

## An exploration into fern genome space

\*Paul G. Wolf<sup>1</sup>, Emily B. Sessa<sup>2,3</sup>, D. Blaine Marchant<sup>2,3,8</sup>, Fay-Wei Li<sup>4</sup>, Carl J. Rothfels<sup>5</sup>, Erin M. Sigel<sup>4,6</sup>, Mathew A. Gitzendanner<sup>2,3</sup>, Clayton J. Visger<sup>2,3</sup>, Jo Ann Banks<sup>7</sup>, Douglas E. Soltis<sup>2,3,8</sup>, Pamela S. Soltis<sup>3,8</sup>, Kathleen M. Pryer<sup>4</sup>, Joshua P. Der<sup>9</sup>

<sup>1</sup>Ecology Center and Department of Biology, Utah State University, Logan UT 84322, USA

<sup>2</sup>Department of Biology, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup>Genetics Institute, University of Florida, Gainesville, FL 32611, USA

<sup>4</sup>Department of Biology, Duke University, Durham, NC 27708, USA

<sup>5</sup>Current address: University Herbarium and Dept. of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720-2465, USA

<sup>6</sup>Current address: Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA

<sup>7</sup>Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47906

<sup>8</sup>Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 USA

<sup>9</sup>Department of Biological Science, California State University Fullerton, Fullerton, CA 92834, USA

\*Paul G. Wolf, Ecology Center and Department of Biology, Utah State University, Logan UT 84322, USA, phone (435) 797 4034, Fax (435) 797 1575, paul.wolf@usu.edu

## Abstract

Ferns are one of the few remaining major clades of land plants for which a complete genome sequence is lacking. Knowledge of genome space in ferns will enable broad-scale comparative analyses of land plant genes and genomes, provide insights into genome evolution across green plants, and shed light on genetic and genomic features that characterize ferns, such as their high chromosome numbers and large genome sizes. As part of an initial exploration into fern genome space, we used a whole genome shotgun sequencing approach to obtain low-density coverage ( $\sim 0.4X$  to  $2X$ ) for six fern species from the Polypodiales (*Ceratopteris*, *Pteridium*, *Polypodium*, *Cystopteris*), Cyatheaales (*Plagiogyria*), and Gleicheniales (*Dipteris*). We explore these data to characterize the proportion of the nuclear genome represented by repetitive sequences (including DNA transposons, retrotransposons, rDNA, and simple repeats) and protein-coding genes, and to extract chloroplast and mitochondrial genome sequences. Such initial sweeps of fern genomes can provide information useful for selecting a promising candidate fern species for whole genome sequencing. We also describe variation of genomic traits across our sample and highlight some differences and similarities in repeat structure between ferns and seed plants.

Keywords: comparative genomics, plastome, chloroplast, mitochondrial, repeat content, transposons.

## Introduction

Recent advances in DNA sequencing technology and improvements in assembly strategies are resulting in rapid growth in the availability of genome sequences for non-model species. Currently, genome sequences are available for over 100 vascular plants, including one lycopod, two gymnosperms, and numerous crop and non-crop angiosperms (Michael and VanBuren 2015). However, genomic resources in other major clades of vascular plants are lagging. The sister group of seed plants is the fern clade (Monilophyta sensu Cantino et al. 2007): these two lineages diverged from a common ancestor approximately 380 million years ago (Schneider et al. 2004). Ferns in the broad sense include horsetails, whisk ferns and ophioglossoid ferns, marattioid ferns, and leptosporangiate ferns. The latter lineage is by far the most diverse, with about 9,000 species (Smith et al. 2006) that occupy many key ecosystems, and comprise, for example, a significant component of tropical forest understories and canopies. There are currently no nuclear genome sequences available for any fern, despite the richness of this clade and its key evolutionary position as sister to the seed plants. Having such a reference sequence, or any information about the content and structure of fern genomes, will enable investigation of several perplexing features of fern biology and evolution, and will facilitate comparative analyses of genome structure and function across vascular plants as a whole.

Ferns differ fundamentally from seed plants in several key biological and genomic features. For example, ferns alternate between free-living, independent gametophyte (haploid) and sporophyte (diploid) phases, whereas the gametophyte phase in seed plants is dependent on the sporophyte and is highly reduced. Thus, a large component of the fern

genome is presumably expressed only in the haploid phase. Further, most ferns are homosporous (apart from one heterosporous clade), whereas all seed plants are heterosporous. This characteristic is apparently correlated with chromosome number across tracheophytes in that homosporous taxa uniformly have more chromosomes. For example, chromosome numbers average  $n = 57.05$  for homosporous ferns and lycophytes,  $n = 15.99$  for flowering plants, and  $n = 13.62$  for heterosporous ferns and lycophytes (Klekowski and Baker 1966). The underlying cause of this association between homospory and high chromosome number is not understood.

Ferns are the only lineage of land plants for which there is a strong positive correlation between chromosome number and genome size (Nakazato et al. 2008; Bainard et al. 2011). Whether this also extends to lycophytes is not yet clear, but this pattern has not been reported from any other group of eukaryotic organisms, and suggests that fern nuclear genomes may possess unique structural characteristics. Ferns are prone to polyploidization (Wood et al. 2009) but may undergo different diploidization processes that are distinct from those in other lineages of land plants (Barker and Wolf 2010; Leitch and Leitch 2012). Information on the nature and relative proportions of various components of fern genomes will help to establish how this group of plants can be used in studies of genome evolution and dynamics across land plants. If fern genomes respond uniquely to changes in genome size, then they could provide useful control models for the study of genome downsizing following whole genome duplication (Leitch and Bennett 2004).

Gathering information on fern genomes will provide an improved phylogenetic context for investigating evolutionary questions across land plants. For example,

knowledge of fern genome content and structure may shed light on the transition from homospority to heterospority that has occurred several times during the evolution of land plants. Ferns are also the most appropriate outgroup for understanding genome structure and evolution in their sister clade, the seed plants.

The research community would benefit from well-assembled, annotated nuclear genomes from several leptosporangiate ferns, as well as representatives of the other early-diverging fern clades (Li and Pryer 2014; Schneider et al. 2015; Sessa et al. 2014). Such nuclear genome sequences are necessary for rigorous tests of most questions about genome and chromosome structure and evolution, and addressing these questions currently awaits completion of one or more fern genome sequencing projects (Li and Pryer 2014; Sessa et al. 2014). Meanwhile, low-coverage genome scans can be used to begin uncovering broad patterns of fern genome content, allowing, for example, preliminary estimates of protein-coding and repetitive content. Here we use such scans at  $\sim 0.4\text{--}2X$  coverage for species from six different fern lineages, each representing a major leptosporangiate clade (Fig. 1). We use these data to ask how much variation exists in gene and repeat content across ferns, and we compare these with data from existing angiosperm and gymnosperm genome sequencing projects. Although our focus is on nuclear genomes of ferns, the data obtained include many organellar genome sequences, and we use these to assemble plastomes and identify contigs carrying putative mitochondrial genes. The latter are the first such resources available for ferns, for which no mitochondrial genome has been sequenced to date.

