Deciphering N-fixing symbiosis signaling in *Medicago* with dynamic regulatory module networks (DRMNs)

Daniel Conde,[2] Sara Knaack[1,]* (saknaack@wisc.edu), Lucas Gontijo Silva Maia,[3] Thomas Irving,[3] Kelly Balmant,[2] Matthew Crook,[4] Christopher Dervinis,[2] Heather Kates,[5] Wendell Pereira,[2] Ryan Folk,[6] Robert Guralnick,[5] Douglas Soltis,[5,7]Pamela Soltis,[5] Jean-Michel Ane,[3] Sushmita Roy[1,8]. Matias Kirst,[2]

Wisconsin Institute for Discovery, University of Wisconsin at Madison, Madison, WI; [2] School of Forest Resources and Conservation, University of Florida, Gainesville, FL; Departments of Bacteriology and Agronomy, University of Wisconsin at Madison, Madison, WI; [4] Department of Microbiology, Weber State University, Ogden, UT; [5] Florida Museum of Natural History, University of Florida, Gainesville, FL; [6] Department of Biology, Mississippi State University, Mississippi State, MS; [7] Department of Biology, University of Florida, Gainesville, FL; and [8] Department of Biostatistics and Medical Informatics, University of Wisconsin at Madison, Madison, WI

http://NitFix.org

Project Goals: Our research goal is to identify genomic elements required for the symbiotic relationship between nodulating plants and nitrogen (N)-fixing bacteria. In order to identify these elements, we measured and analyzed novel RNAseq and ATACseq time course data obtained from Medicago plants subjected to treatment with lipo-chitooligosaccharide (LCO) Nod factors. The gene regulatory network involved in the response to LCOs were discovered by applying a novel computational method that defines dynamically transitioning genes and predicts key regulators of these genes. Prioritized regulators and their target genes will be validated experimentally.

Nitrogen fixation occurs naturally in a limited lineage of plants (N-fixing clade) through a symbiotic relationship where plant root nodules are colonized by N-fixing bacteria. The vast majority of plants do not have this ability, requiring the application of fertilizers for agricultural production. Some molecular components of the symbiosis pathways are known, yet the gene regulatory network involved in this process is not well understood. A detailed characterization of the regulatory network controlling symbiosis can provide important insights into the key regulators that are missing from non N-fixing species. More importantly, this can prioritize the elements and regulators that could be engineered into plants lacking the ability to form symbiotic relationships.

We measured transcriptomic (with RNA-seq) and chromatin accessibility (with ATAC-seq) profiles in *Medicago* roots treated with LCOs over a 24 hour time course. LCOs are a component of the symbiotic pathway and part of the early signaling processes needed to trigger the establishment of symbiosis in *Medicago*.

We analyzed the expression data using a time-series clustering algorithm (Escarole[1]) and identified groups of genes exhibiting coherent dynamic change over time, several of which are associated with coordinated changes in accessibility. In parallel, we examined the signal profile of peak calls from the ATAC-seq data collected over the same time points, and identified 81,114 putative regulatory regions, including clusters of those regions that exhibit coherent dynamic change over time. These dynamically changing peaks are proximal to 39,755 genes (of 51,316 in the v1.6 annotation of the v5 *Medicago* genome). Furthermore, the accessibility profile around the gene promoter is correlated (Pearson correlation of 0.50 or greater) to transcriptomic changes detected over the duration of the experiment in 28.5% of genes.

We next applied a novel approach, Dynamic Regulatory Module Networks (DRMNs), for integrating RNA-seq and ATAC-seq time courses along with known sequence-specific motif instances. With this approach we identified modules of co-expressed genes at each time point, in addition to per-module regulatory networks predictive of gene expression within each module. The network edges are based on predicting gene expression using gene promoter and motif site accessibility values. We examined known regulators of the nodulation pathway and found them to be enriched within specific groups of genes exhibiting transitions in their expression state. Several of our modules are enriched for root hair elongation, defense response to bacterium, chromatin organization and MAPK cascade processes. We identified transitioning genes using DRMN results and predicted regulators for these genes using a network inference algorithm[2]. The regulators include IBM1, EDN3, MTF1, EIN3, SHY2, BHLH, AB15, RRB15, MTF1, NSP1, RRB9, AB14, and SMB. Several of these regulators (EIN3, NSP1) are known to be involved in nitrogen fixation providing early promising support to our predictions. We are currently identifying targets for these regulators and experimentally testing them.

Taken together our dataset and associated analyses provide a valuable resource to the plant community to understand the gene regulatory programs controlling Nitrogen fixation.

References

1. Inferring Regulatory Programs Governing Region Specificity of Neuroepithelial Stem Cells during Early Hindbrain and Spinal Cord Development. Deborah Chasman, Nisha Iyer, Alireza Fotuhi Siahpirani, Maria Estevez Silva, Ethan Lippmann, Brian McIntosh, Mitchell D. Probasco, Peng Jiang, Ron Stewart, James A.Thomson, Randolph S. Ashton, Sushmita Roy. Cell Systems, 2019.

2. Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks. Sushmita Roy, Stephen Lagree, Zhonggang Hou, James A. Thomson, Ron Stewart, Audrey P. Gasch. PLoS Comput Biol, 2013.

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).