Identification of the Minimal Genetic Toolkit Required for Nitrogen Fixation Using Comparative Genomics and Single-Cell Transcriptomics

Daniel Conde,[1] Kelly Balmant,[1] Paolo Triozzi,[1] Wendell J. Pereira,[1] Thomas B. Irving,[2] Lucas G. Maia,[2] Sanhita Chakraborty,[2] Heather R. Kates,[1] Sara A. Knaack,[2] Henry W. Schmidt,[1] Pamela Soltis,[1,3] Douglas Soltis,[1] Robert Guralnick,[3] Ryan Folk,[4] Sushmita Roy,[2] Jean-Michel Ané,[2] and **Matias Kirst**[1]* (mkirst@ufl.edu)

[1]University of Florida, Gainesville, FL; [2]University of Wisconsin, Madison, WI; [3]University of Florida, Florida Museum of Natural History, Gainesville, FL; Mississippi State University, Starkville, MS.

http://nitfix.org; Twitter: @nit_fix

Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" aims to identify the evolutionary events that enabled the symbiotic relationship between nodulating plants and nitrogen-fixing bacteria. Identifying a minimum genetic toolkit required for nodulation is a first step towards genetically engineering this trait into bioenergy crops. We are applying multiple methods to discover the genome novelties of nitrogen fixers. Here we present the outcome of two approaches: comparative genomics of nodulating and non-nodulating species and single-cell transcriptomics of nodule development in *Medicago*. Because multiple genes are likely to be required to engineer nitrogen fixation, we also describe a newly devised synthetic biology method for high-throughput parallel testing of multigene combinations for their function on nodule formation and nitrogen-fixation.

Nitrogen (N) is essential for plant growth because of its fundamental role as a component of DNA, RNA, and amino acids. Most plants cannot obtain N directly from the atmosphere but depend on its availability in the soil, in the form of nitrate, ammonium, or amino acids. Nitrogen fertilization comes at a high financial and environmental cost. In contrast, several species of four angiosperm orders (Fabales, Fagales, Rosales, and Cucurbitales) can establish a symbiotic relationship with bacteria that convert atmospheric N to ammonium. This group of species is referred to as the nitrogen-fixing clade (NFC). After bacteria infect the root, they become established in newly developed nodules. While the plant benefits by absorbing the produced ammonia, it provides nutrients to the bacteria.

A single evolutionary event may have created a predisposition for the symbiotic relationship between N-fixing bacteria and plants in the NFC. Our phylogenetic results show that many gains and losses followed this early event. In this complex evolutionary scenario, several genes and regulatory elements are likely critical for engineering N-fixation into bioenergy crops. We are applying approaches such as comparative genomics of nodulating and non-nodulating species and transcriptome analysis of single cells in nodules to identify them.

Comparative genomics – To detect genomic novelties that enable nodulation using comparative genomics, we aligned plant genomes from N-fixing and non-fixing species within the NFC and an outgroup. Next, we identified loci that display evolutionary signatures compatible with gains or losses of the trait. Based on phylogeny and analysis of coding sequences, we find strong evidence that a deletion in a Lysin Motif Receptor Like Kinases is associated with symbiosis in the NFC,

having appeared multiple times in nodulating species. A similar strategy was used to survey the regulatory regions of genes implicated in nodulation. This analysis showed that a set of conserved non-coding sequences immediately upstream of putative homologs of the *Mt*CRE1 Cytokinin Receptor are absent among species in the outgroup. This cytokinin receptor has been previously shown to be an essential part of the signaling mechanism that triggers nodule formation. The effect of these and other genes and regulators are being functionally validated to verify their molecular role in nodulation.

Single-cell transcriptomics – The cell-specific gene expression program that leads to cell division activation in the root pericycle and cortex and, posteriorly, the nodule primordia formation remains largely unknown. Uncovering the regulators of this program may determine the genetic elements critical for engineering nodule formation in non-nodulating species. We developed a high-quality nuclei isolation protocol suitable for use in the 10× Genomics Chromium system to address this question. This method was used to characterize the single-cell transcriptome of *Medicago* roots after treatment with the N-fixing bacteria *Sinorhizobium meliloti* (Fig. 1). Analysis of the single-cell transcriptome data identified cell-type-specific gene expression programs, including potential regulators of cell division initiation in the inner root cells, such as *SCARECROW*, *YUCCA 1*, and *PLETHORA 5*. Most strikingly, gene expression signatures specific to mature nodule compartments appear to emerge within few hours of treatment with *S. meliloti*. Thus, cell fate may be determined very early in this developmental process. We are now applying pseudotime, trajectory analysis to detect additional regulators for validation, as described below.

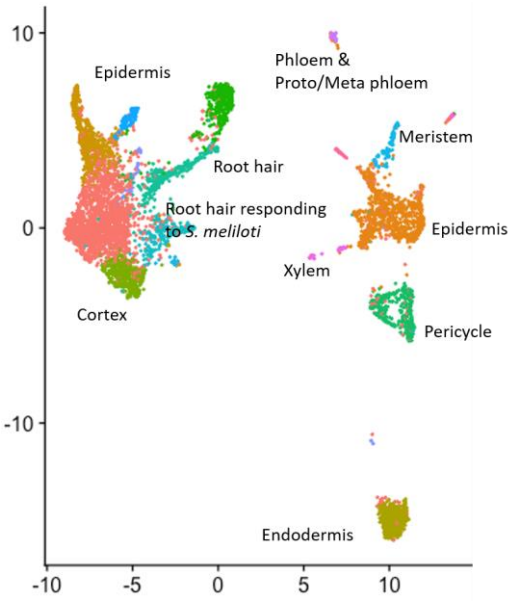

Fig. 1. Clustering of single-cell RNA-seq data generated from *Medicago* roots treated with *Sinorhizobium meliloti* for 24 hrs. Cell type were annotated according to the expression of known marker genes.

Nodule organogenesis is controlled by multiple genes that act in concert to regulate this complex developmental program. Evidence from work described above indicates that multiple genetic elements will have to be engineered to introduce nodule organogenesis in bioenergy crops. This fact motivated us to design a Golden Gate-based system to screen multigene combinations for engineering nodule-like structures in poplar. This system allows the random combination of genes and promoters up to six transcriptional units simultaneously to be tested in *Agrobacterium rhizogenes*-mediated poplar roots. Preliminary results from the application of this system indicate the possibility of engineering nodule-like structures in poplar roots.

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).