## Methods

## Samples

We selected six leptosporangiate ferns from across a range of major clades (Fig. 1): *Dipteris conjugata* (Gleicheniales), *Plagiogyria formosana* (Cyatheales), *Pteridium aquilinum* (Dennstaedtiaceae), *Ceratopteris richardii* (Pteridaceae), *Polypodium glycyrrhiza* (eupolypods I), and *Cystopteris protrusa* (eupolypods II). Details of species used, collections, and vouchers are provided in Table 1.

## Genome size estimation

Genome size estimates for three taxa were derived from the literature: *Polypodium glycyrrhiza* (Murray 1985), *Pteridium aquilinum* (Bainard et al. 2011) and *Cystopteris protrusa* (Bainard et al. 2011, for the related diploid, *C. bulbifera*). We estimated genome size for *Pteridium aquilinum* by chopping approximately 0.75 cm<sup>2</sup> of fresh fern leaf tissue and 0.5 cm<sup>2</sup> of the standard *Vicia faba* (26.9pg) or *Pisum sativum* (9.09pg; Doležel et al. 1998) on a chilled surface using a fresh razor blade and adding 500 $\mu$ L of ice cold extraction buffer (0.1 M citric acid, 0.5% v/v Triton X-100) (Hanson et al. 2005) with 1% w/v PVP-40 (Yokoya et al. 2000). Tissue was chopped into a semi-fine slurry, and the resulting mixture was swirled by hand until the liquid obtained a light-green tinge. The suspension was poured through a cell strainer (BD Falcon; Becton, Dickinson and Company, Franklin Lakes, NJ). We added RNaseA (1mg/ml) and 350 $\mu$ L of propidium iodide staining solution (0.4 M NaPO<sub>4</sub>, 10 mM sodium citrate, 25 mM sodium sulfate, 50  $\mu$ g/ml propidium iodide) to 140 $\mu$ L of filtrate, incubated the solution at 25 °C for 30 minutes, and then stored it for up to four hours on ice. We ran the stained solutions on an Accuri C6 (BD Biosciences, Franklin Lake, New Jersey) using a 488 nm laser and captured 10,000 events. For estimating genome size

of *Plagiogyria formosana* and *Dipteris conjugata*, we carried out flow cytometry analyses following the protocol of Ebihara et al. (2005) on a BD FACSCan system (BD Biosciences, Franklin Lake, New Jersey). We calculated the relative genome content using the ratio of the mean fluorescent peak of the sample to the internal standard, multiplied by the genome size of the standard, and converted to an estimate of the number of bases using  $1\text{ pg} = 980\text{ Mb}$ .

### DNA Sequencing

Genomic libraries for *Polypodium*, *Cystopteris*, *Plagiogyria*, and *Dipteris* were prepared with the KAPA Illumina library preparation kit (KAPA Biosystems, Wilmington, Massachusetts, USA) using fragment sizes of 300–400 bp. Barcodes were added with the NEBNext Multiplex Oligos for Illumina kit (New England Biolabs, Ipswich, Massachusetts, USA). These four taxa were run on a single lane of Illumina HiSeq 2500. Libraries for *Pteridium* and *Ceratopteris* (average insert size = 300 bp) were prepared using the Illumina TruSeq library preparation kit (Illumina, San Diego, California, USA) and run together on a second lane of Illumina HiSeq 2500. All Illumina sequencing (paired-end reads) was performed at the Duke University Center for Genomic and Computational Biology, which also performed pre-run library quality control. Illumina data for *Pteridium aquilinum* were supplemented with sequences from a Roche 454 GS-FLX Titanium run (Der, 2010).

### Data processing and assembly

Unwanted adapter sequences were removed from our Illumina reads using cutadapt (Del Fabbro et al. 2013; Martin 2011). We used Sickle (Joshi and Fass 2011) to assess read

quality used a sliding window approach. Sections of reads with an average quality score of < Q25 were trimmed, and reads with less than 50 bp remaining were also removed.

Sequence data were sorted by barcode. Quality-trimmed reads for each species were assembled into contigs using CLC Assembly Cell (v4.2.1), specifying a library insert size between 275 and 425 bp (for paired-end reads) and a word size (kmer length) of 31 bp.

### Assembly and analysis of organellar genome components

To assess the structures of the nuclear genomes, we first separated plastid and mitochondrial contigs, which comprised a portion of the assemblies. To identify plastid contigs, we performed BLASTX (Altschul et al. 1997) searches (using an e-value threshold of 1e-10) of the CLC contigs against a custom database of fern proteins extracted from complete sequenced plastomes obtained from GenBank: *Adiantum* (NC\_004766), *Pteridium* (NC\_014348), *Angiopteris* (NC\_008829), *Lygodium* (NC\_024153), *Alsophila* (NC\_012818), and *Cheilanthes* (NC\_014592). The remaining non-plastome contigs were then queried against several plant mitochondrial genomes (there are currently no fern mitochondrial genomes available for such searches): *Zea mays* (NC\_007982), *Pleurozia purpurea* (NC\_013444), *Nicotiana tabacum* (NC\_006581), *Mesostigma viride* (NC\_008240), *Megaceros aenigmaticus* (NC\_012651), *Marchantia polymorpha* (NC\_001660), *Cycas taitungensis* (NC\_010303), *Chara vulgaris* (NC\_005255), *Arabidopsis thaliana* (NC\_001284), *Physcomitrella patens* (NC\_007945), and *Vitis vinifera* (NC\_012119). We searched against both the above complete mitochondrial genome sequences plus a collection of core mitochondrial genes conserved across 27 plant mitochondrial genomes. Details for



extracting these core genes are provided at Digital Commons

(<http://dx.doi.org/10.15142/T39G67>).

The putative plastome contigs from the initial BLASTX search were used to build, iteratively and manually, plastome assemblies in Geneious v7.1 (Biomatters, Auckland, New Zealand). First, we performed manual reference-guided alignments to the most closely related available fern plastome, to orient and order contigs based on the general structure of the reference plastome. The boundaries of the inverted repeat (IR) were manually identified using small cut and paste alignments in Geneious. Next, we used Mauve (Darling et al. 2004) to align these rough plastome assemblies to one or more published reference sequences (listed above). We then transferred preliminary gene annotations from the references to the new assemblies, and manually adjusted reading frames, introns, and putative RNA editing sites. We then used these plastome assemblies as queries in another round of BLASTX searches with the entire CLC contig set as the database, to identify additional possible plastome contigs, or sections of plastid DNA inserted in nuclear or mitochondrial contigs. We then filtered contigs, retaining those with > 95% sequence similarity and > 90% of the contig length with a match to the plastome assembly. This was to exclude possible small portions of plastid DNA that had been inserted into the nuclear genomes. Any additional contigs not incorporated into the plastome assembly were removed from subsequent analyses of the plastome. Annotated plastomes were deposited on GenBank, and contigs identified as putatively containing mitochondrial genes are available at Digital Commons (<http://dx.doi.org/10.15142/T39G67>). We further explored possible organellar contigs by examining the average weighted (by contig length) depth of coverage of putative organellar contigs relative to the entire assembly. Because of the

higher copy number of organellar relative to nuclear genomes in cells, we expect organellar sequences to be detected at higher depth. Thus, low depth organellar sequences can indicate possible regions that have been transferred to the nucleus. All contigs remaining in the original assemblies after the removal of plastid and mitochondrial DNA were considered to be nuclear genomic DNA and were used to estimate repeat content in the nuclear genome.

