

## **NitFix: Global-Scale Phylogenomics of the Nitrogen-fixing Clade**

Heather Rose Kates,<sup>1</sup> **Ryan Folk**\*<sup>1</sup> (rfolk@flmnh.ufl.edu), **Robert P. Guralnick**,<sup>1</sup> **Douglas E. Soltis**,<sup>1</sup> **Pamela S. Soltis**,<sup>1</sup> **Jean-Michel Ane**,<sup>2</sup> **Sushmita Roy**,<sup>2</sup> **Matias Kirst**<sup>3</sup>

<sup>1</sup>University of Florida, Florida Museum of Natural History, Gainesville, FL; <sup>2</sup>University of Wisconsin, Madison, WI; <sup>3</sup>University of Florida, Gainesville, FL

[www.nitfix.org](http://www.nitfix.org)

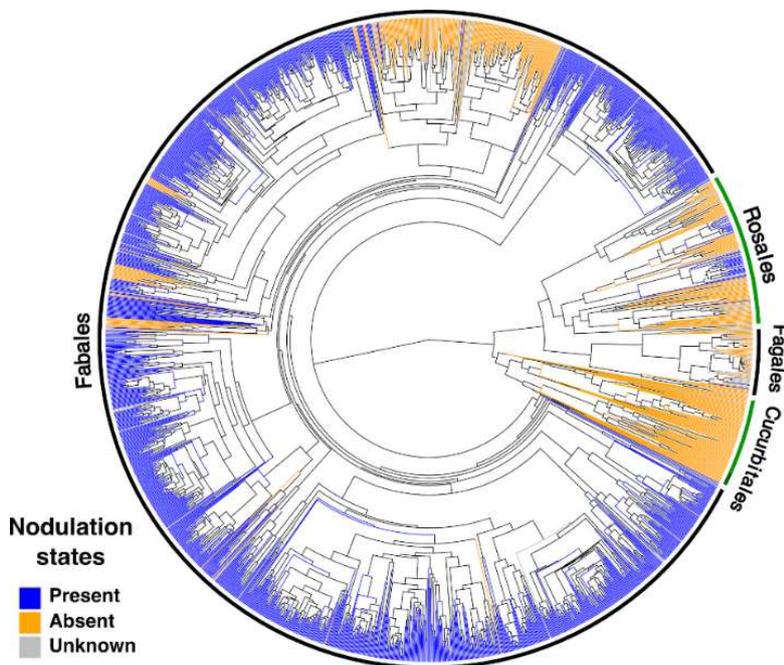
**Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" is focused on identifying the genomic novelties that enable the symbiotic relationship between nodulating plants and Nitrogen-fixing bacteria to support genetically 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 these root nodule symbioses. A massively improved phylogenetic framework for the Nitrogen-fixing clade that includes genetic data for nodulation genes for all species in the phylogeny will provide a robust, improved framework for understanding the origin and evolution of nodulation. These data will also inform a series of experiments comparing close relatives that do or don't to identify the genes underlying nodulation. To achieve these goals, we have developed a cutting-edge phylogenetic approach with an unprecedented sampling effort which is illustrated by the project described below.**

The robust inference of the origin of nodulation, and of how nodulation has been gained and/or lost, are key to understanding the evolutionary lability and thus the likelihood of successful transferability of Nitrogen-fixing symbioses to diverse angiosperm crops. A well-resolved and well-sampled Nitrogen-fixing clade phylogeny is therefore a prerequisite to identifying this evolutionary history of symbiosis and to jump-starting the discovery of a core set of genes shared across this diverse group that determine nodule development. Multiple phylogenetic analyses have been conducted on the Nitrogen-fixing clade with the aim of elucidating the origins of Nitrogen-fixing symbioses; however, these analyses have relied on trees estimated using a few genes and in which species sampling in the large Nitrogen-fixing clade (30,000 species) 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 Nitrogen-fixing symbioses, gain and/or loss events, and potential transferability of the capability to crop plants not in the Nitrogen-fixing clade. The final phylogeny will include 15,000 species in the Nitrogen-fixing clade.

This ambitious phylogenetic study requires novel strategies for rapid specimen sampling and DNA data generation. Novel elements of our strategy that enable geographically and taxonomically comprehensive sampling include (1) a protocol for rapid tissue sampling of 15,000 historical (herbarium) specimens, (2) a high-throughput, high-yield DNA extraction protocol specifically suited to degraded DNA, (3) a targeted-enrichment DNA sequencing kit that works across phylogenetic scales and includes key functional genes, and (4) a custom-developed, scalable laboratory information management system (LIMS) together with custom

analysis tools. The protocols and methods developed for this project can be broadly 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 (13% of our ultimate effort) highlight the rapid progress we have made in sampling and data generation and demonstrate the utility of our phylogenetic products for meeting project goals. Our preliminary phylogeny is based on 100 nuclear loci sequenced for 1,900 species from across the Nitrogen-fixing clade (Fig. 1); this tree informs our interpretation of patterns of presence/absence of ~129 nodulation genes inferred from genome screening of all species in our phylogeny. We highlight a set of genes for which there is evidence of correlated evolution with nodulation phenotypes. We show that these evolutionary relationships differ among legume and non-legume nodulators, suggesting more limited genetic homology among these organs than suggested with fewer taxa. We will leverage the total phylogenetic and comparative genomic results to discover gene candidates that potentially underlie nodule development and test these for function in nodulating and non-nodulating model systems.



**Fig. 1.** Preliminary phylogeny of a subset of the Nitrogen-fixing clade based on our gene capture data available at the time of the abstract submission. This tree represents ~13% of the complete phylogeny that will be available for meeting project goals. Terminal branches are colored by presence or absence of nodulation.

*This project is funded 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).*