Delving into the C-Fern Genome and Euphyllophyte Evolution

D Blaine Marchant[1,2], Emily B Sessa[1], W Brad Barbazuk[1], Paul G Wolf[3], Joshua P Der[4,5], Matias Kirst[6], Pamela S Soltis[2], Douglas E Soltis[1,2]

[1]Department of Biology and Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[2]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[3]Ecology Center and Department of Biology, Utah State University, Logan, UT 84322, USA
[4]Department of Biology and Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA
[5]Department of Biological Sciences, California State University Fullerton, Fullerton, CA 92831, USA
[6]School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611, USA
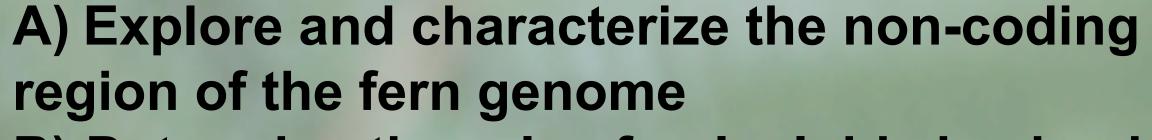
Project Objectives

We will use *Ceratopteris richardii*, or C-Fern, a fast-growing, homosporous fern with a genome size of 11.26 Gb and chromosome count of $n = 39$, used in K-12 and undergraduate classes around the globe, as the "fern model organism."[1] By sequencing the genome of *C. richardii,* this project addresses three aims:

A) **Explore and characterize the non-coding region of the fern genome**
B) **Determine the role of polyploidy in shaping the genomic history of ferns**
C) **Compare the fern genome with available sequenced seed plant genomes to better understand euphyllophyte genomic evolution**

Why Sequence Ferns?

Investigating the genomic characteristics and complexities of ferns is critical for understanding the evolutionary genomics of land plants as a whole. As sister to the seed plants, ferns are the required outgroup for comparative studies among the more economically significant gymnosperms and angiosperms. Investigation into the fern genome can provide insight into the genetic progenitors of traits such as wood and flower development, phenology, etc.

Singularly, ferns pose a number of biological paradoxes. The average fern genome has a haploid size of 9.3 Gb[2], compared to the minute 0.16 Gb genome of the model flowering plant, *Arabidopsis thaliana*. Ferns are notorious for numerous chromosomes, on average containing three times more chromosomes than the average angiosperm[3]. Ferns are one of the few lineages comprising both homosporous and heterosporous species, as well as the most recently diverged lineage to have an independent gametophyte. While these life history traits correlate with chromosome count, the causes are unclear.
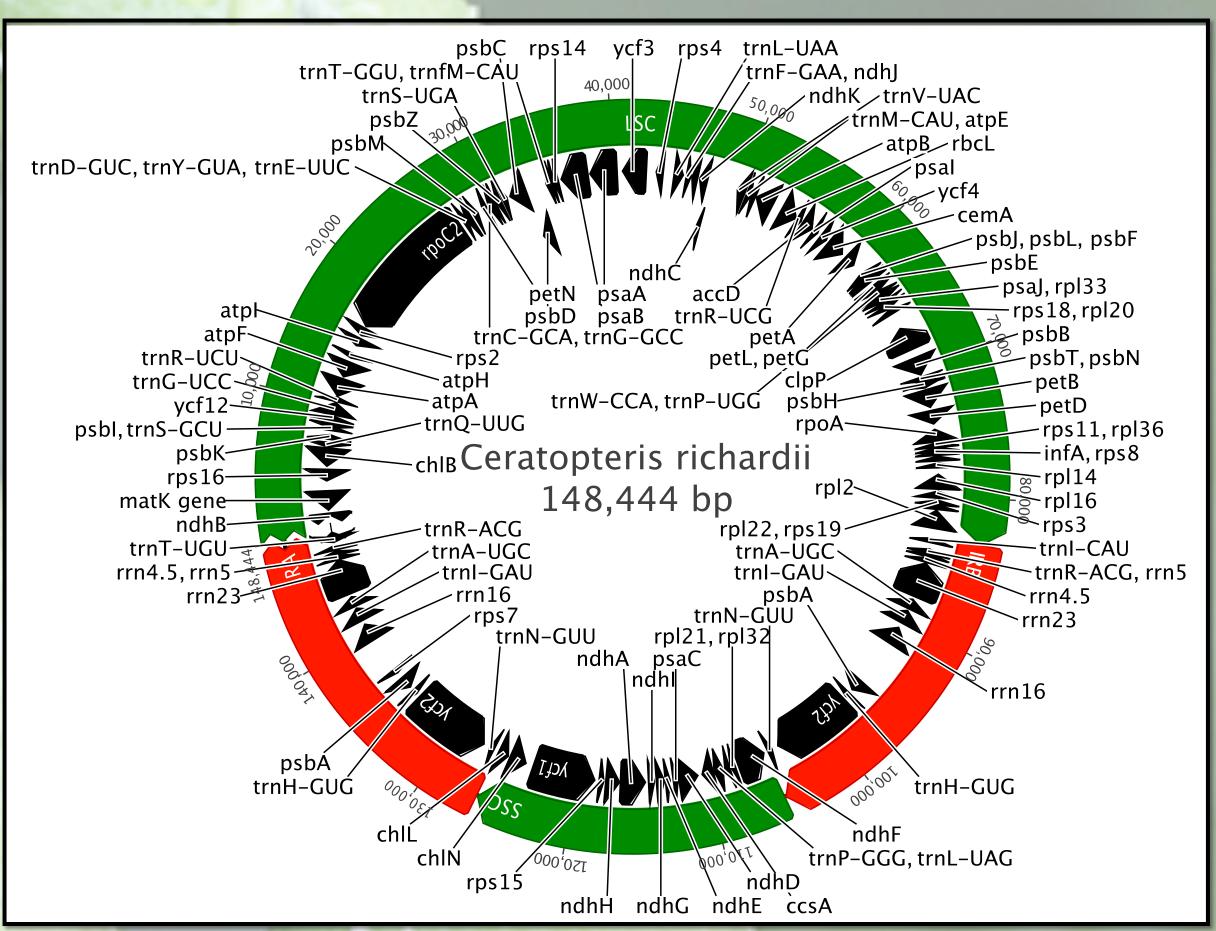
Although recent advances in genomics and bioinformatics have allowed insight into the genomes of more than 30 different plant species, ferns have remained conspicuously untouched in terms of genomic resource development. Implementing an amalgamation of next generation sequencing technology, this project will provide crucial insight into the genome composition of ferns and the evolutionary genomics of euphyllophytes.