### Assessing repeat content in the nuclear genome

To assess repeat structure, we analyzed repeat content in these six fern genomes and made comparisons to six seed plants. We downloaded datasets from six phylogenetically representative seed plants from the NCBI SRA (*Amborella trichopoda*, *Vitis vinifera*, *Zea mays*, *Gnetum gnemon*, *Pinus taeda*, and *Taxus baccata*). *Pinus taeda* reads were first trimmed to 100 bp and reads for all 12 taxa were quality-filtered using the fastx toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) to exclude reads that did not contain at least 70% of the bases with quality scores higher than Q20. Quality-filtered paired-end reads were reassociated and interleaved. Three replicate samples of 0.05X coverage of the estimated genome size (approximately 5% of each genome) were used in subsequent analyses to identify highly represented repeat clusters using RepeatExplorer (Novák et al. 2013). Default parameters were used in RepeatExplorer analyses, except that paired reads were used and a domain search was performed using an e-value threshold of 1e-5. The genome proportions represented by 10 sequence-based repeat classes, plus unknown repeats and non-repetitive sequences, were summed for each species based on the most abundant RepeatMasker (<http://www.repeatmasker.org>) hits identified for each cluster.

We used standard linear regressions, performed in R (R Core Team, 2014), to test for correlations between genome size and genome proportions inferred to belong to the 10 repeat classes, for ferns and seed plants separately.

Differences in genomic repeat content between clades were assessed using one-way analysis of variance for a completely randomized design with subsamples. Clade was incorporated in the model as a fixed effects factor, and species within clades as a random effects factor; triplet samples of each species were considered to be subsamples. Each repeat class was analyzed separately, and proportion data were logit-transformed prior to analysis to better meet assumptions of normality and homogeneity of variance. Data calculations were made using the GLIMMIX procedure in SAS/STAT 13.2 in the SAS System for Windows Version 9.4. This statistical model assumes that our choice of species was a random sample from their respective clades; although not strictly true, we chose species to be representative of the breadth within clades and so are comfortable using these analyses as exploratory indicators.

### **Assessing protein-coding content**

We estimated the size of the protein-coding portion of the nuclear genome in two ways: one based on the proportion of assembled contigs with protein hits, and the other based on the proportion of reads that contained BLASTX hits to a protein database.. First, we used BLASTX to query all (previously assembled) non-organelle contigs against a database of annotated protein sequences from 22 plant species (Amborella Genome Project 2013). We then calculated the total length of BLASTX hits longer than 100 bp and divided this by the total assembly length to get estimates of the proportion of each assembly with

protein hits. Because low-copy genes are less likely to be represented in assembled contigs, we also used a second read-based approach. Working with the original read files for each species, we used Bowtie2 (Langmead and Salzberg 2012) to map the reads against the organellar assemblies (plastome assemblies and mitochondrial contigs) as described above. We then removed all reads with organellar hits from the original read file and selected a random sample of the remaining reads to represent an estimated 0.0025X of the genome. We partitioned these random samples of reads into ten equal sets and queried each sequence against a database of annotated protein sequences from 22 plant species (Amborella Genome Project 2013) using BLASTX (e-value < 1e-5). The numbers of reads in each set with hits to known plant proteins were used to calculate mean protein-coding coverage within each set and standard deviation across the ten sets for each species.

## Results

### Genome sizes and assemblies

Our estimates of genome size ranged from 2.45 Gb to 15.65 Gb (Table 2). Illumina sequencing generated between 40,830,366 and 207,771,644 raw reads per species, and between ~4.0 and 19.4 billion bp of quality-filtered data (Table 2). For *Pteridium*, we also included an additional set of 454 data from a previous study (Der 2010), for a total of almost 20 billion bp of data for *Pteridium* (Table 2). GC content in the assemblies ranged from 37.9 to 42.9% (Table 2). Assemblies included from 116,508 to 1.5 million contigs, and were between 42.8 and 620.5 million bp in total length summed across contigs for each species (Table 3). Depth of coverage ranged from 0.39 to 2.06 X, and the proportion of the

nuclear genome represented by our assemblies ranged from ~0.3 % to ~9% (Table 3). All sequence data are available at Digital Commons (<http://dx.doi.org/10.15142/T39G67>).

### Organelar genomes

We assembled and annotated plastomes, consisting of 3–4 contigs for each species, and ranging from 123,674 to 158,508 total bp in length (including both copies of the inverted repeat; Table 4). Plastomes sequences are archived in GenBank (see Table 4 for accession numbers). Each of the six species sampled here appears to have the same gene order as observed in its nearest relatives with previously published plastomes.

We also detected regions with strong sequence similarity to plastid DNA in contigs that did not appear to be part of the plastome assembly. We infer these to be plastome-like genes residing within the nuclear (or perhaps mitochondrial) genomes. For each fern species, we detected 20–241 contigs containing such regions, ranging from a total of about 10 Kb to 55 Kb of plastome-like sequences (Table 4). This amounted to no more than 0.025% of the presumed nuclear genome assembly for each species (Table 4).

Putative mitochondrial contigs had on average approximately 6 times more weighted mean depth of coverage than the entire assembly, and putative plastid contigs had on average 33 times more coverage than entire assembly (Table 5, Fig. 2). These average coverage values enabled us to identify additional mitochondrial contigs with very low relative depth of coverage; these were subsequently transferred to the collection of nuclear contigs. One unusual putative mitochondrial contig of 1380 bp was detected at 2040X in *Plagiogyria*, considerably higher than other contig depths. We suspect that this may be a piece of mitochondrial DNA that was inserted into an active transposon.

We identified 17–36 contigs per species containing putative mitochondrial genes, with the total length of sequences (>100 bp) ranging from approximately 23–415 Kb (Table 4). Most, but not all, known mitochondrial genes were detected in each species (Table 6).

Contigs containing these sequences are available from Digital Commons:

<http://dx.doi.org/10.15142/T39G67>.

