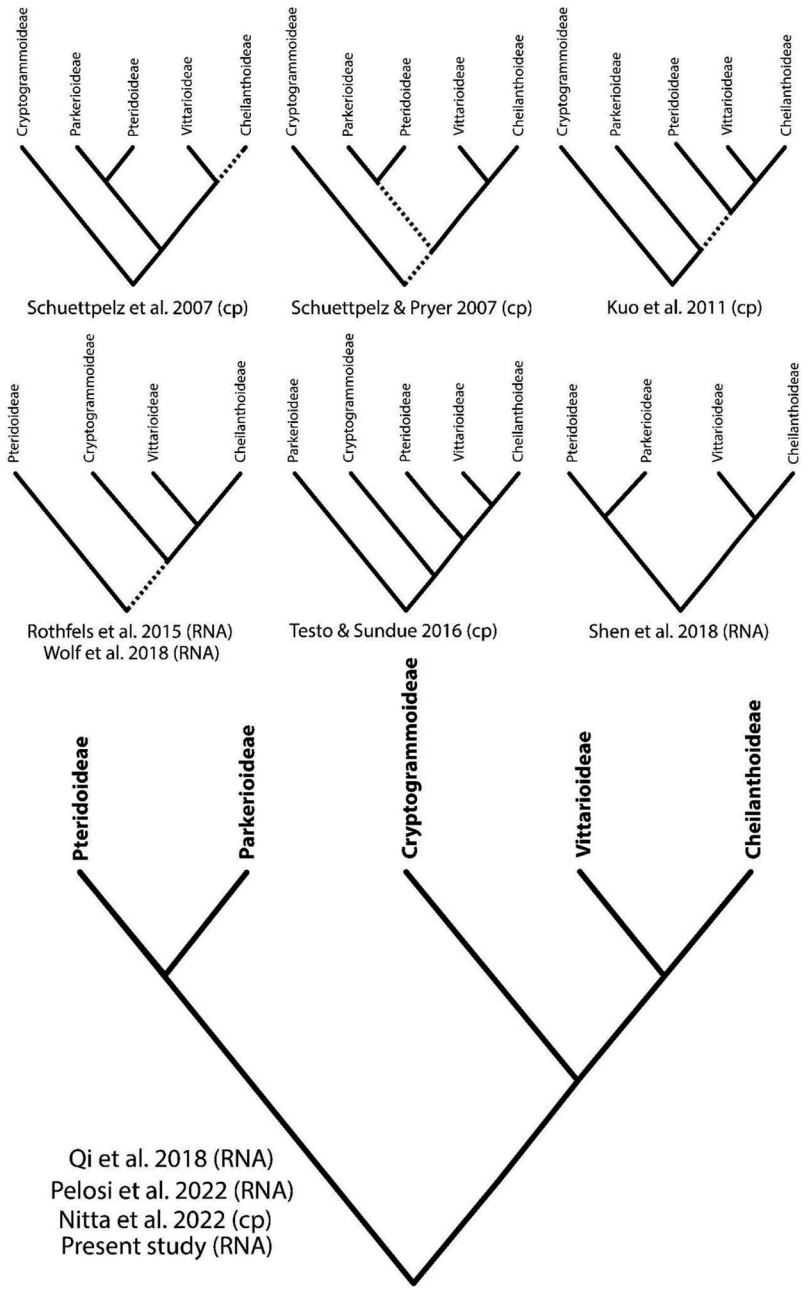

FIG. 2. Cladograms showing relationships among Pteridaceae subfamilies found in sources mentioned in the text—Schuettpelz *et al.* (2007), Schuettpelz and Pryer (2007), Kuo *et al.* (2011), Rothfels *et al.* (2015), Wolf *et al.* (2018), Testo and Sundue (2016), Shen *et al.* (2018), Qi *et al.* (2018), Pelosi *et al.* (2022), Nitta *et al.* (2022)—and in the present study. Studies are marked as chloroplast (cp) or transcriptome (RNA) datasets.

study but lacked sampling from Cryptogrammoideae. Notably, Qi *et al.* (2018) who also used transcriptome data, inferred the same topology found in the present study albeit with a sampling of only eight Pteridaceae taxa. Using transcriptome data and largely overlapping sampling, Pelosi *et al.* (2022) recovered the same backbone topology as the present study, with notable differences including a divergent placement for *Onychium* as well as the addition in our study of five genera: *Actiniopteris*, *Aspidotis*, *Bommeria*, *Pentagramma*, and *Vaginularia* (Appendix). Nitta *et al.* (2022), using both plastomes and Sanger-sequenced genes and comparatively dense, though again different, taxonomic sampling, found the same backbone topology as the present study.

