

PAGS Project Proposal

Joel Nitta

2020-06-24

An open, continuously-updated Fern Tree of Life (FTOL)

Ferns (ca. 12,000 spp.) are the second most diverse clade of vascular plants after seed plants (ca. 260,000 spp.), and are a useful study system for understanding processes of biogeography (Tryon 1986; Kato 1993), community ecology (e.g., Hennequin et al. 2014; Lehtonen et al. 2015), and speciation (e.g., Kao et al. 2020). Ferns have received relatively intense focus from molecular systematists, resulting in the publication of molecular trees for nearly all major clades and DNA sequences for ca. 1/3 of all fern species deposited in GenBank. Global trees including all ferns have been published periodically (Pryer et al. 2004; Lehtonen 2011; Schuettpelez and Pryer 2007; Testo and Sundue 2016), but as molecular systematics research continues at a rapid pace, such trees quickly become outdated. A continuously-updated fern tree of life (FTOL) built from sequence data would be useful for anybody studying this important group of plants at scales from smaller (i.e., genus or lower) clades to the entire tree.

There exist several projects to automatically or semi-automatically generate trees for any arbitrary part of the tree of life (Antonelli et al. 2016), a tree for all plants (Eiserhardt et al. 2018), or even the entire tree of life at once (Hinchliff et al. 2015). The question arises then, why the need for a fern tree? First, such approaches must take some shortcuts because of their broad scope. For example, Hinchliff et al. (2015) use a super-tree approach and therefore their tree lacks branch lengths (which are frequently needed for comparative analyses) and cannot integrate new sequence data. Such broad approaches also do not allow for custom tailoring of taxonomy or input data. By focusing methods and datasets specifically on ferns, it should be possible to generate a higher quality end-product (tree) that can be used “as-is” by biologists studying these organisms.

Furthermore, there is much to be gained from integrating FTOL with the fern systematics community that would not be possible with a “tree of all life” or “tree of all plants”. Recently, a genus-level taxonomy of ferns and lycophytes was established following a open, community-driven model (Pteridophyte Phylogeny Group I 2016). Thanks to this community-first approach, the taxonomy (PPG I) has been widely accepted and used (cited 475 times on Google Scholar as of writing). However, there were clearly problematic (non-monophyletic) genera at the time PPG I was published, and many new taxonomic changes have been (and will continue to be) made since (e.g., Almeida et al. 2017; Shang et al. 2018; Zhang et al. 2020). It is anticipated that “PPG II” will be an online, open resource that can be updated as necessary (Schuettpelez, pers. comm.). The FTOL could be directly integrated with PPG II so that taxonomic decisions reflecting phylogeny can be made based on community consensus using the most recently available data.

Here, I propose constructing an automatically-updated pipeline that will pull whole plastomes and selected plastid genes from GenBank to infer a global fern tree of life (FTOL). In addition to tree files, a web portal will be provided that visualizes the tree, provides links to other data sources like the Global Biodiversity Information Facility (GBIF) and the Pteridophytes Collections Consortium (<http://www.pteridoportal.org/portal/>), and allows users to query their own DNA sequences and visualize them on the tree.

The FTOL will have several impacts for the field of fern systematics and evolution: 1) It will always provide the most up-to-date snapshot of our collective understanding of fern phylogeny; 2) It will allow for continuous assessment of taxonomy, and indicate what parts of the tree are in need of taxonomic revision; 3) It will be an important source of data for any phylogenetic comparative study of ferns.

Methods

The workflow (summarized in Fig. 1) leverages the fact that only a handful of genes have been intensively sequenced in ferns, and account for the vast majority of sequences on GenBank: *rbcL*, *atpA*, *atpB*, and *rps4*. These four genes (ca. 5 kb total) provide maximum phylogenetic breadth, but lack the resolving power necessary to infer a robustly supported tree across all ferns. Therefore, I also sample all available fern plastomes for single-copy genes (ca. 70 genes for 100 species), and combine the two data sources to create a mostly empty supermatrix. The plastome data provide a well-supported backbone, which is filled-in by the much more phylogenetically diverse Sanger sequences.

