Leveraging temporal change in chromatin accessibility to predict regulators of N-fixing symbiosis in *Medicago* with dynamic regulatory module networks (DRMNs)

Sara Knaack[1,*] (saknaack@wisc.edu), Daniel Conde,[2] 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: We seek to identify genomic elements required for the symbiotic relationship between nodulating plants and nitrogen (N)-fixing bacteria. To identify such elements we generated parallel RNAseq and ATACseq time course data from *Medicago* plant subjected to lipo-chitooligosaccharide (LCO) (Nod factors) treatment. The gene regulatory network involved in LCO response was identified by applying a novel computational method to identify dynamically expressed (transitioning) genes, and predict key regulators of these genes. Prioritized regulators and their target genes were examined based on current literature and will be validated experimentally in the future.

Nitrogen fixation occurs naturally in a small number of plant species through a symbiotic relationship in plant root nodules colonized by N-fixing bacteria. Components of the symbiosis pathways are known, yet the gene regulatory network controlling this process is not thoroughly understood. 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 the experiment emulates the early signaling processes in the establishment of symbiosis in *Medicago*.

To define the gene regulatory network from this parallel time course we applied a novel computational method, Dynamic Regulatory Module Networks (DRMNs). With DRMN we inferred modules of similarly-expressed genes at each time point, and per-module regulatory networks predictive of gene expression within each module. Module regulatory network edges are based on predicting gene expression from the accessibility (ATAC-seq) of gene promoters (+/- 2 kbp) and motif sites of known regulatory proteins mapped to within 10 kbp upstream to 1 kbp down-stream of a given gene transcription start site (TSS). Among the top regulators

identified in the module networks include IBM1, EDN3, MTF1, EIN3, SHY2, NSP1, and RRB9. Several (EIN3, NSP1) are known to be involved in nitrogen fixation. Our modules are furthermore enriched for root hair elongation, defense response to bacterium, chromatin organization and the MAPK cascade, recapitulating key features of symbiosis.

We leveraged the results of DRMN to define 10,176 transitioning genes (those with changing module assignment across time) and then clustered into 79 clusters. We predicted regulators of these transitioning genes using accessibility of motifs in their promoters using a structured sparsity framework, MTG-LASSO.[1] This approach leverages a group structure of similar transitioning genes to learn a statistically robust model compared to standard expression-based network inference. We identified a high confidence set of regulatory network edges which included 126 regulators (motifs) and 5,978 genes. Of the regulators identified with a large number of connections, many were identified from the module-level results (EIN3, IBM1, and MTF1), and some were particular to this analysis (including CAMTA1 and CYCRE).

To validate our regulatory network we overlapped predicted target genes of the EIN3 motif to sets of differentially expressed (DE) genes called between SKL/EIN2Δ mutant and wild type time course data from Larrainzar et al.[2] The SKL/EIN2Δ data was used for this validation because EIN2 and EIN3 function is closely related. Our inferred EIN3 targets overlap significantly (hypergeometric test $p<0.05$) with DE gene sets inferred from each time point of the Larrainzar et al.[2] dataset. Despite the differing techniques of our experiment and that of Larrainzar et al.,[2] these results demonstrate the effectiveness of our DRMN/MTG-LASSO framework for inferring regulatory networks. As future work, we will experimentally validate such predictions by functional perturbations to the regulator and measuring nodulation phenotype and downstream targets. Our dataset and predictions are a valuable resource to the plant community to study the gene regulatory programs controlling Nitrogen fixation.

References

1. Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks. S. Roy, et al. PLoS Comput Biol, 2013.

2. Deep Sequencing of the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. E. Larrainzar, et al. Plant Physiology, 2015.

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).