Another historically contentious backbone relationship, that between Pteridaceae, Dennstaedtiaceae, and eupolypods (summarized in Shen *et al.*, 2018), can be examined here through the inclusion of our outgroup samples. In contrast to some other studies where Dennstaedtiaceae was supported as sister to Pteridaceae + eupolypods (Schuettpelz and Pryer, 2007; Kuo *et al.*, 2011; Testo and Sundue, 2016) or as sister to Pteridaceae (Rothfels *et al.*, 2013; Du *et al.*, 2021, 2022), Dennstaedtiaceae is here inferred as sister to the eupolypods by our BI and ASTRAL analysis (Figure 1). The same result was supported by Rothfels *et al.* (2015), Shen *et al.* (2018), Qi *et al.* (2018), Wolf *et al.* (2018), Huang *et al.* (2020), Nitta *et al.* (2022), and Pelosi *et al.* (2022). However, our ML tree supports Dennstaedtiaceae as sister to Pteridaceae, and demonstrates the potential difficulties reconciling big data approaches with more curated approaches. Nonetheless, increased taxonomic sampling will likely be necessary to fully resolve these relationships.

Intrageneric polyploidization is common in Pteridaceae (Windham and Yatskievych, 2003; Rothfels, 2008; Grusz, Windham, and Pryer, 2009; Beck, Windham, and Pryer, 2011; Sigel *et al.*, 2011; Schuettpelz *et al.*, 2015; Kao *et al.*, 2019; Adjie *et al.*, 2007; Chao *et al.*, 2012a; Chao *et al.*, 2012b; Jaruwattanaphan, Matsumoto, and Watano, 2013), but the extent of older, backbone polyploidization is unknown. The One Thousand Plant Transcriptomes Initiative (OneKP, 2019) used the tree-based Multi-tAxon Paleopolyploidy Search (MAPS, Li *et al.*, 2015) and *Ks* plots to test for WGDs across the fern phylogeny; they found recent (*i.e.*, terminal in their sampling) WGD events supported by *Ks* plots, but not by MAPS, in *Vittaria lineata*, *Adiantum raddianum*, and *Ceratopteris thalictroides*. Our *wgd* analyses support the same WGD event in *Ceratopteris*, though we sampled *C. richardii* (Fig. 1). Similarly, using MAPS and *Ks* plots, Marchant *et al.* (2019) found evidence for a WGD $>$ 100 Mya in *Ceratopteris richardii*. The placement of this WGD event was revised in a subsequent study (Marchant *et al.*, 2022) incorporating additional genomes and using MAPS, *Ks* plots, and NOTUNG (Chen *et al.*, 2000), to the stem of *Ceratopteris richardii* + *C. pteridoides* ~60 Mya. Our analyses also support a WGD in *Vittaria lineata*; our sampling lacks *Adiantum raddianum* and we found no evidence of WGD in our *Adiantum* samples. Likewise, Fang *et al.* (2022) analyzed the genome of *Adiantum capillus-veneris* and only found evidence for an ancient WGD on the branch leading to core leptosporangiate ferns.

Using both *Ks* plots and gene tree-species tree reconciliation, Huang *et al.* (2020) found no evidence for WGD events in the family. In contrast, Pelosi *et al.* (2022), using MAPS and *Ks* plots, found evidence for multiple WGDs in a duplication at the base of *Antrophyum* + *Vittaria* + *Haplopteris* that was supported by both their analysis approaches. Our OrthoFinder analyses support a WGD at the same place on the tree (although, notably, our sampling includes *Vaginularia*; Fig. 1). Pelosi *et al.* (2022) also found, supported by *Ks* plots, two more WGD events, in *Adiantum raddianum* and *Ceratopteris thalictroides*, the same as supported in OneKP (2019). Our various approaches (ChromoEvol, *Ks* plots, and OrthoFinder) show a cluster of inferred WGDs in the *Antrophyum* + *Vittaria* + *Haplopteris* lineage. Given the difficulties of pinpointing exactly where an ancient WGD event occurs (Zwaenepoel and Van de Peer, 2019), we may interpret these results together as supporting at least one ancient WGD event in Vittarioideae. In some groups, such as Cheilanthoideae, which similarly have high rates of neo-polyploidy (newly formed polyploid lineages that arise in diploid populations and may face struggles toward establishment due, e.g., to minority cytotype exclusion), OrthoFinder identifies one non-terminal WGD event—in the common ancestor of *Aspidotis* and *Gaga*—which is not corroborated by any other approach or study. The *Ks* plots and ChromEvol analysis found five separate events on terminal branches in this group. This novel finding is difficult to explain; while we generated transcriptome data for *Aspidotis carlotta-halliae* and *Gaga angustifolia*, and others used *Gaga arizonica* sequence data (OneKP, 2019; Pelosi *et al.*, 2022), an ancestral duplication event in the history of the lineage should be detectable in any species of *Gaga* or *Aspidotis*. Similarly, OrthoFinder supports a WGD in the common ancestor of *Actiniopteris* and *Onychium* while the ChromoEvol results infer a WGD event shared by all of the Pteridoideae; again, the *Actiniopteris* data are newly collected for this study. Despite the fact that *Ks* plots should in theory be insensitive to taxon sampling, selection of taxa appears to play a significant role in the conclusions of WGD inference studies. This is evidenced by the different results among studies that have analyzed Pteridaceae at different sampling depths and with different, but overlapping, sampling (OneKP, 2019; Huang *et al.*, 2020; Pelosi *et al.*, 2022; present study). In addition to their value in the phylogenetic analyses, multiple complementary approaches were critical to our WGD inferences, as each of the WGD analyses found evidence of an ancient WGD event somewhere in Pteridaceae, but no event was corroborated by all of the approaches. These findings could be because these approaches are not very powerful and may not necessarily reflect conflict between the results. What could have been interpreted as an unequivocal result is now evidence for the large uncertainty of the occurrence and placement of WGDs in this clade (and, perhaps, in other studies using these tools).