The workflow is implemented in R (R Core Team 2019) with the Drake package (Landau 2018). After downloading plastid genes and genomes, taxonomic name resolution (detection of synonyms) is carried out using the Catalog of Life (COL) (<https://www.catalogueoflife.org/>) as a taxonomic standard. Next, rogue taxa are detected by an all-by-all BLAST search (Camacho et al. 2009); any sequences whose top three hits match a different family are excluded. The current workflow only includes one specimen per species, so sequences are selected prioritizing those that come from the same specimen with maximum combined length. Genes are then aligned individually using MAFFT (Katoh et al. 2002), then concatenated into a supermatrix (only sequences originating from the same specimen are concatenated). The tree is inferred with IQTREE (Nguyen et al. 2015), and dated with treePL (Smith and O’Meara 2012) using 28 fossil calibration points.

In the future, the tree will be available in several versions: one tip per species, multiple tips per species, and with the option to include unsampled species based on taxonomy.

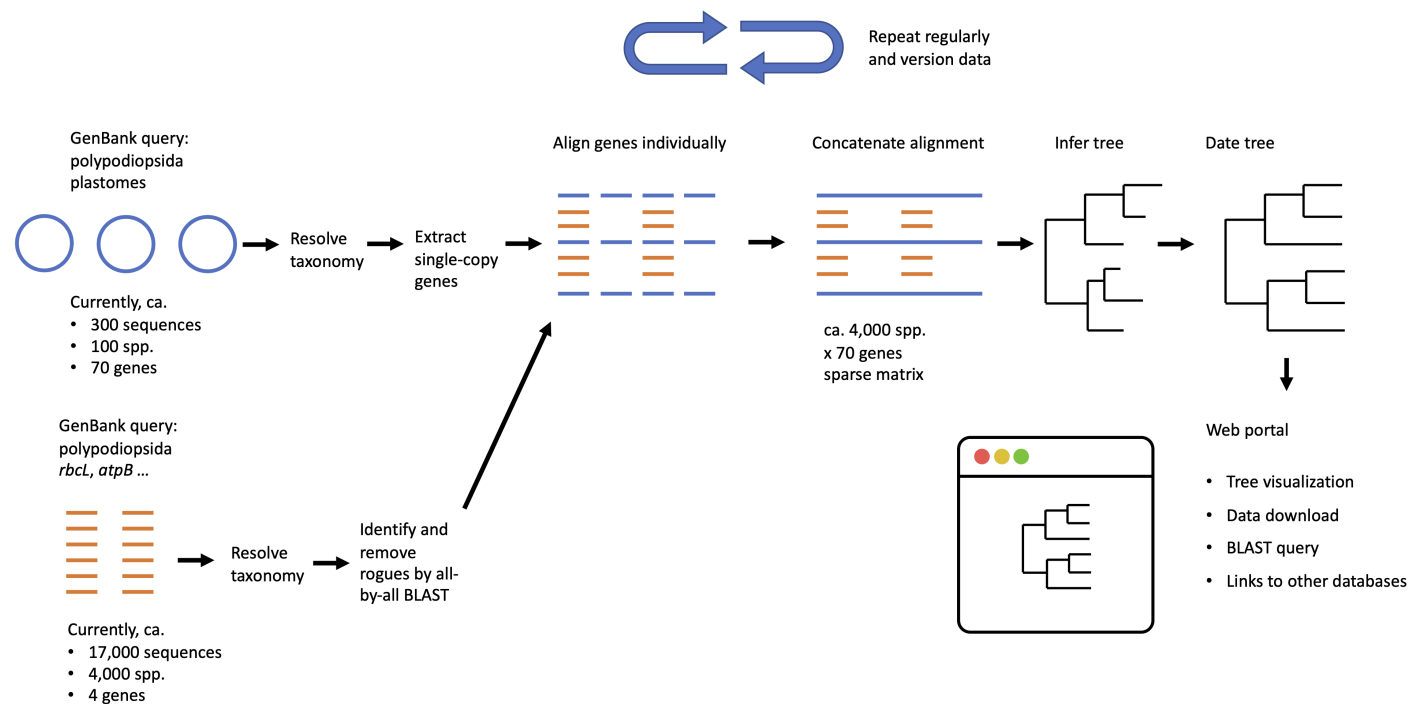


Figure 1: Workflow for generating a continuously-updated global fern tree.