### Repeat content analyses

We compared genomic repeat content of our sample of ferns to a similar data set of seed plants. We report here possible differences that can be examined further in the future when high coverage assemblies become available for more taxa. Compared to seed plants, ferns had a higher proportion of their genomes in three main repeat classes (Fig. 3): DNA transposons (mean±S.E. of  $3.2\pm 0.72\%$  in ferns;  $0.83\pm 0.19\%$  in seed plants,  $p(F) = 0.001$ ), long interspersed nuclear elements (henceforth LINES;  $2.2\pm 0.75\%$  in ferns;  $0.49\pm 0.17\%$  in seed plants,  $p(F) = 0.006$ ), and simple repeats ( $15.5\pm 1.5\%$  in ferns;  $1.19\pm 0.89\%$  in seed plants,  $p(F) = 0.007$ ). Satellite DNA (comprised of tandem arrays, including centromeres and telomeres) was on average lower in ferns ( $0.1\pm 0.03\%$ ) compared to seed plants ( $0.8\pm 0.34\%$ ), but both groups in our analyses are low for this class, and the differences were not significant ( $p(F) = 0.214$ ); differences for all other repeat classes were also not significant ( $p(F) > 0.1$ ). Figure 4 illustrates relative proportions of the genome for each class of repeat, the two estimates of protein-coding content (see below), and the remaining non-repetitive component, versus genome size. These plots reveal the similarity across the three samples from each taxon, indicating that our subsampling method is effective.

Standard linear regressions revealed that genome size is not significantly correlated with the size of any class of repetitive element.

### Protein-coding content analyses

Based on sampling reads representing 0.0025X of each genome, we estimated the protein-coding content as 2.85%  $\pm$ 0.03 (*Pteridium*)–6.61%  $\pm$ 0.03 (*Ceratopteris*) of the reads (Table 7, Fig. 4). Estimates obtained by examining all non-organelle contigs in the assemblies, were lower, ranging from 1.11 (*Pteridium* and *Ceratopteris*)–1.90% (*Plagiogyria*) of the assemblies (Table 7, Fig. 4). All supplementary information, including assemblies and sequence reads, is available at Digital Common: <http://dx.doi.org/10.15142/T39G67>.

### Discussion

Here we describe the first investigation into the comparative genome content of ferns. Although low-coverage genome scans are unsuitable for obtaining high-quality genome assemblies or revealing higher-level aspects of genome structure, these data can provide key initial insights into genome content (Rasmussen and Noor 2009; Weitemier et al. 2014). Based on these first analyses of six species, it appears that several aspects of genome content are consistent across ferns, some of which appear to be similar to seed plants, while other aspects may be unique to ferns.

The six species we sampled have genome sizes typical for homosporous leptosporangiate ferns. Estimates of fern genome sizes range from 0.77 pg for *Azolla microphylla* (heterosporous leptosporangiate) to 65.55 pg for *Ophioglossum reticulatum*

and 72.68 pg for *Psilotum nudum* (two eusporangiate ferns; Bennett and Leitch 2001; Obermayer et al. 2002). Our six species have genome sizes of 2.45–15.65 pg, on the lower end of the above range (Table 2). The three species for which we performed flow cytometry experiments, *Plagiogyria formosana*, *Ceratopteris richardii*, and *Dipteris conjugata*, had no previous genome size estimates. Although several surveys of genome size have been made in ferns (Bainard et al. 2011; Bennett and Leitch 2001, 2012; Obermayer et al. 2002), published C-value estimates exist for only 104 fern species (Bennett and Leitch 2012) out of the ca. 9,000 known ferns (Smith et al. 2006; Smith et al. 2008). Additional studies are needed to establish the full range of genome sizes in ferns, which will be particularly useful for determining whether currently unsampled species have small genomes that may be suitable candidates for high-coverage sequencing and assembly. The large genome sizes and high chromosome numbers in ferns, and the concomitant challenges they represent for assembly, have contributed significantly to the difficulty in obtaining a reference genome for ferns (Pryer et al. 2002; Sessa et al. 2014).

Although this is the first genome-wide comparative analysis in ferns, several previous studies have made inferences about fern genome structure or content. For example, Pichersky et al. (1990) reported defective copies of CAB (chlorophyll a/b-binding) genes in the homosporous fern *Polystichum munitum*. The authors hypothesized that the defective genes are the result of gene silencing or loss of gene function in duplicated gene copies. Other nuclear genes with silenced copies have been detected in genetically diploid fern genomes (Gastony 1991; McGrath and Hickok 1999; McGrath et al. 1994). While the presence of these putatively silenced genes appeared consistent with a history of paleopolyploidy in ferns, a high-resolution genetic linkage map generated for



*Ceratopteris richardii* (Nakazato et al. 2006) failed to recover evidence of large-scale synteny that would support extensive ancient polyploidy, though 76% loci were duplicated. Nakazato and colleagues (2006) concluded that small-scale gene duplication was likely the primary mode of duplication in *C. richardii*. Meanwhile, Rabinowicz et al. (2005) examined genome-wide methylation in *C. richardii* and found that, as in other plant groups (Bennetzen et al. 1994; Rabinowicz et al. 2003) gene-rich regions are less methylated than other genomic regions. They also determined that *C. richardii* has roughly the same numbers of genes as angiosperms, but a much lower gene density due to its larger genome size (Rabinowicz et al. 2005).

From calculations based on full, annotated genome sequences, the proportion of a genome that is protein-coding varies considerably among taxa. For example, *Picea abies* (19.6 Gb) is made up of approximately 2.4% protein-coding genes (Nystedt et al., 2014) whereas *Utricularia gibba* (83 Mb), the smallest plant genome sequenced to date, may be composed of as much as 97% protein-coding sequence (Ibarra-Laclette, et al. 2013). In general, estimates of total gene number vary within an order of magnitude: usually between 20,000–40,000 per diploid genome (Rabinowicz et al. 2005; Sterck et al. 2007). Thus, the proportion of a genome that is protein-coding will tend to reflect the inverse of genome size. The only published estimate of protein-coding content in a fern genome is for *Ceratopteris richardii* (11Gb), estimated by Rabinowicz, et al. (2005) to be 0.49% or 6%, depending on the approach used. This is very similar to our estimate for the same species (1.1% or 6.61%). Rabinowicz (2005) used fewer than 600 reads at an average length of >600 bp. Therefore, that their estimates of protein coding content are similar to ours provides us with some confidence in our estimation approaches. Our estimates based on

the proportion of reads were three to six times greater for all species than those based on the proportion of all assembled contigs containing protein-coding sequences (Fig. 4). With our low coverage, we had expected that the assemblies might underestimate the protein-coding component. This could occur if the assemblies were biased towards repetitive parts of the genome, with the excluded, unassembled regions more likely to be single-copy. Despite the difference in results from the two methods, all of our estimates fall towards the low end of protein content measured in (non-fern) vascular plants, ranging from 1.11–1.90% or 3.07–6.61%. These low estimates probably reflect the relatively large genome sizes of the species we sampled. We also suspect that both our estimates could be low because of a lack of reference proteins from ferns. Although genome size and protein-coding gene density may be negatively correlated across plants in general, no such relationship exists among the six fern species that we sampled (Fig. 4).