Our study offers insight into the developing field of phylotranscriptomics, which utilizes large datasets derived from RNA sequencing to infer evolutionary relationships. Such large datasets pose risks of systematic error due to challenges

such as homology inference (Walker *et al.*, 2018), ortholog inference (Brown and Thomson, 2017), and technical problems introduced through the difficulties of model selection and of performing inference under complex models with large datasets (Redmond and McLysaght, 2021). In our analyses, a large unpartitioned phylotranscriptomic dataset was analyzed as well as a smaller curated dataset with a partitioned more complex and biologically realistic model and found the same topology. However, our analyses found different relationships among the outgroups. While fitting models is difficult, new methods such as incorporating site-heterogeneous models and amino acid recoding into partitioned analyses could be used to assess the impact of systematic errors and would be a future direction to resolve this outgroup relationship. As phylotranscriptomic methods are positioned as a one-stop shop for resolving phylogenetic problems across every level of organizational hierarchy from species or genera (*e.g.*, Yu *et al.*, 2017; Zhang *et al.*, 2022) to all plants (Wickett *et al.*, 2014; OneKP, 2019), our study suggests that it is helpful to also support such findings with a smaller curated dataset; approaches that are designed for minimizing stochastic error are thereby complemented by those that are designed for minimizing systematic error.

Acknowledgements

The authors would like to thank Gabe Johnson, David Adelhelm, and Mike May for technical support, Forrest Freund for sample collection and handling, D. Blaine Marchant for *Ceratopteris* transcriptomes, Ya Yang for help with transcriptome assembly, and Chien-Hsun Huang and Hong Ma for sharing assembled transcriptomes. We are grateful to Editor Chris Haufler and two anonymous reviewers, who provided constructive comments on an earlier version of the manuscript.—M. J. Song, C. J. Rothfels, E. Schuettpelz, J. Nitta, L. Huiet, F.-W. Li, and K. M. Wefferling.

Literature Cited

Adjie, B., S. Masuyama, H. Ishikawa, and Y. Watano. 2007. Independent origins of tetraploid cryptic species in the fern *Ceratopteris thalictroides*. Journal of Plant Research 120:129–138.

Ammal, L. and K. Bhavanandan. 1992. Studies on the cytology of some ferns from south India. Indian Fern Journal 9:94–101.

Beck, J. B., M. D. Windham, and K. M. Pryer. 2011. Do asexual polyploid lineages lead short evolutionary lives? a case study from the fern genus *Astrolepis*. Evolution 65:3217–3229.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brown, J. M. and R. C. Thomson. 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Systematic Biology 66:517–530.

Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59–60.

Carpenter, E. J., N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F. R. Jimenez Vieira, C. Bowler, R. G. Dorrell, M. A. Gitzendanner, L. Li, W. Du, *et al.* 2019. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant Transcriptomes Initiative (OneKP). GigaScience 8: giz126.

Cave, M. S. 1959. *Index to plant chromosome numbers supplement*. California Botanical Society, Berkeley, CA, USA.

Cave, M. S. 1964. *Index to Plant Chromosome Numbers for 1963*. California Botanical Society, Berkeley.

CHAO, Y. S., S. Y. DONG, Y. C. CHIANG, H. Y. LIU, and W. L. CHIOU. 2012. Extreme multiple reticulate origins of the *Pteris cadieri* complex (Pteridaceae). International Journal of Molecular Sciences 13:4523–4544.