Proof-of-concept

I used the workflow described above to generate a 4,413-tip tree of ferns (one tip per species). Overall coverage (percent of species in the tree out of all accepted species) was 35%. Coverage by major clade (order or suborder; taxonomy follows Pteridophyte Phylogeny Group I (2016)) varied from 18% (Saccolomatineae) to 61% (Salviniales). Larger clades tended to have lower coverage, with the exception of Pteridineae (49% of 1,532 accepted names). The topology of the tree generally agrees with other broadly sampled phylogenies (Pryer et al. 2004; Lehtonen 2011; Schuettpelz and Pryer 2007; Testo and Sundue 2016). The previously largest sampled fern tree included 3,973 tips (Testo and Sundue 2016), but these were not standardized to the same taxonomy used here (COL). When I standardized names in the tree of Testo and Sundue (2016) to COL, 221 were synonyms; thus, the current tree represents an increase in sampling of at least 17.6%.

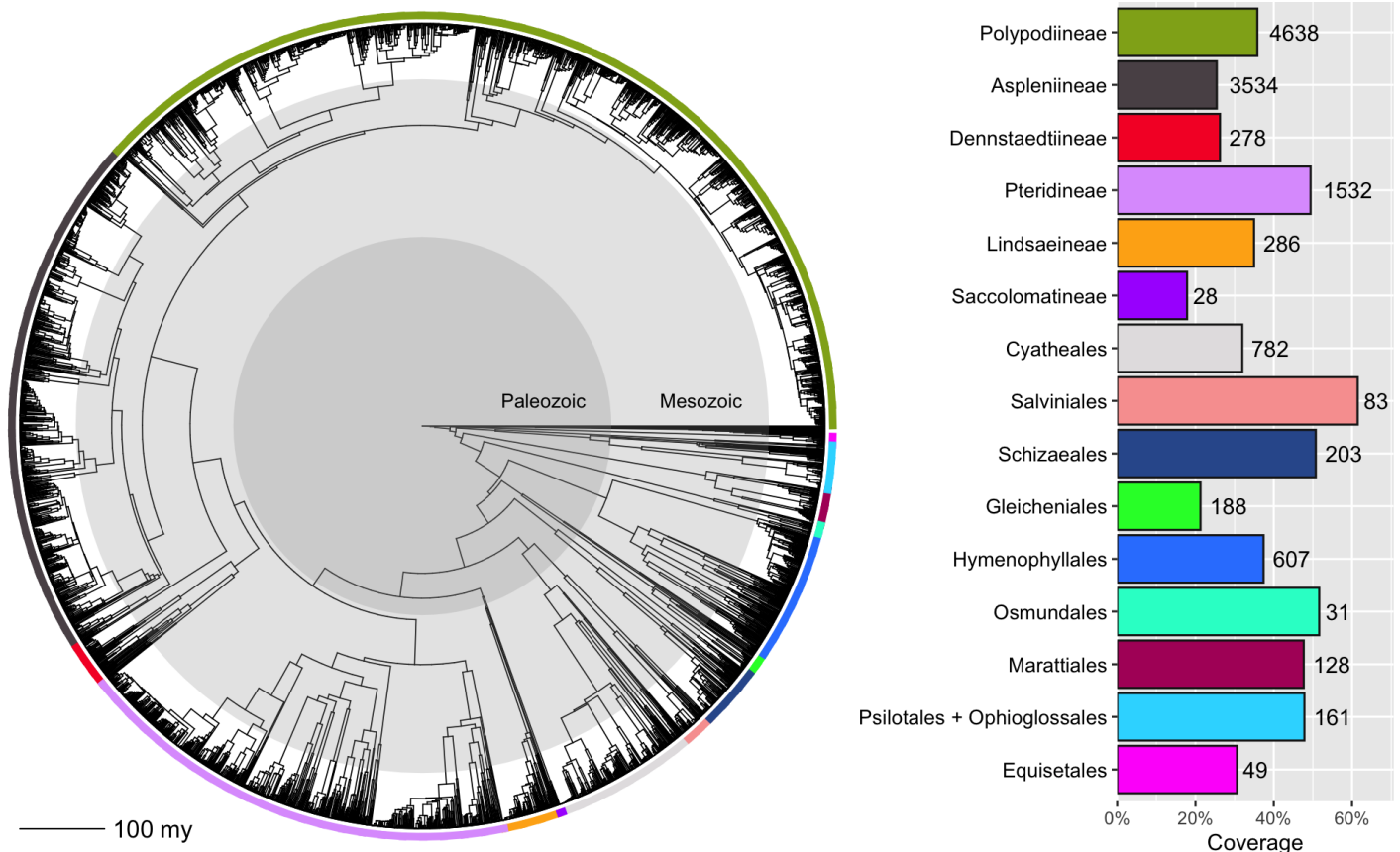


Figure 2: Fern tree of life (FTOL) based on plastid data. Colors correspond to major clade (order or suborder). Barplot shows coverage (percent of species in tree out of all accepted species); numbers next to bars indicate the total number of accepted species in each major clade. Shading on tree indicates geologic era (Cenozoic not labeled).

Future directions: Genomic data