We detected several differences in the repetitive elements of the genomes of ferns compared to seed plants. As a group, our six fern samples had higher proportions of DNA transposons, LINES, and simple repeats, and lower proportions of satellite DNA than seed plants (Figs. 3 and 4). It may be premature to infer that these represent real, biologically significant differences between clades, but given the unusual characteristics of fern genomes it seems reasonable to expect that such differences may exist. Future work should focus on identifying specific subclasses of repeat elements in a broader taxon sampling, to explore patterns of genome repeat structure across land plants in more detail.

Although overall genome coverage was low, coverage for organellar reads was, as expected, much higher (Table 5, Fig. 2). Thus, we were able to approach full assembly of fern plastomes and detect the majority of known plant mitochondrial genes. Resolution of

assemblies might be improved with the addition of longer-read sequence data (such as PacBio) or higher coverage with shorter Illumina reads, or both. Better-resolved assemblies should help to distinguish true plastid and mitochondrial genes from those that have been transferred to a different compartment (Matsuo et al. 2005). However, our low coverage data should be ample for studies requiring just the gene sequences (such as phylogenetic analyses). Currently, even reference-guided assemblies require the manual step of establishing plastome inverted repeat boundaries. Given the relative conservation of these positions, it should be possible to automate this assembly step. The plastomes that we assembled here were all from previously sampled major clades of ferns, and we detected no unusual gene order; all plastomes appeared to have structures consistent with other members of the same clades.

Detailed information for mitochondrial genomes is limited from many groups of plants, including ferns. We are aware of three studies that have examined fern mitochondrial genomes. Palmer et al. (1992) isolated restriction fragments of two fern mitochondrial genomes: those of *Equisetum arvense* (>200 Kb) and *Onoclea sensibilis* (~300 Kb). The authors also detected repeat structure and several known plant mitochondrial genes. In another study, a large fosmid clone (290 kb) of the mitochondrial genome of *Gleichenia dicarpa* was found to contain fragments of foreign DNA including transposons, retrotransposons, and transposed introns (Grewe 2011). Several fragments of mitochondrial DNA have also been isolated from *Adiantum nidis* (Panarese et al. 2008). A fragment of almost 21 kb contained inverted repeats and several genes that appear to have been transferred from the plastome. Our analysis of fern mitochondrial DNA identified a large proportion of known plant mitochondrial genes (Table 6). However, coverage was not

sufficient to assemble large fragments containing more than about three genes. We also cannot be sure if undetected genes are absent from mitochondrial genomes or the result of low coverage. Nevertheless, the sequences of the fragments detected provide an excellent starting point for further studies of fern mitochondrial genes and studies of horizontal gene transfer in plants.

Across sequenced plant genomes, there is a positive correlation between genome size and the proportion of a genome that is made up of repeats (Michael 2014). This is because larger genomes tend to be larger because of the presence of expanding repeat elements. However, we do not observe such a pattern in ferns. One possibility is that the range of genome sizes here (2.45–15.65 pg) is too small, compared to the full range for ferns (0.77 - 65.55 pg), to detect a relationship, and studies of more fern species may reveal a positive relationship. It is also possible that ferns in general, or the species we chose, do not have many recently expanding repetitive elements. A third possibility is that ferns are indeed different from other plants when it comes to patterns of genome downsizing (Barker and Wolf 2010; Leitch and Leitch 2012). In ferns, there is a positive correlation between genome size and chromosome number (Nakazato et al. 2008; Bainard et al. 2011). In most other organisms this relationship has been found to be weak (Leitch and Bennett 2004) or even negative (Vinogradov 2001). This lack of a relationship has been explained by rapid (in evolutionary terms) genome downsizing following polyploidy, so that tetraploids (except very recent ones) have genome sizes less than double that of their diploid ancestors (Leitch and Bennett 2004). This downsizing presumably involves the loss of extra genomic material that is not needed. In contrast, the genomic patterns we observe in ferns suggest that they lack the mechanisms for jettisoning excessive and redundant

genomic regions. If this pattern holds with examination of other species, then ferns may provide useful control cases for studies of the underlying mechanisms of genome downsizing in other lineages. High quality assembly and annotation of a fern genome will go a long way to assist researchers in the study of plant genome dynamics.

Our low-coverage genome scans enabled us to make some general statements about the relative content of homosporous, leptosporangiate fern genomes. However, these plants differ in many ways from the heterosporous land plants that have been examined to date, and higher coverage assemblies are critical for detailed comparative analyses of fern and land plant genome structure. Such studies are essential for addressing questions about the evolution of land plant genomes. Furthermore, ferns are the sister group to seed plants (Pryer et al. 2001) and evolutionarily comparative statements about seed plant genomes would benefit from comparisons to a fern genome. Currently, researchers are assembling the first fern genome, that of the heterosporous fern *Azolla* (Li and Pryer 2014). Also underway is higher coverage assembly of the model homosporous fern, *Ceratopteris richardii* (Sessa et al. 2014; Marchant et al unpubl.). Meanwhile, here we have presented a first exploration into comparative genome content of ferns.

### **Acknowledgements**

This research was supported by National Science Foundation Doctoral Dissertation Improvement Grant sDEB-1407158 (to K.M.P and F.-W.L.) and DEB-1110767 (to K.M.P. and C.J.R.), and a National Science Foundation Graduate Research Fellowship to F.-W.L. We also thank Li-Yaung Kuo for assisting with flow cytometry, Susan Durham for help with statistical analyses, and Carol Rowe for comments on the manuscript.

## Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Amborella Genome Project 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342: 1241089.
- Bainard JD, Henry TA, Bainard LD, Newmaster SG 2011. DNA content variation in monilophytes and lycophytes: large genomes that are not endopolyploid. *Chromosome Res* 19: 763-775.
- Barker MS, Wolf PG 2010. Unfurling Fern Biology in the Genomics Age. *Bioscience* 60: 177-185.
- Bennett MD, Leitch IJ 2001. Nuclear DNA amounts in pteridophytes. *Ann Bot* 87: 335-345.
- Pteridophyte DNA C-values database (Release 5.0, Dec. 2012). 2012. Available from: <http://data.kew.org/cvalues/>
- Bennetzen JL, Schrick K, Springer PS, Brown WE, Sanmiguel P 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37: 565-576.
- Cantino PD, et al. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon*: 822-846.
- Darling AC, Mau B, Blattner FR, Perna NT 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14: 1394-1403.
- Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM 2013. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE* 8: e85024.