CHAO, Y. S., H. Y. LIU, Y. C. CHIANG, and W. L. CHIOU. 2012. Polyploidy and speciation in *Pteris* (Pteridaceae). Journal of Botany 2012:1–7.

CHEN, K., D. DURAND, and M. FARACH-COLTON. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. Journal of Computational Biology 7:429–447.

DARNAEDI, D. 1992. A preliminary cytological study of fern flora of Gede-Pangrango National Park (West Java). In *Proceedings of the Second Seminar on Asian Pteridology, Taiwan*. National Chung Hsiang University and National Science Council, pp. 73–78.

DU, X.-Y., L.-Y. KUO, Z.-Y. ZUO, D.-Z. LI, and J.-M. LU. 2022. Structural variation of plastomes provides key insight into the deep phylogeny of ferns. Frontiers in Plant Science 13. https://doi.org/10.3389/fpls.2022.862772

DU, X.-Y., J.-M. LU, L.-B. ZHANG, J. WEN, L.-Y. KUO, C. M. MYNSSEN, H. SCHNEIDER, and D.-Z. LI. 2021. Simultaneous diversification of Polypodiales and angiosperms in the Mesozoic. Cladistics 37:518–539.

EMMS, D. M. and S. KELLY. 2017. STRIDE: Species tree root inference from gene duplication events. Molecular Biology and Evolution, 34:3267–3278.

EMMS, D. M. and S. KELLY. 2018. STAG: Species tree inference from all genes. BioRxiv 267914.

EMMS, D. M. and S. KELLY. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology 20:1–14.

FABRETI, L. G. and S. HÖHNA. 2022. Bayesian inference of phylogeny is robust to substitution model over-parameterization. bioRxiv 2022-02.

FANG, Y., X. QIN, Q. LIAO, R. DU, X. LUO, Q. ZHOU, Z. LI, H. CHEN, W. JIN, Y. YUAN, and P. SUN. 2022. The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. Nature Plants 8:1024–1037.

FREYMAN, W. A. and S. HÖHNA. 2018. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Systematic Biology 67:195–215.

GARDNER, E. M., M. G. JOHNSON, D. RAGONE, N. J. WICKETT, and N. J. ZEREGA. 2016. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Applications in Plant Sciences 4:1600017.

GASTONY, G. J. and W. P. JOHNSON. 2001. Phylogenetic placements of *Loxoscaphe thecifera* (Aspleniaceae) and *Actiniopteris radiata* (Pteridaceae) based on analysis of *rbcL* nucleotide sequences. American Fern Journal 91:197–213.

GLICK, L. and I. MAYROSE. 2014. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular Biology and Evolution 31.7: 1914–1922.

GRABHERR, M. G., B. J. HAAS, M. YASSOUR, J. Z. LEVIN, D. A. THOMPSON, I. AMIT, X. ADICONIS, L. FAN, R. RAYCHOWDHURY, Q. ZENG, Z. CHEN, E. MAUCELI, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644.

GRUSZ, A. L., M. D. WINDHAM, and K. M. PRYER. 2009. Deciphering the origins of apomictic polyploids in the *Cheilanthes yavapensis* complex (Pteridaceae). American Journal of Botany 96: 1636–1645.

HAAS, B. 2021. TransDecoder. https://github.com/TransDecoder/TransDecoder. [Online; accessed 01-March-2021].

HASEBE, M., P. G. WOLF, K. M. PRYER, K. UEDA, M. ITO, R. SANO, G. J. GASTONY, J. YOKOYAMA, J. R. MANHART, N. MURAKAMI, E. H. CRANE, C. H. HAUFLER, and W. D. HAUK. 1995. Fern phylogeny based on *rbcL* nucleotide sequences. American Fern Journal 85:134–181.

HÖHNA, S., M. J. LANDIS, T. A. HEATH, B. BOUSSAU, N. LARTILLOT, B. R. MOORE, J. P. HUELSENBECK, and F. RONQUIST. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726–736.

HUANG, C.-H., X. QI, D. CHEN, J. QI, and H. MA. 2020. Recurrent genome duplication events likely contributed to both the ancient and recent rise of ferns. Journal of Integrative Plant Biology 62:433–455.

JARUWATTANAPHAN, T., S. MATSUMOTO, and Y. WATANO. 2013. Reconstructing hybrid speciation events in the *Pteris cretica* group (Pteridaceae) in Japan and adjacent regions. Systematic Botany 38: 15–27.

JERMY, A. C. and T. G. WALKER. 1985. *Cytotaxonomic studies of the ferns of Trinidad*. British Museum (Natural History).

