TEXAS A&M UNIVERSITY
Engineering

Texas A&M Engineering
Experiment Station

Engineering Approaches for Agricultural Genomics

Aniruddha Datta, email: datta@ece.tamu.edu
Website: https://engineering.tamu.edu/electrical/profiles/adatta.html
Center for Bioinformatics and Genomic Systems Engineering (CBGSE)
Department of Electrical and Computer Engineering

Challenges in Agriculture

- Plants: adaptation to stress for survival

- Abiotic Stresses: Extreme temperature, drought, soil toxicity, lodging.

- Biotic Stresses: Viruses, bacteria, insects, weeds, animals.

Gene Regulatory Networks

- Defense response against stress: extensive biochemical pathways
- Originates at the genomic level but manifests as observable phenotypic traits.
- Examples: closing of stomata under drought, reduced transpiration rates under saline stress.
- Biochemical interactions: model using Gene Regulatory Networks (GRNs).

Gene Regulatory Networks

- GRNs also underlie other biological processes, e.g. grain protein content, flowering.
- Intricate networks of transcription factors, proteins, and target genes.
- Identifying key **trait controlling regulators** in GRNs: breed plants with desirable traits.
- Examples: drought resistant cotton, lysine (Protein) rich rice, lodging resistant sorghum.

Trait-Regulator Identification Methods

- Most Bioinformatics approaches use only data: e.g. differential gene expression (DE) analysis

- Univariate approaches, assigning cause-effect (e.g. drought resistance) relationships to most significant gene (cause).

Trait-Regulator Identification Methods

- Modeling GRNs using Probabilistic Graphical Models (PGMs): can capture multivariate interactions

- Can capture prior knowledge (pathways)

- Also can be learned from data

- PGMs: a good modeling paradigm for GRNs

PGMs

- PGMs consist of two major components :
    1. The network structure (network topology): connectivity
    2. The network parameters (stochastic component): quantify interactions

- Network structure from pathway info: lacking in agricultural context

- Can also be learned from data

- Network parameters: estimated using data

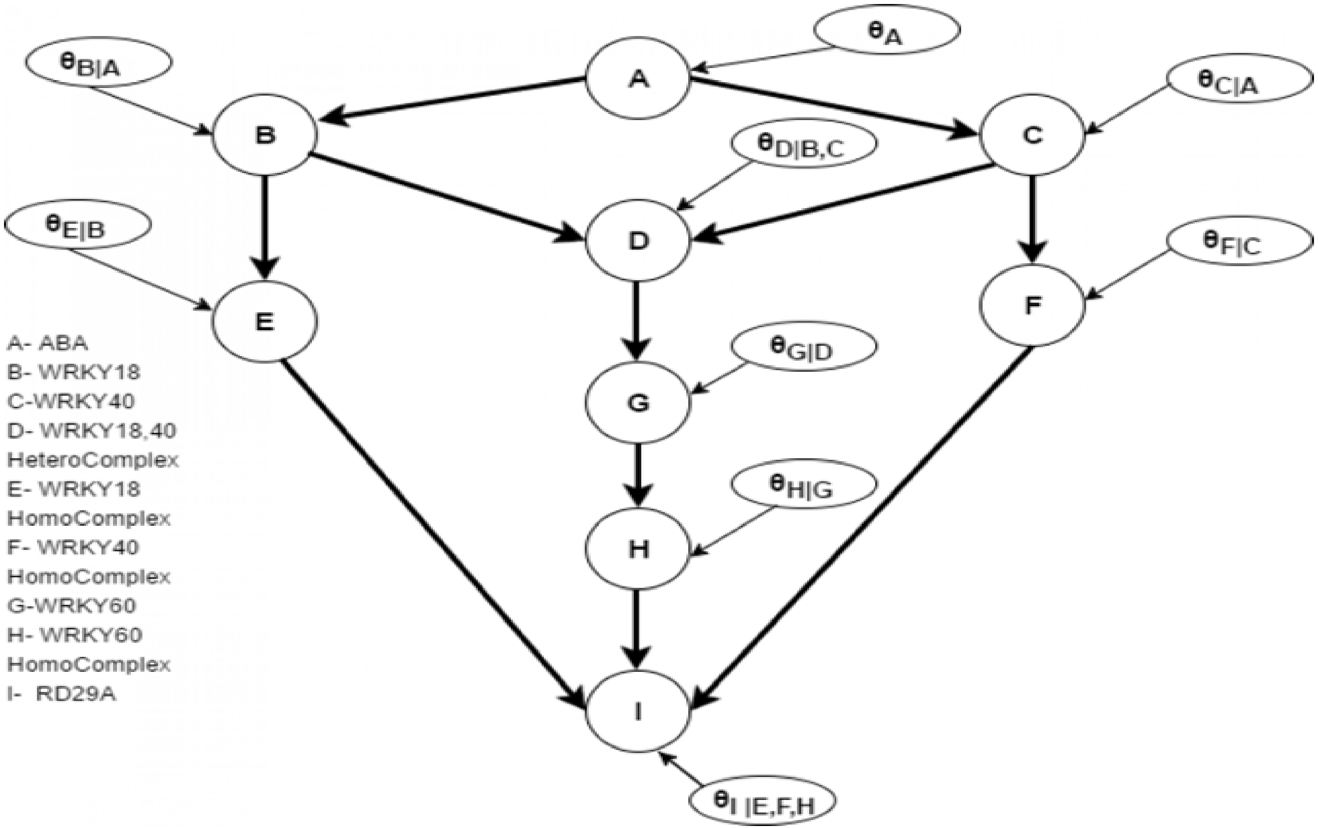
Example 1: Parameter Estimation in BN (WRKY Transcription Factor)