Currently, the vast majority of publicly available DNA sequence data for ferns is Sanger-sequenced plastid gene regions, which is what FTOL will use. Plastid sequences are convenient for phylogenetic analysis because they are essentially a single, uniparentally inherited linkage group, thus free from recombination. However, it is possible that the species tree may differ significantly from the plastid tree due to gene conflict in other parts of the genome. Recent studies using transcriptomics are clarifying the backbone of the fern phylogeny using many (25–2400) nuclear genes from representative species spanning the tree (Rothfels et al. 2015; Qi et al. 2018; Shen et al. 2018). The most comprehensively sampled phylogenomic study targeting ferns is the on-going Genealogy of

Flagellate Plants (GoFLAG) (<http://flagellateplants.group.ufl.edu/>), which seeks to generate genomic data (ca. 300 single-to-low copy nuclear gene regions) for all flagellate plants (bryophytes, lycophytes, ferns, and gymnosperms). The GoFLAG tree can also be thought of a “fern tree of life,” but it differs in only having one specimen per species, significantly less species coverage, and using nuclear (not plastid) genes. The growth of genomic data notwithstanding, many subclades of ferns are under active investigation using plastid markers, and plastid data for previously unsampled species will likely continue to grow at a rapid pace. Thus, genomic studies are complementary, not competing, with FTOL. Analysis combining both nuclear and plastid data should provide new insights into the role of introgression, lineage sorting, and hybridization at deep phylogenetic levels in ferns.