- Der JP 2010. Genomic perspectives on evolution in bracken fern. PhD. Dissertation. Utah State University, Logan, UT, USA.
- Doležel J, et al. 1998. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann Bot* 82 (suppl 1): 17-26.
- Ebihara A, et al. 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. *Am J Bot* 92: 1535-1547.
- Gastony GJ 1991. Gene silencing in a polyploid homosporous fern: Paleopolyploidy revisited. *Proc Nat Acad Sci USA* 88: 1602-1605.
- Grewe F 2011. Die mitochondriale DNA basaler Tracheophyten: Molekulare Evolution komplexer Genomstrukturen. [Bonn: Rheinischen Friedrich-Wilhelms-Universität.
- Hanson L, Boyd A, Johnson MAT, Bennett MD 2005. First nuclear DNA C-values for 18 eudicot families. *Ann Bot* 96: 1315-1320.
- Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. [github.com/najoshi/sickle](https://github.com/najoshi/sickle).
- Klekowski EJJ, Baker HG 1966. Evolutionary significance of polyploidy in the Pteridophyta. *Science* 153: 305-307.
- Langmead B, Salzberg SL 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods* 9: 357-359.
- Li FW, Pryer KM 2014. Crowdfunding the *Azolla* fern genome project: a grassroots approach. *GigaScience* 3: 16.
- Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82: 651-663.

- Leitch, AR, IJ Leitch. 2012. Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytol* 194:629–646.
- Martin M 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.jour* 17: 10-12.
- 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.
- McGrath JM, Hickok LG 1999. Multiple ribosomal RNA gene loci in the genome of the homosporous fern *Ceratopteris richardii*. *Can J Bot* 77: 1199-1202.
- McGrath JM, Hickok LG, Pichersky E 1994. Assessment of gene copy number in the homosporous ferns *Ceratopteris thalictroides* and *C. richardii* (Parkeriaceae) by restriction fragment length polymorphisms. *Plant Syst Evol* 189: 203-210.
- Michael TP 2014. Plant genome size variation: bloating and purging DNA. *Brief. Funct. Genomics* 13: 308-317.
- Michael TP, VanBuren R 2015. Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol* 24: 71-81.
- Murray BG 1985. Karyotypes and nuclear DNA amounts in *Polypodium* L. (Polypodiaceae). *Bot J Linn Soc* 90: 209-216.
- Nakazato T, Barker MS, Rieseberg LH, Gastony GJ. 2008. Evolution of the nuclear genome of ferns and lycophytes. In: Ranker TA, Haufler CH, editors. *Biology and Evolution of Ferns and Lycophytes*. Cambridge, UK: Cambridge University Press.



- Nakazato T, Jung MK, Housworth EA, Rieseberg LH, Gastony GJ 2006. Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. *Genetics* 173: 1585–1597.
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29: 792–793.
- Obermayer R, Leitch IJ, Hanson L, Bennett MD 2002. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann Bot* 90: 209-217.
- Palmer JD, Soltis D, Soltis P 1992. Large size and complex structure of mitochondrial DNA in two nonflowering land plants. *Curr Genet* 21: 125-129.
- Panarese S, Rainaldi G, De Benedetto C, Gallerani R 2008. Sequencing of a segment of a monilophyte species mitochondrial genome reveals features highly similar to those of seed plant mtDNAs. *Open Plant Sci J* 2: 15-20.
- Pichersky E, Soltis D, Soltis P 1990. Defective chlorophyll a/b-binding protein genes in the genome of a homosporous fern. *Proc Nat Acad Sci USA* 87: 195-199.
- Pryer KM, et al. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409: 618-622.
- 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, Schneider H, Zimmer EA, Banks JA 2002. Deciding among green plants for whole genome studies. *Trends Plant Sci* 7: 550-554.

- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Rabinowicz PD, et al. 2005. Differential methylation of genes and repeats in land plants. *Genome Res* 15: 1431-1440.
- Rabinowicz PD, et al. 2003. Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res* 13: 2658-2664.
- Rasmussen D, Noor M 2009. What can you do with 0.1x genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC Genomics* 10: 382.
- Schneider H, et al. 2015. Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns. *New Phytol* doi: 10.1111/nph.13330.
- Schneider H, et al. 2004. Ferns Diversified in the shadow of angiosperms. *Nature* 428: 553-557.
- Sessa E, et al. 2014. Between two fern genomes. *GigaScience* 3: 15.
- Smith AR, et al. 2006. A classification for extant ferns. *Taxon* 55: 705-731.
- Smith AR, et al. 2008. Fern Classification. In: Ranker TA, Haufler CH, editors. *Biology and Evolution of Ferns and Lycophytes*: Cambridge University Press.
- Sterck L, Rombauts S, Vandepoele K, Rouze P, Van de Peer Y 2007. How many genes are there in plants (. . . and why are they there)? *Curr Opin Plant Biol* 10: 199-203.
- Vinogradov AE 2001. Mirrored genome size distributions in monocot and dicot plants. *Acta Biotheor.* 49: 43-51.

- Weitemier K, et al. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl Plant Sci* 2: apps.1400042.
- Wood TE, Takebayashi N, Barker MS, Greenspoon PB, Rieseberg LH 2009. The frequency of polyploid speciation in vascular plants. *Proc Nat Acad Sci USA* In Press.
- Yokoya K, Roberts AV, Mottley J, Lewis R, Brandham PE 2000. Nuclear DNA amounts in roses. *Ann Bot* 85: 557–561.

**Figure legends:**

Figure 1. Phylogeny of ferns summarized from Pryer et al. (2004). Numbers of sequenced nuclear genomes are indicated for the lineages that have them. Lineages in pink are the eusporangiate ferns; the leptosporangiate fern clade is in green. Taxa in this study are given in parentheses. Photos of representative ferns are included: A) *Ceratopteris richardii*, Pteridaceae, B) *Salvinia* sp., Salviniiales (heterosporous water ferns), C) *Cystopteris protrusa*, Cystopteridaceae, D) *Ophioglossum* sp., Ophioglossales (rattlesnake ferns), E) *Equisetum* sp. (horsetails).

Figure 2. Depth of coverage for primary CLC assemblies plotted as a function of contig length. Axes are log scale and contour lines (blue) show the density of over-plotted contigs. Chloroplast contigs are shown in green and mitochondrial contigs are shown in orange.

Figure 3. Genome proportions represented by 10 sequence-based repeat classes, plus unknown repeats and non-repetitive sequences, in six fern and six seed plant taxa, with three samples per taxon.

Figure 4. Scatter plots showing the relationship between proportion of different classes of genomic elements and genome size for ferns and seed plants

Table 1. Locality and voucher information for the six ferns sampled.