KAO, T.-T., K. M. PRYER, F. D. FREUND, M. D. WINDHAM, and C. J. ROTHFELS. 2019. Low-copy nuclear sequence data confirm complex patterns of farina evolution in notholaenid ferns (Pteridaceae). Molecular Phylogenetics and Evolution 138:139–155.

KATO, M. 1999. A cytotaxonomic study of Hainan (S. China) pteridophytes with notes on polyploidy and apogamy of Chinese species. Ching Memorial Volume 1–19.

KATO, M., N. NAKATO, X. CHENG, and K. IWATSUKI. 1992. Cytotaxonomic study of ferns of Yunnan, southwestern China. The Botanical Magazine=Shokubutsu-gaku-zasshi 105:105–124.

KATOH, K. and D. M. STANDLEY. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772–780.

KELCHNER, S. A. and M. A. THOMAS. 2007. Model use in phylogenetics: nine key questions. Trends in Ecology and Evolution 22:87–94.

KNOBLOCH, I. W., W. TAI, and T. N. ADANGAPPURAM. 1975. Chromosome counts in *Cheilanthes* and *Aspidotis* with a conspectus of the cytology of the Sinopteridaceae. American Journal of Botany 62:649–654.

KRAMER, K. U., P. S. GREEN, and E. GÖTZ. 2013. *Pteridophytes and Gymnosperms, Volume 1*. Springer.

KUO, L.-Y., F.-W. LI, W.-L. CHIOU, and C.-N. WANG. 2011. First insights into fern *matK* phylogeny. Molecular Phylogenetics and Evolution 59:556–566.

LAMESCH, P., T. Z. BERARDINI, D. LI, D. SWARBRECK, C. WILKS, R. SASIDHARAN, R. MULLER, K. DREHER, D. L. ALEXANDER, M. GARCIA-HERNANDEZ, A. S. KARTHIKEYAN, C. H. LEE, *et al.* 2012. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40:D1202–D1210.

LEHTONEN, S. 2011. Towards resolving the complete fern tree of life. *PLoS ONE* 6: e24851.

LI, W. and A. GODZIK. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

LI, Z. and M. S. BARKER. 2020. Inferring putative ancient whole-genome duplications in the 1000 plants (OneKP) initiative: access to gene family phylogenies and age distributions. GigaScience 9:giaa004.

LI, F. W., P. BROUWER, L. CARRETERO-PAULET, S. CHENG, J. DE VRIES, P. M. DELAUX, A. EILY, N. KOPPERS, L.-Y. KUO, Z. LI, M. SIMENC, I. SMALL, *et al.* 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nature Plants 4:460–472.

LI, Z., A. E. BANIAGA, E. B. SESSA, M. SCASCITELLI, S. W. GRAHAM, L. H. RIESEBERG, and M. S. BARKER. 2015. Early genome duplications in conifers and other seed plants. Science Advances 1: e1501084.

LÖVE, A., D. LÖVE, and R. PICHI SERMOLLI. 1977. *Cytotaxonomical Atlas of the Pteridophyt*a, Vaduz: J. Cramer, 398p.

MARCHANT, D. B., E. B. SESSA, P. G. WOLF, K. HEO, W. B. BARBAZUK, P. S. SOLTIS, and D. E. SOLTIS. 2019. The C-Fern (*Ceratopteris richardii*) genome: Insights into plant genome evolution with the first partial homosporous fern genome assembly. Scientific Reports 9:1–14.

MARCHANT, D. B., G. CHEN, S. CAI, F. CHEN, P. SCHAFRAN, J. JENKINS, S. SHU, C. PLOTT, J. WEBBER, J. T. LOVELL, G. HE, L. SANDOR, *et al.* 2022. Dynamic genome evolution in a model fern. Nature Plants 8:1038–1051.

MARCON, A. B., I. C. BARROS, and M. GUERRA. 2003. A karyotype comparison between two closely related species of *Acrostichum*. American Fern Journal 93:116–125.

MCKAIN, M. R., M. G. JOHNSON, S. URIBE-CONVERS, D. EATON, and Y. YANG. 2018. Practical considerations for plant phylogenomics. Applications in Plant Sciences 6:e1038.

MILLER, M. A., W. PFEIFFER, and T. SCHWARTZ. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: 2010 *Gateway Computing Environments Workshop (GCE)*:1–8.

MORALES-BRIONES, D. F., G. KADEREIT, D. T. TEFARIKIS, M. J. MOORE, S. A. SMITH, S. F. BROCKINGTON, A. TIMONEDA, W. C. YIM, J. C. CUSHMAN, and Y. YANG. 2021. Disentangling sources of gene tree

discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae s.l. Systematic Biology 70:219–235.

