

Global-Scale Phylogenomics of the Nitrogen-Fixing Clade

Heather Rose Kates*¹ (hkates@ufl.edu), Jean-Michel Ane², Kelly Balmant³, Daniel Conde³, Matthew Crook⁴, Christopher Dervinis³, Robert P. Guralnick¹, Thomas Irving², **Matias Kirst**³, Sara Knaack², Lucas Maia², Sushmita Roy², Ryan Folk¹, Douglas E. Soltis¹, Pamela S. Soltis¹

¹University of Florida, Florida Museum of Natural History, Gainesville, FL; ²University of Wisconsin, Madison, WI; ³University of Florida, Gainesville, FL; ⁴Weber State University, Ogden, UT

www.nitfix.org

Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" is focused on identifying the genomic novelties that enable the symbiotic relationship between nodulating plants and N-fixing bacteria to support genetically engineering this capability into bioenergy crops. The first aim of this project is a comparative phylogenomic study of the nodulating clade to uncover the genomic novelties that were required for the evolution of the root nodule symbioses. A massively improved phylogenetic framework for the N-fixing clade that includes genetic data for nodulation genes for all species in the phylogeny will provide a robust, revised understanding of the exact ancestral origin of nodulation and the evolution of the predisposition to nodulate. This framework will also inform a experiments comparing close relatives that nodulate or do not nodulate to identify the genes underlying this trait. To achieve these aims we developed a cutting-edge phylogenetic approach with an unprecedented sampling effort that is described below.

Robust phylogenetic inferences on the origin of the predisposition to nodulation, and events of nodulation gain and loss, are key to understanding the evolutionary lability and, consequently, the likelihood of successful transferability of N-fixing symbioses among lineages of angiosperms. A well-resolved and well-sampled N-fixing clade phylogeny is therefore a prerequisite to the discovery of genes that determine nodule development. Multiple phylogenetic analyses have been conducted on the N-fixing clade with the aim of elucidating the origins of N-fixing symbioses; however, these analyses have relied on trees estimated using a few genes and in which species sampling in the N-fixing clade was limited. We present our first steps toward a revised phylogeny based on deliberate and extensive sampling, phylogenomic data, and rigorous statistical analysis. These will allow more accurate inference of precursors of N-fixing symbioses, gain and/or loss events, and potential transferability of the capability to crop plants not in the N-fixing clade. The final phylogeny will include 15,000 of 30,000 species of the N-fixing clade, making it the best sampled and phylogenetically studied major clade of plants.

This ambitious phylogenetic study requires novel strategies for rapid specimen sampling and DNA data generation. Elements of our strategy that enable geographically and taxonomically comprehensive sampling include a protocol for rapid tissue sampling of 15,000 historical specimens, a high-throughput, high-yield DNA extraction protocol specifically suited to degraded DNA, a targeted-enrichment DNA sequencing kit that works across phylogenetic

scales and includes key functional genes, and scalable information management and processing using a project database and custom bioinformatics tools. The protocols and tools developed for this project can be leveraged as a toolkit for phylogenetic researchers to generate very large phylogenetic datasets and will facilitate greater utilization of historical specimens in phylogenetic research.

The results so far include our completion of rapid sampling, data generation, and data assembly and highlight preliminary phylogenetic results and their usefulness for meeting project goals. Our preliminary phylogeny is based on 100 nuclear loci sequenced for 9,876 species from the N-fixing clade (Fig. 1) and informs comparisons between nodulating and non-nodulating species based on novel phylogenetic relationships. We also highlight a set of functional genes that may be correlated with nodulation based on our preliminary results.

This result represents an initial analysis comprising two thirds of our genomic and taxonomic sampling effort. We will leverage the total phylogenetic and comparative genomic results to discover gene candidates that potentially underlie nodule development so that these genes can be tested for function in nodulating and non-nodulating model systems.

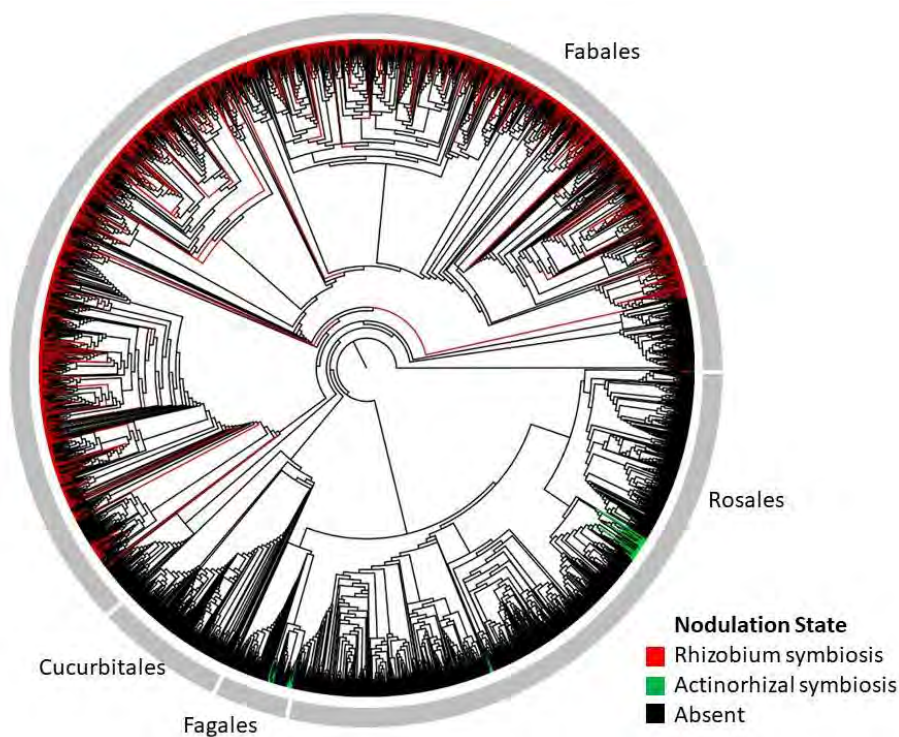


Fig. 1. Preliminary phylogeny of a subset of the N-fixing clade based on our gene capture data available at the time of the abstract submission. This tree represents two thirds of the complete phylogeny that will be available for meeting project goals. Terminal branches are colored by presence or absence of nodulation based on scorings from Werner et al. (2014) and Afkhami et al. (2018).

This work is supported by the Biosystems Design Program from the Biological and Environmental Research (BER) Office of Science at the U.S. Department of Energy (grant #DE-SC0018247).