C. richardii Plastome

- Plastome Size: 148,444 bp
- Inverted repeats (IR): 22,020 bp each
- Large Single-Copy (LSC) region: 83,178 bp
- Small Single-Copy (SSC): 21,226 bp
- 118 genes assigned when IRs considered once
- 77 encode proteins, 33 encode tRNAs, 8 encode rRNAs
- 45 putative RNA editing sites in 31 stop codons, 13 non-ATG possible start codons, and 1 "repaired" stop codon
- Plastome gene content and order are identical to all other sequenced polypod ferns

Genbank Accession: KM052729

Preliminary Genome Assembly Statistics

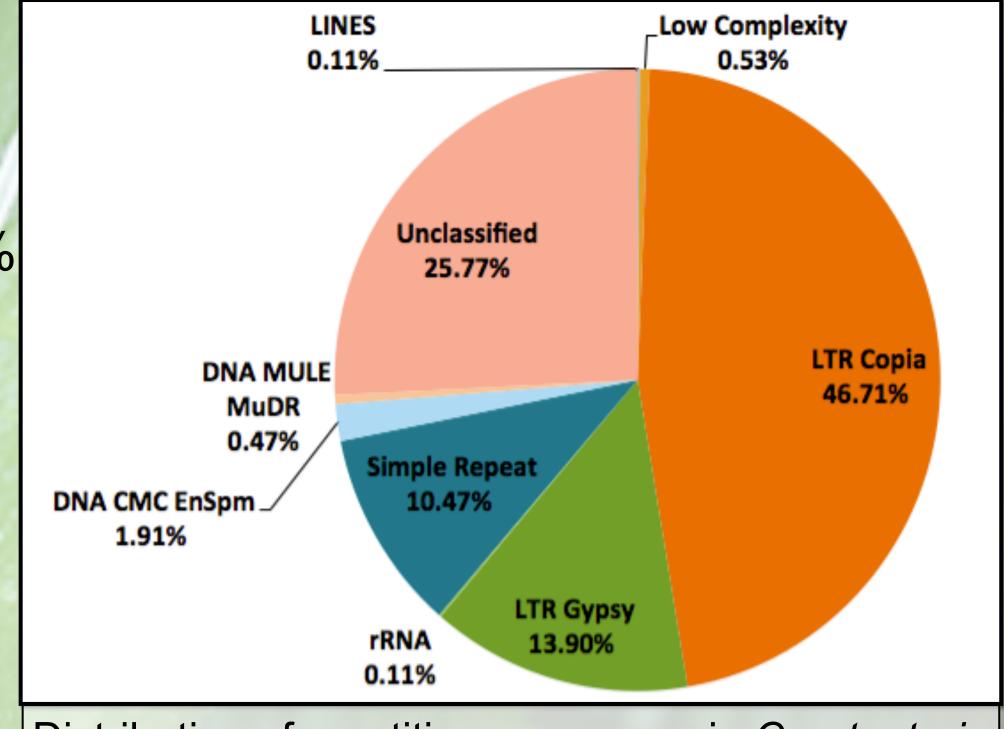| 1N Genome Size (Gb)	Total Clean Data (Gbp)	Number of contigs in total assembly	Assembly Size (Mbp)	N50 Length (bp)	Proportion of assembly aligned with BLAST	% GC	Proportion of genome assembled
11.26	19.45 (1.81X)	944,561	350.03	365	0.111	0.38	0.033

Genome Repeat Composition

- Estimated from 0.1x coverage of 100 bp paired-end Illumina reads using RepeatExplorer[4]
- 37.6% of the genome consists of repeats, 62.4% non-repetitive low copy sequences
- Long terminal repeats (LTR) comprise 22.8% of the genome

Distribution of repetitive sequences in *Ceratopteris* based on repeat type

Future Directions

1. **Gametophyte (1n) and sporophyte (2n) transcriptome sequencing**
2. **Shallow (~15x) gametophyte genome sequencing**
3. **Targeted capture of areas of interest (i.e. pseudogenes, transposable elements, paralogs)**

References: [1]Hickok LG, Warne TR, Slocum MK. 1987. *Ceratopteris richardii*: applications for experimental plant biology. *American Journal of Botany* 74: 1304-1316; [2]Bennett MD, Leitch IJ. 2012. Pteridophyte DNA C-values database; [3]Klekowski EJ, Baker HG. 1966. Evolutionary significance of polyploidy in the Pteridophyta. *Science* 153: 305-307; [4]Novak P et al. 2013. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. *Bioinformatics*