Species	Collection locality	Voucher (herbarium)
<i>Dipteris conjugata</i> (Kaulf.) Reinw.	Pahang, Malaysia	Schuettpelz 770 (DUKE)
<i>Cystopteris protrusa</i> (Weatherby) Blasdell	Ashe County, North Carolina, USA	Rothfels 4168 (DUKE)
<i>Plagiogyria formosana</i> Nakai	Nantou County, Taiwan	Schuettpelz 1083A (DUKE)
<i>Ceratopteris richardii</i> Brongn.	Cuba (Accession: Hnn)	dbmarchant01 (FLAS)
<i>Pteridium aquilinum</i> L. Kuhn	Manchester, UK	E. Sheffield S48 (UTC)
<i>Polypodium glycyrrhiza</i> D.C. Eaton	Squamish-Lillooet, British Columbia, Canada	Rothfels 4086 (DUKE)

Table 2. Amount of sequence data (bp and reads), GC content, and estimates of genome size for six ferns. Clean data refers to reads processed by removing adapters and trimming low quality regions

Taxon	Number of raw reads	bp of clean data	GC content	Number of Ns in assembly	Genome size (Gb) from flow cytometry
<i>Cystopteris</i>	47,005,020	4,580,023,307	0.42	645,022	4.23
<i>Dipteris</i>	51,232,072	5,023,794,762	0.42	1,825,278	2.45
<i>Plagiogyria</i>	58,488,796	5,717,123,738	0.43	1,376,069	14.81
<i>Polypodium</i>	40,830,366	4,000,482,565	0.43	503,779	10.02
<i>Ceratopteris</i>	204,001,778	19,445,093,728	0.38	8,287,893	11.25
<i>Pteridium</i> 454		216,194,085			
<i>Pteridium</i> Illumina	207,771,644	19,437,952,758			
<i>Pteridium</i> Both	208,482,822	19,654,146,843	0.39	24,208,821	15.65
Totals	610,040,854	58,420,664,943			

Table 3. Information on genome assemblies (in bp) and genome coverage. Asterisk indicates that these are combined Illumina and 454 assemblies for *Pteridium*.

Taxon	Contigs in assembly	Total length of assembly	N50	Assembly size minus organellar	Proportion of genome covered by assembly	Depth (x) of coverage
<i>Cystopteris</i>	125,022	42,821,163	326	42,691,902	0.01001	1.082
<i>Dipteris</i>	628,061	232,459,611	366	232,459,008	0.09507	2.055
<i>Plagiogyria</i>	116,508	46,007,615	365	46,000,412	0.00311	0.386
<i>Polypodium</i>	162,707	53,369,105	313	53,369,105	0.00532	0.399
<i>Ceratopteris</i>	944,561	350,037,872	365	349,866,779	0.03111	1.729
<i>Pteridium</i> *	1,497,826	620,490,875	460081	620,488,482	0.06344	1.256

Table 4. Characteristics of organellar genome sequences and assemblies in six ferns. Asterisk indicates that these are combined Illumina and 454 assemblies for *Pteridium*. mt = mitochondrial; nc = nuclear; pl = plastid

Taxon	Plastome length (bp)	Number of pl contigs	GenBank accession	Number of contigs with putative mt sequences	Length of mt-like sequences > 100 bp detected	Number of nc contigs containing pl-like sequences	Total length of pl-like sequences (bp)	Proportion of nuclear assembly with pl-like seq
<i>Cystopteris</i>	158,508	3	KP136830	19	27,868	45	10,166	0.000238
<i>Dipteris</i>	123,674	3	KP136829	36	413,081	29	1,852	0.000077
<i>Plagiogyria</i>	150,106	4	KP136831	33	387,300	35	1,105	0.000242
<i>Polypodium</i>	152,982	4	KP136832	34	339,724	36	1,324	0.000194
<i>Ceratopteris</i>	126,823	3	KM052729	22	22,776	260	5,243	0.000158
<i>Pteridium</i> *	152,362	3	NC_014348	17	27,463	166	36,892	0.000059



Table 5. Weighed mean depth of coverage for all, plastid (pl), and mitochondrial (mt) contigs normalized by contig length. Chloroplast coverage exceeds mitochondrial coverage by an order of magnitude and mitochondrial coverage exceeds the overall mean coverage for the assembly in all cases.

	<i>Ceratopteris</i>	<i>Cystopteris</i>	<i>Dipteris</i>	<i>Plagiogyria</i>	<i>Polypodium</i>	<i>Pteridium</i>
Weighed mean coverage	48.51	76.98	16.33	78.36	48.58	26.42
pl weighted mean coverage	3112.58	1641.57	427.09	365.90	1248.29	2878.43
mt weighted mean coverage	348.57	393.61	144.92	93.81	171.25	608.08

Table 6 – list of putative mitochondrial genes detected in six fern species

<i>Polypodium</i>	<i>Cystopteris</i>	<i>Dipteris</i>	<i>Plagiogyria</i>	<i>Ceratopteris</i>	<i>Pteridium</i>
atp1	atp1	atp1	atp1	atp1	atp1
atp4		atp4	atp4		
		atp6	atp6	atp6	
atp8		atp8	atp8		
				atp9	atp9
cob	cob	cob	cob	cob	cob
cox1	cox1	cox1	cox1	cox1	cox1
cox2	cox2		cox2	cox2	cox2
cox3	cox3			cox3	cox3
		ccmB			
matR		matR	matR	matR	matR
		mttB			
nad1		nad1	nad1	nad1	
nad2	nad2	nad2	nad2	nad2	nad2
	nad3	nad3	nad3	nad3	nad3
nad4	nad4	nad4		nad4	nad4
nad4L	nad4L	nad4L	nad4L	nad4L	nad4L
nad5	nad5	nad5	nad5	nad5	nad5
nad6	nad6	nad6	nad6	nad6	
nad7	nad7	nad7	nad7	nad7	nad7
nad9	nad9	nad9	nad9	nad9	nad9
					rpl5
rpl16	rpl16	rpl16	rpl16	rpl16	rpl16
	rps13				
		rps12	rps12		
		rps2B	rps2B		

