Infernet: Gene Function Inference By Leveraging Large, Organ-Specific Expression Datasets And Validation Of Non-Redundant Regulators

Rajeev Ranjan*(ranjan9@purdue.edu, rajeevmbbtac@gamil.com)[1], Ying Li[1], Karen Hudson[2] and **Kranthi Varala**[1],

[1]Purdue University, West Lafayette, IN; [2] USDA-ARS, West Lafayette, IN

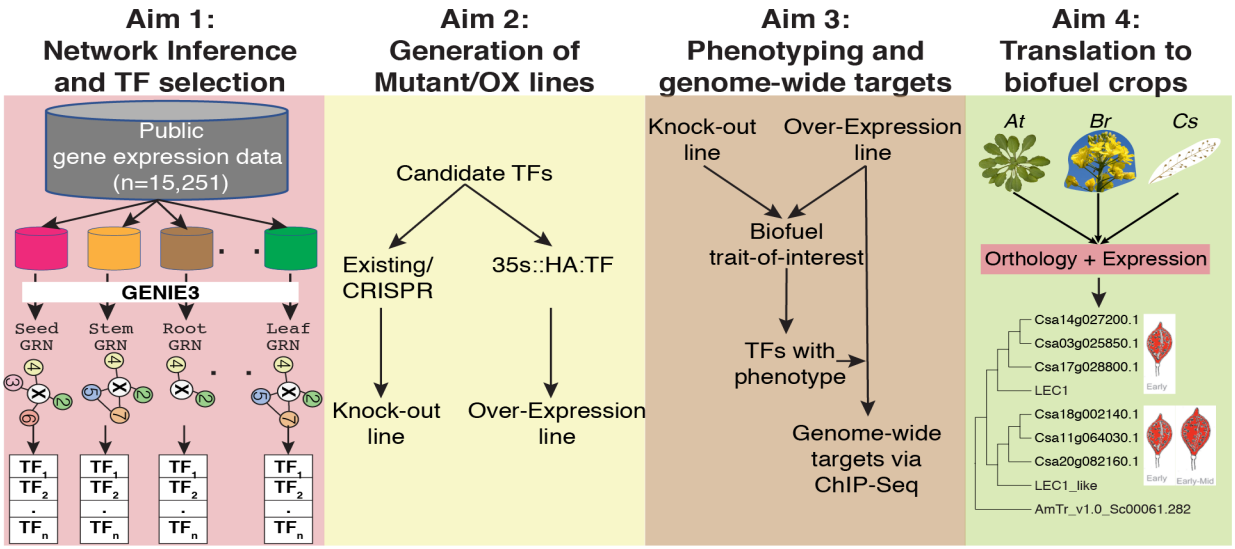
https://www.purdue.edu/hla/sites/varalalab/infernet

Project Goals

This project combines computational approaches, e.g., machine learning, network inference and phylogenomics, with molecular approaches, e.g., metabolite profiling and ChIP-Seq, to find novel transcription factors (TF) that regulate traits of agronomic or biofuel interest. This project focuses on the biofuel trait of seed oil synthesis as a proof of concept that is extensible to any agronomic/biofuel trait of interest. This project focuses on regulation of a biological process of interest (e.g., lipid biosynthesis) in an organ specific manner (e.g., in seeds) and by estimating the likelihood of a given TF being redundant in its function (Aim 1). We then validate our functional predictions, using transgenic lines (Aim 2), via phenotypic assays (Aim 3a) and by identifying the specific targets these TFs regulate (Aim 3b). Finally, we translate the validated TF regulation knowledge gained in a model species (*Arabidopsis*) to biofuel crops (e.g., *Camelina sativa*) (Aim 4).

Abstract

Gene regulatory network inference [1] from public RNA-Seq data (Aim 1) predicted TF regulators of seed lipid biosynthesis. The list of top 10 candidate TFs included four known regulators of this process and many novel TFs that are predicted to have a strong influence on seed lipid biosynthesis. We have identified and collected mutant lines and completed phenotyping for traits like seed size and seed weight. Phenotyping for alteration in seed lipid profile



is under progress (Table 1). In addition, we generated over-expression lines for each of these candidate TFs using a set of 3 promoters: i. a generic 35S promoter for robust plant-wide over-expression ii. a seed-specific NapinA promoter for robust seed-specific over-expression and iii. native promoter driven expression. Screening of 35S and Napin promoter driven T_1 over-expression transgenic lines is completed and isolation of homozygous T2 lines is under progress. While screening of native promoter driven T1 lines is under progress. T2 overexpression lines showing higher level of TF expression will be used for phenotyping of seed lipid profile along with seed size and weight. Preliminary results show a significant change in total FA content for one novel TF regulator, and significant changes in FA composition for another novel TF regulator.

Table 1 *Status of mutant identification and phenotyping, along with generation of over-expression lines for each candidate TF. All TFs were conjugated with a HA tag to enable ChIP-Seq assays to identify global targets of TF binding.*

TF Name	Mutant				35Spro:TF-HA			NapApro: TF-HA		
	Mutant Name	ABRC ID	Seed phenotype	Lipid profile	Transfo ration	Transgene Screening (T1)	Transgene Screening (T2)	Transfor mation	Transgene Screening (T1)	Transgene Screening (T2)
bHLH93	*bhlh093_3*	SALK_121082	Done	Ongoing	Done	Done	Ongoing	Done	Done	Ongoing
HB25	*hb25_2/hb25_4*	SALK_014023/SAIL_517_E03	Done	Ongoing	Done	Done	Ongoing	Done	Done	
DIV2	*div2_2*	SALK_208938								
SRM1	*srm1_2*	SALK_206518	Done	Ongoing	Done	Done	Ongoing	Done	Done	Ongoing
MYB30	*myb30_1*	SALK_122884	Done	Ongoing	Done	Done		Done	Done	
DAG2	*dag2_1*	SALK_201125	Done	Ongoing				Done	Done	Ongoing
CESTA	*cesta_1/cesta_3*	SALK_124840/SAIL_674_A01			Done	Done	Ongoing			
TGA4	*tga4_1*	SALK_127923	Done	Ongoing	Done	Done	Ongoing	Done	Done	Ongoing
SPL12	*spl12_1*	SALK_142295	Done	Ongoing						
AGL18	*agl18_2/agl18_5*	SALK_004483//SALKseq_69591	Done	Ongoing						
CAMTA6	*camta6_1*	SALK_078900	Done	Ongoing						
MYB118	*myb118_1*	SALK_111812	Done	Ongoing						
MYB115	*myb115_1/myb115_2*	SALK_202795/SALK_044168	Done	Ongoing	Done	Done	Ongoing	Done	Done	Ongoing
WRI 1	*wri1_1*	CS69538			Done	Done	Ongoing			

The GRN inference used in Aim 1 only considered TFs as regulators of gene expression. Subsequently, we repeated the inference pipeline using TFs + epigenetic (i.e., DNA and histone) modifiers as potential regulators of gene expression. Using the shoot and root apical meristems as use cases we identified many known and novel non-TF regulators of plant development. A manuscript describing this study was recently published (McCoy *et. al.,* 2021).

Publications

McCoy, R.M., Julian, R., Kumar, S.R.V., Ranjan, R., Varala, K., Li, Y. A systems biology approach to identify essential epigenetic regulators for specific biological processes in plants. Plants, 2021, 10 (2), 364.

References

1. Huynh-Thu, V.A., A. Irrthum, L. Wehenkel and P. Geurts, Inferring regulatory networks from expression data using tree-based methods. PLoS One, 2010. 5(9).

The InferNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.