References

- Almeida, Thaís Elias, Alexandre Salino, Jean-Yves Dubuisson, and Sabine Hennerquin. 2017. “Adetogramma (Polypodiaceae), a New Monotypic Fern Genus Segregated from Polypodium.” *PhytoKeys* 78: 109–31. <https://doi.org/10.3897/phytokeys.78.12189> (<https://doi.org/10.3897/phytokeys.78.12189>).
- Antonelli, Alexandre, Hannes Hettling, Fabien L. Condamine, Karin Vos, R. Henrik Nilsson, Michael J. Sanderson, Hervé Sauquet, et al. 2016. “Toward a Self-Updating Platform for Estimating Rates of Speciation and Migration, Ages, and Relationships of Taxa.” *Systematic Biology*, September, syw066. <https://doi.org/10/f3t3b8> (<https://doi.org/10/f3t3b8>).
- Camacho, C, G Coulouris, V Avagyan, N Ma, J Papadopoulos, K Bealer, and T Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10 (1): 421. [papers://ebd7d3df-d8f1-4be8-bcf7-4fb07ababf30/Paper/p1417](https://doi.org/10.1186/1471-2107-10-421) ([papers://ebd7d3df-d8f1-4be8-bcf7-4fb07ababf30/Paper/p1417](https://doi.org/10.1186/1471-2107-10-421)).
- Eiserhardt, Wolf L., Alexandre Antonelli, Dominic J. Bennett, Laura R. Botigué, J. Gordon Burleigh, Steven Dodsworth, Brian J. Enquist, et al. 2018. “A Roadmap for Global Synthesis of the Plant Tree of Life.” *American Journal of Botany* 105 (3): 614–22. <https://doi.org/10.1002/ajb2.1041> (<https://doi.org/10.1002/ajb2.1041>).
- Hennequin, Sabine, Michael Kessler, Stuart Lindsay, and Harald Schneider. 2014. “Evolutionary Patterns in the Assembly of Fern Diversity on the Oceanic Mascarene Islands.” Edited by José María Fernández-Palacios. *Journal of Biogeography* 41 (May): 1651–63. <https://doi.org/10.1111/jbi.12339> (<https://doi.org/10.1111/jbi.12339>).
- Hinchliff, Cody E., Stephen A Smith, James F. Allman, J. Gordon Burleigh, Ruchi Chaudhary, Lyndon M. Coghill, Keith a. Crandall, et al. 2015. “Synthesis of Phylogeny and Taxonomy into a Comprehensive Tree of Life.” *Proceedings of the National Academy of Sciences*, 201423041. <https://doi.org/10.1073/pnas.1423041112> (<https://doi.org/10.1073/pnas.1423041112>).
- Kao, Tzu-Tong, Carl J. Rothfels, Alicia Melgoza-Castillo, Kathleen M. Pryer, and Michael D. Windham. 2020. “Intraspecific Diversification of the Star Cloak Fern (*Notholaena Standleyi*) in the Deserts of the United States and Mexico.” *In Review* 107 (4): 1–18. <https://doi.org/10.1002/ajb2.1461> (<https://doi.org/10.1002/ajb2.1461>).
- Kato, M. 1993. “Biogeography of Ferns: Dispersal and Vicariance.” *Journal of Biogeography* 20 (3): 265–74. [https://doi.org/Doi 10.2307/2845634](https://doi.org/10.2307/2845634) (<https://doi.org/Doi%2010.2307/2845634>).
- Katoh, Kazutaka, Kazuharu Misawa, Keiichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66. <https://doi.org/10.1093/nar/gkf436> (<https://doi.org/10.1093/nar/gkf436>).
- Landau, William Michael. 2018. “The Drake R Package: A Pipeline Toolkit for Reproducibility and High-Performance Computing.” *The Journal of Open Source Software* 3 (21): 550. <https://doi.org/10.21105/joss.00550> (<https://doi.org/10.21105/joss.00550>).