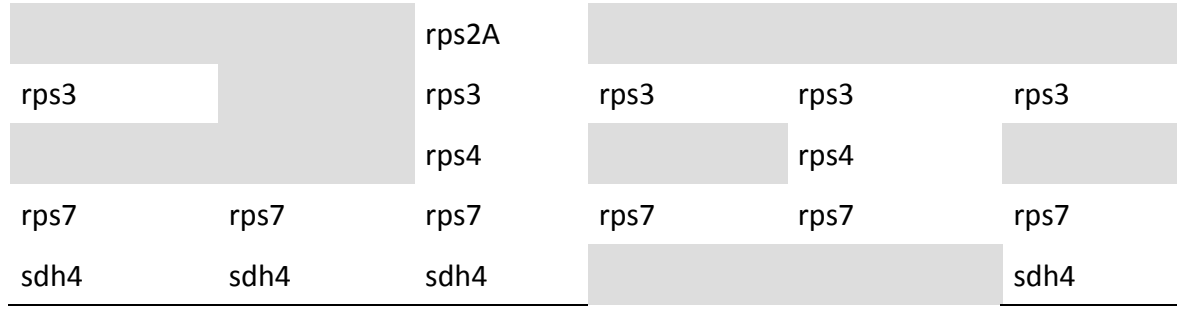
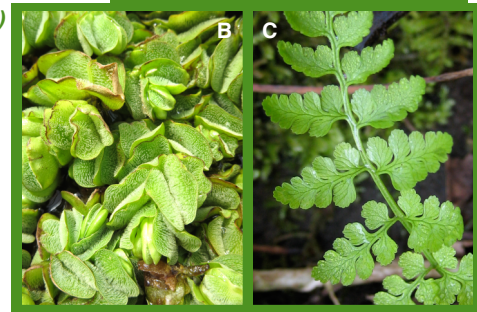
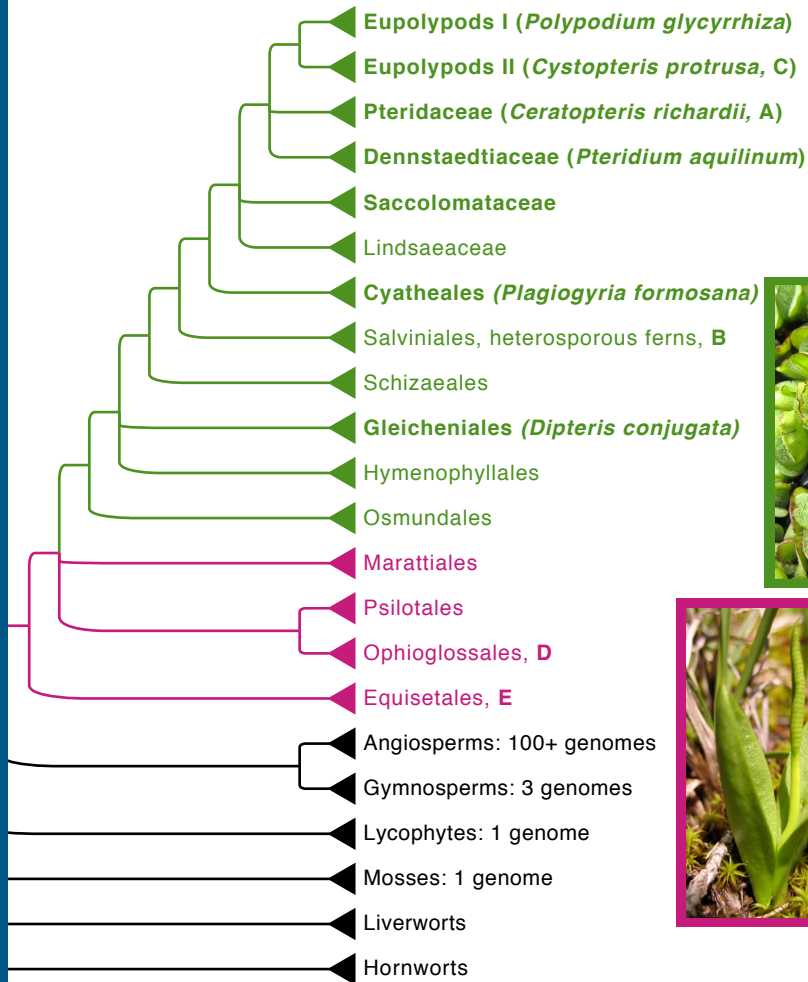


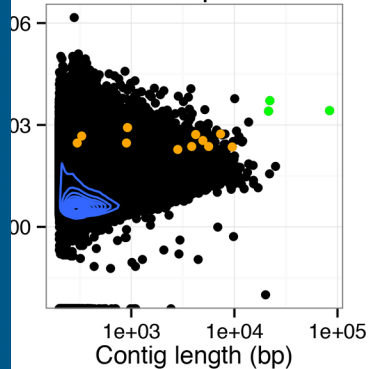
Table 7. Estimated percent protein-coding content

Species	Method 1: Mean % protein content $\pm$ Standard error of the mean	Method 2: Percent of assemblies with blast hits > 100 bp
<i>Ceratopteris</i>	6.61 $\pm$ 0.03	1.11
<i>Cystopteris</i>	5.22 $\pm$ 0.07	1.78
<i>Dipteris</i>	4.82 $\pm$ 0.06	1.12
<i>Plagiogyria</i>	3.07 $\pm$ 0.02	1.90
<i>Polypodium</i>	4.01 $\pm$ 0.03	1.40
<i>Pteridium</i>	2.85 $\pm$ 0.03	1.11

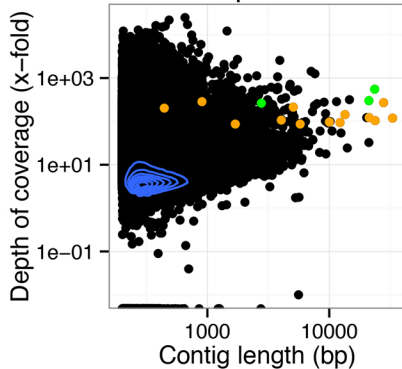




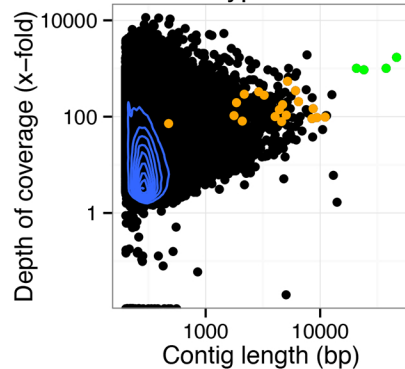
Ceratopteris



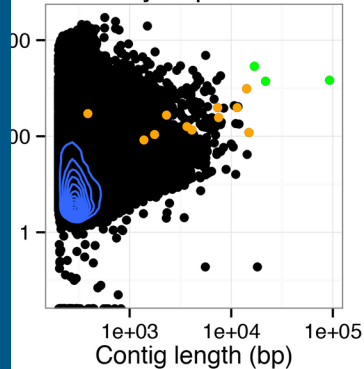
Dipteris



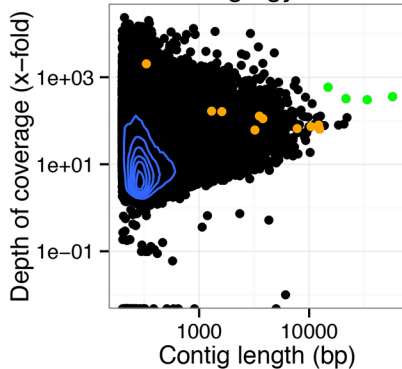
Polypodium



Cystopteris



Plagiogyria



Pteridium