Nitta, J. 2020. *baitfindR*. https://github.com/joelnitta/baitfindR. [Online; accessed 01-March-2021].

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

One Thousand Plant Transcriptomes Initiative (OneKP). 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574:679–685.

Pajaron, S., E. Pangua, and L. Garcia Alvarez. 1999. Sexual expression and genetic diversity in populations of *Cryptogramma crispa* (Pteridaceae). American Journal of Botany 86:964–973.

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

Pelosi, J. A., E. H. Kim, W. B. Barbazuk, and E. B. Sessa. 2022. Phylotranscriptomics illuminates the placement of whole genome duplications and gene retention in ferns. Frontiers in Plant Science 2356.

Phillips, M. J., F. Delsuc, and D. Penny. 2004. Genome-scale phylogeny and the detection of systematic biases. Molecular Biology and Evolution 21:1455–1458.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biology 9: e1000602.

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

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One 5:e9490.

Pryer, K. M., A. R. Smith, and J. E. Skog. 1995. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. American Fern Journal 85:205–282.

Qi, X., L.-Y. Kuo, C. Guo, H. Li, Z. Li, J. Qi, L. Wang, Y. Hu, J. Xiang, C. Zhang, J. Guo, C. Huang, H. Ma. 2018. A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. Molecular Phylogenetics and Evolution 127:961–977.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67:901.

Redmond, A. K. and A. McLysaght. 2021. Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. Nature Communications 12:1783.

Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J. Mayzel, O. Chay, and I. Mayrose. 2015. The chromosome counts database (CCDB)–a community resource of plant chromosome numbers. New Phytologist 20619–26.

Rothfels, C. 2008. Pteridaceae EDM Kirchn. 1831. Brake Ferns, Maidenhair Ferns, and allies. http://tolweb.org/Pteridaceae. [Online; accessed 01-March-2021].

Rothfels, C. J., A. Larsson, L.-Y. Kuo, P. Korall, W.-L. Chiou, and K. M. Pryer. 2012. Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. Systematic Biology 61:490–509.

Rothfels, C. J., A. Larsson, F.-W. Li, E. M. Sigel, L. Huiet, D. O. Burge, M. Ruhsam, S. W. Graham, D. W. Stevenson, G. K.-S. Wong, *et al.* 2013. Transcriptome-mining for single-copy nuclear markers in ferns. PloS One 8:e76957.

Rothfels, C. J., F.-W. Li, E. M. Sigel, L. Huiet, A. Larsson, D. O. Burge, M. Ruhsam, M. Deyholos, D. E. Soltis, C. N. Stewart Jr, S. W. Shaw, L. Pokorny, T. Chen, C. dePamphilis, L. DeGironimo, *et al.* 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. American Journal of Botany 102:1089–1107.

Schneider, H., E. Schuettpelz, K. M. Pryer, R. Cranfill, S. Magallon, and R. Lupia. 2004. Ferns diversified in the shadow of angiosperms. Nature 428:553–557.

Schuettpelz, E. and K. M. Pryer. 2007. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56:1037–1050.

Schuettpelz, E. and K. M. Pryer. 2009. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. Proceedings of the National Academy of Sciences 106:11200–11205.

Schuettpelz, E., K. M. Pryer, and M. D. Windham. 2015. A unified approach to taxonomic delimitation in the fern genus *Pentagramma* (Pteridaceae). Systematic Botany 40:629–644.

Schuettpelz, E., H. Schneider, L. Huiet, M. D. Windham, and K. M. Pryer. 2007. A molecular phylogeny of the fern family Pteridaceae: Assessing overall relationships and the affinities of previously unsampled genera. Molecular Phylogenetics and Evolution 44:1172–1185.

Shen, H., D. Jin, J.-P. Shu, X.-L. Zhou, M. Lei, R. Wei, H. Shang, H.-J. Wei, R. Zhang, L. Liu, Y. Gu, X. Zhang, Y. Yan. 2018. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. GigaScience 7:gix116.

Sigel, E. M., M. D. Windham, L. Huiet, G. Yatskievych, and K. M. Pryer. 2011. Species relationships and farina evolution in the cheilanthoid fern genus *Argyrochosma* (Pteridaceae). Systematic Botany 36:554–564.

Smith, A. R. 1974. Taxonomic and cytological notes on ferns from California and Arizona. Madroño 22:376–378.

Smith, A. R., K. M. Pryer, E. Schuettpelz, P. Korall, H. Schneider, and P. G. Wolf. 2006. A classification for extant ferns. Taxon 55:705–731.