- Lehtonen, Samuli. 2011. "Towards Resolving the Complete Fern Tree of Life." *PLoS ONE* 6 (10): e24851. <https://doi.org/10.1371/journal.pone.0024851> (<https://doi.org/10.1371/journal.pone.0024851>).
- Lehtonen, Samuli, Mirkka M. Jones, Gabriela Zuquim, Jefferson Prado, and Hanna Tuomisto. 2015. "Phylogenetic Relatedness Within Neotropical Fern Communities Increases with Soil Fertility." *Global Ecology and Biogeography* 24: 695–705. <https://doi.org/10.1111/geb.12294> (<https://doi.org/10.1111/geb.12294>).
- Nguyen, Lam Tung, Heiko A. Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74. <https://doi.org/10.1093/molbev/msu300> (<https://doi.org/10.1093/molbev/msu300>).
- Pryer, Kathleen M, Eric Schuettpelz, Paul G. Wolf, Harald Schneider, Alan R. Smith, and Raymond B Cranfill. 2004. "Phylogeny and Evolution of Ferns (Monilophytes) with a Focus on the Early Leptosporangiate Divergences." *American Journal of Botany* 91 (10): 1582–98. <https://doi.org/10.3732/ajb.91.10.1582> (<https://doi.org/10.3732/ajb.91.10.1582>).
- Pteridophyte Phylogeny Group I. 2016. "A Community-Derived Classification for Extant Lycophytes and Ferns." *Journal of Systematics and Evolution* 54 (6): 563–603. <https://doi.org/10.1111/jse.12229> (<https://doi.org/10.1111/jse.12229>).
- Qi, Xinping, Li-Yaung Kuo, Chunce Guo, Hao Li, Zhongyang Li, Ji Qi, Linbo Wang, et al. 2018. "A Well-Resolved Fern Nuclear Phylogeny Reveals the Evolution History of Numerous Transcription Factor Families." *Molecular Phylogenetics and Evolution* 127: 961–77. <https://doi.org/10.1016/j.ympev.2018.06.043> (<https://doi.org/10.1016/j.ympev.2018.06.043>).
- R Core Team. 2019. "R: A Language and Environment for Statistical Computing. 3.6.1."
- Rothfels, Carl J., Fay-Wei Li, Erin M Sigel, Layne Huiet, Anders Larsson, Dylan O. Burge, Markus Ruhsam, et al. 2015. "The Evolutionary History of Ferns Inferred from 25 Low-Copy Nuclear Genes." *American Journal of Botany* 102 (7): 1–19. <https://doi.org/10.3732/ajb.1500089> (<https://doi.org/10.3732/ajb.1500089>).
- Schuettpelz, Eric, and Kathleen M Pryer. 2007. "Fern Phylogeny Inferred from 400 Leptosporangiate Species and Three Plastid Genes." *Taxon* 56 (4): 1037–50. <https://doi.org/10.2307/25065903> (<https://doi.org/10.2307/25065903>).
- Shang, Hui, Michael A Sundue, Ran Wei, Xue Ping Wei, Jun Jie Luo, Li Liu, Pedro B. Schwartzburd, Yue Hong Yan, and Xian Chun Zhang. 2018. "Hiya: A New Genus Segregated from Hypolepis in the Fern Family Dennstaedtiaceae, Based on Phylogenetic Evidence and Character Evolution." *Molecular Phylogenetics and Evolution*. <https://doi.org/10.1016/j.ympev.2018.04.038> (<https://doi.org/10.1016/j.ympev.2018.04.038>).
- Shen, Hui, Dongmei Jin, Jiang-Ping Shu, Xi-Le Zhou, Ming Lei, Ran Wei, Hui Shang, et al. 2018. "Large Scale Phylogenomic Analysis Resolves a Backbone Phylogeny in Ferns." *GigaScience* 7: 1–11. <https://doi.org/10.1093/gigascience/gix116/> (<https://doi.org/10.1093/gigascience/gix116/>).
- Smith, Stephen A, and Brian C. O'Meara. 2012. "treePL: Divergence Time Estimation Using Penalized Likelihood for Large Phylogenies." *Bioinformatics* 28 (20): 2689–90. <https://doi.org/10.1093/bioinformatics/bts492> (<https://doi.org/10.1093/bioinformatics/bts492>).
- Testo, Weston L., and Michael A Sundue. 2016. "A 4000-Species Dataset Provides New Insight into the Evolution of Ferns." *Molecular Phylogenetics and Evolution* 105: 200–211. <https://doi.org/10.1016/j.ympev.2016.09.003> (<https://doi.org/10.1016/j.ympev.2016.09.003>).

Tryon, R. 1986. "The Biogeography of Species, with Special Reference to Ferns." *The Botanical Review* 52 (2): 117–56. papers://ebd7d3df-d8f1-4be8-bcf7-4fb07ababf30/Paper/p1319 (papers://ebd7d3df-d8f1-4be8-bcf7-4fb07ababf30/Paper/p1319).

Zhang, Liang, Xin-Mao Zhou, Zhen-Long Liang, Xue-Ping Fan, Ngan Thi Lu, Min-Shu Song, Ralf Knapp, Xin-Fen Gao, Hang Sun, and Li-Bing Zhang. 2020. "Phylogeny and Classification of the Tribe Lepisoreae (Polypodiaceae; Pteridophyta) with the Description of a New Genus, *Ellipinema* Gen. Nov., Segregated from *Lepisorus*." *Molecular Phylogenetics and Evolution*, 106803. <https://doi.org/10.1016/j.ympev.2020.106803> (<https://doi.org/10.1016/j.ympev.2020.106803>).