- Objective: Estimate the network parameters of the WRKY transcription factor (TF) network (GRN) involved in drought resistance.

- Prior knowledge:
  1. The WRKY TF network is available from pathway literature.
  2. The network structure is a DAG.

- Method:
  1. Use Bayesian approach to estimate parameters in BN.
  2. Discretize microarray data into binary with states 0 (gene inhibition) and 1 (gene activation).

Example 1 : Parameter Estimation in BN : The WRKY TF Network (DAG)



Inference for regulator identification

- Network inference in PGM: aids study of information propagation
- Allows identification and comparison of modulators
- Modeling a GRN as a PGM: identify potential trait regulators

Inference in Bayesian Networks

- Compute Probability (Reporter node in desirable state given status of upstream node)

- Find upstream node that maximizes this conditional probability

- Perform wet lab experiments for validation (or use literature)

Reference for Bayesian Modeling of Drought Resistance

Lahiri, A., Venkatasubramani, P. & Datta, A., “Bayesian modeling of plant drought resistance pathway,” BMC Plant Biol 19, 96 (2019). https://doi.org/10.1186/s12870-019-1684-3

Example 2: Lodging Resistance Network Learning with BN

- Objective: Learn the network responsible for conferring lodging resistance in crops e.g. Sorghum

- Pathway info not available

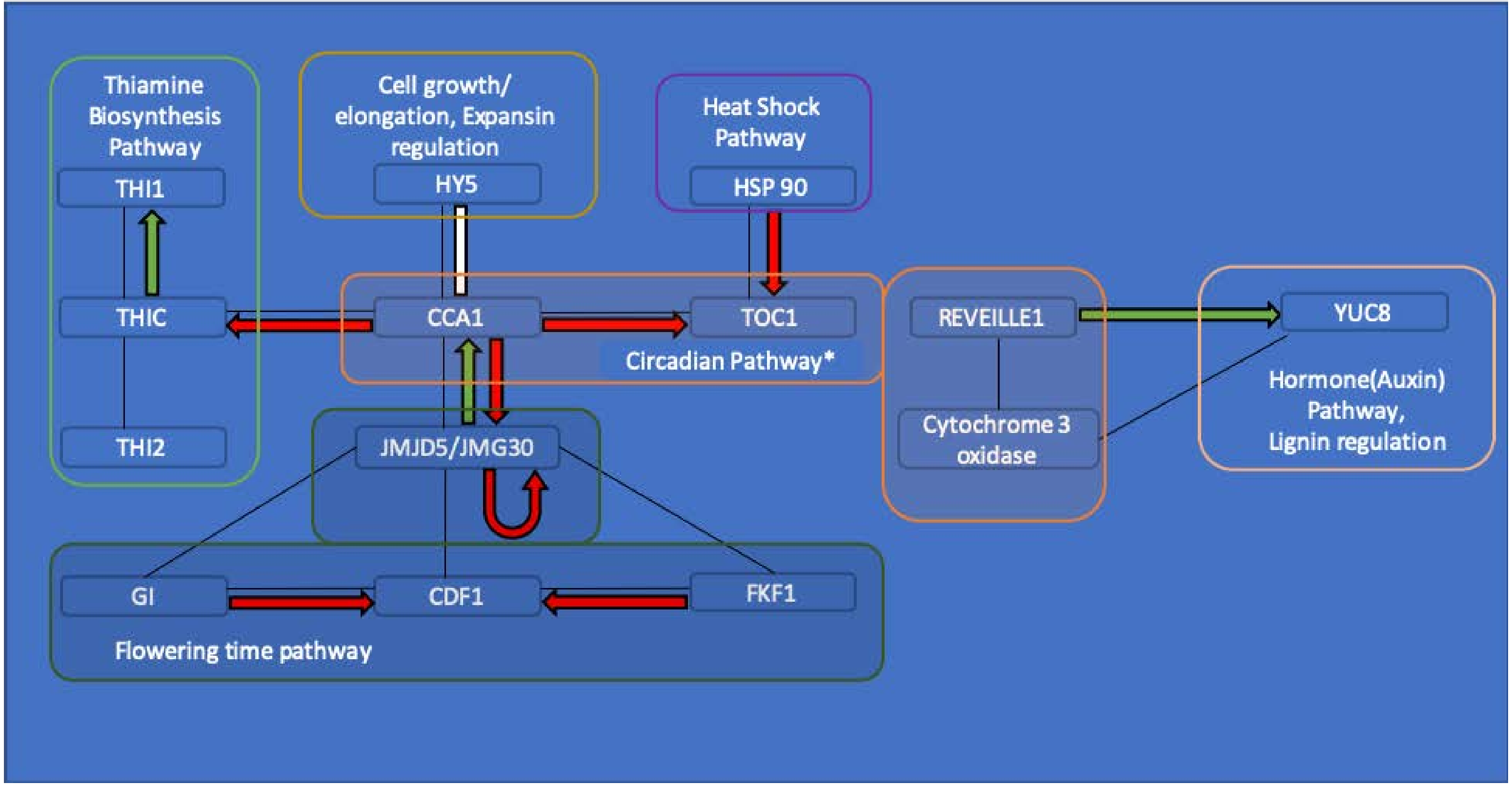
- Design: Learn Bayesian Network (a causal network) from gene expression dataset under lodging stress

- No ground truth to compare against

Example 3: Circadian Network Learning with GGM

- Use Gaussian Graphical Models (GGM)
- Learn circadian network using GGM on Sorghum data.
- Circadian gene network evolutionarily conserved among different crops
- Well characterized Arabidopsis circadian network provides basis for benchmark comparison

Example 3: Circadian Network Learning with GGM



Example 3: Circadian Network Learning with GGM