Srivastava, R. 1985. Ferns of the Indo-Nepal border. Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences 86:471–471.

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

Stamatakis, A. 2015. Using raxml to infer phylogenies. Current Protocols in Bioinformatics 51:6–14.

R Core Team. 2013. R: A language and environment for statistical computing.

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

Tiley, G. P., M. S. Barker, and J. G. Burleigh. 2018. Assessing the performance of *Ks* plots for detecting ancient whole genome duplications. Genome Biology and Evolution 10:2882–2898.

Tribble, C. M., W. A. Freyman, M. J. Landis, J. Y. Lim, J. Barido-Sottani, B. T. Kopperud, S. Höhna, and M. R. May. 2021. RevGadgets: an R Package for visualizing Bayesian phylogenetic analyses from RevBayes. Methods in Ecology and Evolution 13:314–323.

Walker, J. F., Y. Yang, T. Feng, A. Timoneda, J. Mikenas, V. Hutchison, C. Edwards, N. Wang, S. Ahluwalia, J. Olivieri, and N. Walker-Hale. 2018. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. American Journal of Botany, 105:446-462.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, M. S. Barker, J. G. Burleigh, M. A. Gitzendanner, B. R. Ruhfel, E. Wafula, J. P. Der, S. W. Graham, S. Mathews, *et al.* 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences 111:E4859–E4868.

Windham, M. D. 1993. Pteridaceae. In: *Flora of North America North of Mexico*, *Vol.* 2. Flora of North America Editorial Committee, eds. Springer.

Windham, M. D. and G. Yatskievych. 2003. Chromosome studies of cheilanthoid ferns (Pteridaceae: Cheilanthoideae) from the western United States and Mexico. American Journal of Botany 90:1788–1800.

Wolf, P. G., T. A. Robison, M. G. Johnson, M. A. Sundue, W. L. Testo, and C. J. Rothfels. 2018. Target sequence capture of nuclear-encoded genes for phylogenetic analysis in ferns. Applications in Plant Sciences 6:e01148.

Yang, Y., M. J. Moore, S. F. Brockington, J. Mikenas, J. Olivieri, J. F. Walker, and S. A. Smith. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. New Phytologist 217:855–870.

Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K.-S. Wong, E. J. Carpenter, Y. Zhang, L. Chen, Z. Yan, Y. Xie, R. F. Sage, S. Covshoff, J. M. Hibberd, M. N. Nelson, and S. A. Smith. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Molecular Biology and Evolution 32:2001–2014.

Yang, Y. and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. Molecular Biology and Evolution 31:3081–3092.

Yu, Y., Q. Xiang, P. S. Manos, D. E. Soltis, P. S. Soltis, B.-H. Song, S. Cheng, X. Liu, and G. Wong. 2017. Whole-genome duplication and molecular evolution in *Cornus* L. (Cornaceae)–Insights from transcriptome sequences. PLoS One 12:e0171361.

ZHANG, L., X. ZHU, Y. ZHAO, J. GUO, T. ZHANG, W. HUANG, J. HUANG, Y. HU, C.-H. HUANG, and H. MA. 2022. Phylotranscriptomics resolves the phylogeny of Pooideae and uncovers factors for their adaptive evolution. Molecular Biology and Evolution 39:msac026.

ZHANG, C., M. RABIEE, E. SAYYARI, and S. MIRARAB. 2018. Astral-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:15–30.

ZHANG, Z., Z. HE, S. XU, X. LI, W. GUO, Y. YANG, C. ZHONG, R. ZHOU, and S. SHI. 2016. Transcriptome analyses provide insights into the phylogeny and adaptive evolution of the mangrove fern genus *Acrostichum*. Scientific Reports 6:1–10.

ZWAENEPOEL, A. and Y. VAN DE PEER. 2019. wgd—simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics 35:2153–2155.

APPENDIX

Transcriptomes generated in this study:

Bioproject number PRJNA716637.— *Actiniopteris semiflabellata*, collector unknown (UCBG 2006.0030; voucher at UC, Smith 26-I-2006), missing locality (Africa), sporophyte, SAMN18442407; *Antrophyum semicostatum*, E. Schuettpelz 1561 (vouchers at BO, UC, US), Indonesia: West Java, sporophyte, SAMN18442408; *Aspidotis carlotta-halliae*, C. J. Rothfels 4621 (voucher at UC), USA: California, gametophyte, SAMN18442409; *Bommeria hispida*, S. B. Hogan 4925 (UCBG 92.0086, voucher at UC, Welch 7-XI-2005), USA: Arizona, sporophyte, SAMN 18442410; *Coniogramme fraxinea*, E. Schuettpelz 1509 (vouchers at BO, US), Indonesia: West Java, sporophyte, SAMN18442411; *Gaga angustifolia*, C. J. Rothfels 3117B (vouchers at DUKE, MEXU), Mexico: Jalisco, gametophyte, SAMN18442412; *Haplopteris elongata*, E. Schuettpelz 1559 (vouchers at BO, UC, US), Indonesia: West Java, sporophyte, SAMN18442413; *Pentagramma triangularis*, E. Schuettpelz 1277A (voucher at DUKE), USA: California, gametophyte, SAMN18442414; *Pentagramma triangularis*, M. Murphy s.n. (voucher at UWGB), USA: Washington, sporophyte, SAMN18442415; *Vaginularia trichoidea*, E. Schuettpelz 1553 (vouchers at BO, UC, US), Indonesia: West Java, sporophyte, SAMN18442416.

Bioproject number PRJNA821853.— *Adiantum* cf. *davidii*, H. Hansen s.n. (UCBG 2006.0341; voucher at UC, 2014-08-05), China: Sichuan, sporophyte, SAMN28561067; *Adiantum hispidulum*, collector unknown, from Java Botanic Garden (UCBG 57.0774; voucher at UC, Huiet 101), Indonesia, sporophyte, SAMN28561064; *Adiantum jordanii*, B. Anderson s.n. (UCBG 2011.0496, no voucher), USA: California, sporophyte, SAMN28561065; *Adiantum macrophyllum*, M. Grantham and J. Parsons 0270-90 (UCBG 90.2361; voucher at UC, Huiet 102), Costa Rica: Puntarenas, sporophyte, SAMN28561066.

Transcriptomes from other studies:

Carpenter et al. (2019), raw, cleaned, SRA reads: *Argyrochosma nivea*, PRJEB 21674, SAMEA104170982; *Cryptogramma acrostichoides*, PRJEB21674, SAMEA104170976; *Cystopteris fragilis* (outgroup, OG), PRJEB21674, SAME A104170967; *Deparia lobato-crenata* (OG), PRJEB21674, SAMEA104170964;

Lindsaea linearis (OG), PRJEB21674, SAMEA104170986; *Myriopteris rufa*, PRJEB21674, SAMEA104170983; *Notholaena montieliae*, PRJEB21674, SAMEA 104170981; *Parahemionitis cordata*, PRJEB21674, SAMEA104170977; *Pityrogramma trifoliata*, PRJEB21674, SAMEA104170980; *Polystichum acrostichoides* (OG), PRJEB21674, SAMEA104170951; *Pteris ensiformis*, PRJEB 21674, SAMEA104170979; *Pteris vittata*, PRJEB21674, SAMEA104170978; *Struthiopteris spicant* (OG), PRJEB21674, SAMEA104170961; *Vittaria lineata*, PRJEB21674, SAMEA104170984.

Marchant et al. (2019), Blaine Marchant shared an assembled transcriptome: *Ceratopteris richardii*, PRJNA511033, SAMN10638562.

Qi et al. (2018), Chien-Hsun Huang and Hong Ma shared assembled transcriptomes: *Alsophila podophylla* (OG), PRJNA422112, SAMN08805118; *Cheilanthes nitidula*, PRJNA422112, SAMN08805103; *Dennstaedtia hirsuta* (OG), PRJNA422112, SAMN08805100*; Deparia petersenii* (OG), PRJNA422112, SAMN08805075; *Gymnocarpium oyamense* (OG), PRJNA422112, SAMN08805089; *Haplopteris heterophylla*, PRJNA422112, SAMN08805106; *Onychium japonicum*, PRJNA422112, SAMN08805109; *Pteridium revolutum* (OG), PRJNA422112, SAMN 08805099; *Saccoloma campylurum* (OG), PRJNA422112, SAMN08805115; *Tectaria nayarii* (OG), PRJNA422112, SAMN08805054.

Shen et al. (2018), raw, cleaned, SRA reads: *Adiantum caudatum*, PRJNA281136, SAMN03575931; *Aleuritopteris chrysophylla*, PRJNA281136, SAMN03575929; *Antrophyum callifolium*, PRJNA281136, SAMN03575934; *Cheilanthes chusana*, PRJNA281136, SAMN03575930; *Haplopteris amboinensis*, PRJNA281136, SAMN03575935*; Taenitis blechnoides*, PRJNA281136, SAMN03575927.

Wolf et al. (2018), raw, cleaned, SRA reads: *Cystodium sorbifolium*, PRJNA432105, SAMN08434973.

Zhang et al. (2016), raw, cleaned, SRA reads: *Acrostichum aureum*, PRJNA276721, SAMN03380083.