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

Kranthi Varala[1]* (kvarala@purdue.edu), Ying Li[1], and Karen Hudson[2]
[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).
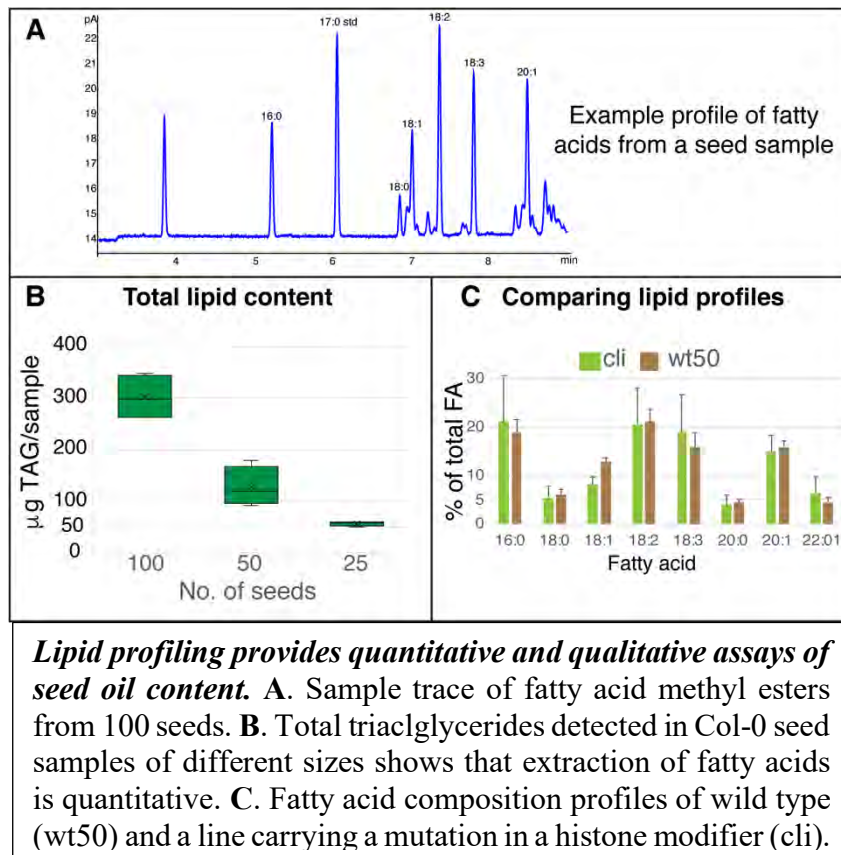
Project Goals

This DOE project focuses on the seed lipid biosynthesis trait as a proof-of-concept and aims to identify multiple novel regulators of this trait. For a given trait of interest, we predict novel TF regulators by leveraging the power of large expression data sets in a model species and then translate that knowledge to a target biofuel species for immediate benefit to the DOE mission. Two of the principal challenges in determining gene function are: 1. Subtle phenotypes that occur in
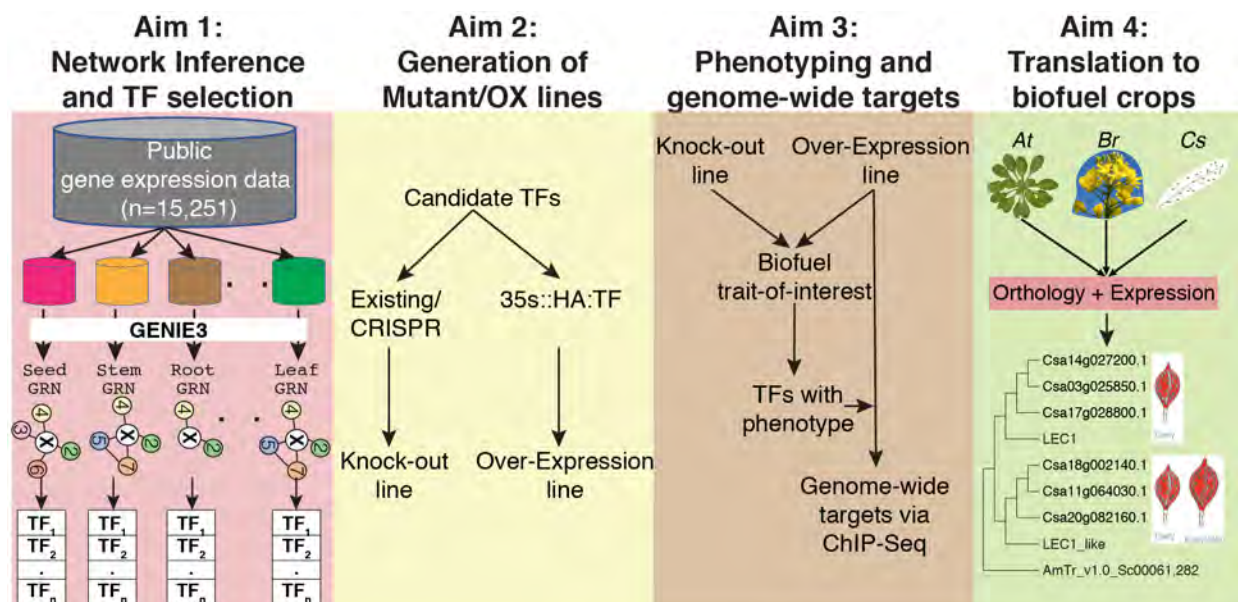
organ and/or developmental stage specific manner and 2. Functional redundancy among gene family members masks phenotypic effects via genetic or dosage compensation. The data and algorithms developed in this project should prove valuable to any research group interested in discovering novel TF regulators of any plant process. This project provides the data and tools to discover such regulators in the model plant Arabidopsis as well as a computational pipeline to translate the knowledge gained in Arabidopsis to a biofuel species.

Progress

Gene regulator network inference [1] from public RNA-Seq data (Aim 1) identified the top predicted regulators of seed lipid biosynthesis. This list included four known regulators of this process in the top 10 predicted TFs and many novel TFs that are predicted to have a strong influence on seed lipid biosynthesis. We have identified and collected mutant lines in most of the novel TFs and are in the process of phenotyping them to detect changes in seed lipid profile. In addition efforts to generate over-expression lines for each of these candidate TFs is ongoing.

Lipid profiling provides quantitative and qualitative assays of seed oil content. **A**. Sample trace of fatty acid methyl esters from 100 seeds. **B**. Total triaclglycerides detected in Col-0 seed samples of different sizes shows that extraction of fatty acids is quantitative. **C**. Fatty acid composition profiles of wild type (wt50) and a line carrying a mutation in a histone modifier (cli).

In parallel, we have built node-vicinity-networks (NVNs) for each of these TFs from the full gene expression data set (n=15,251). These NVNs are being used to identify non-redundant TFs and prioritize those for phenotyping assays.

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).