- We used GGM to create differential graphical networks: day vs night (10AM-4PM vs 10PM-4AM) and morning vs evening (4-10 AM vs 4-10 PM)

- Many **inter -pathway relationships** uncovered using GGM inference.

- Interaction between CCA1 and THIC genes:
  1. CCA1 is key gene involved in circadian pathway while THIC is a key gene involved in thiamine (vitamin B1) biosynthesis pathway.
  2. In plants, thiamin plays a role in resistance to biotic and abiotic stresses.
  3. The causal nature of the interaction is verified experimentally.

Example 3: Circadian Network Learning with GGM

- Relationship between auxin biosynthesis pathway and circadian pathway:
    1. Shown in the network as indirect relationship between circadian gene REVEILLE1 and auxin biosynthesis gene YUC8.
    2. Auxins are thought to play an important role in root organization.
    3. REVEILLE1 has been shown to directly positively regulate YUC8 gene.
    4. Although the network did not predict a direct relationship, it still uncovered a potential link between the two pathways.
- The causal interactions were verified from literature survey of Arabidopsis circadian genes.
- Similar approach to be used for learning interactions between different pathways contributing to lodging resistance by creating differential GGM using gene expression for contrasting genotypes.

Conclusion

- Network modeling can capture the multivariate stochastic interaction taking place in GRNs in plants.

- Incorporates biological prior info when available

- Learning network from data: potential to discover new causal biological relationships.

- Find trait regulators in GRNs: help breed crops with desirable traits

Acknowledgements

- Aditya Lahiri
- Rajan Kapoor
- TEES
- TAMU X-Grant Program
- All collaborators on 2018 CRISPR Gene Editing X-Grant at TAMU (M. Thomson, L. Shan, J. Mullet, B.-J. Yoon, Y. Shan, X. Qian, J. Pellois, K. Rathore, C. Johnson, M. Molina, M. Molina, P. He, J. Awika, S. Okumoto and E. Septiningsih)
- Anastasia Muliana
- Bill Rooney
- Scott Finlayson