

# Bayesian Networks: With Applications for Analyzing Causal Relationships in Rice

Jared Richardson<sup>1</sup>, Dr. Shannon R. M. Pinson<sup>2</sup>, Dr. Jeremy D. Edwards<sup>2</sup>,  
Dr. Jianzhong Su<sup>1</sup>, Dr. Yulin Jia<sup>2</sup>

<sup>1</sup>The University of Texas at Arlington  
<sup>2</sup>USDA-ARS Dale Bumpers National Rice Research Center

## Introduction

Rice sustains over half the global population but faces challenges from rising demand and environmental variability. This study employs statistical tools to analyze casual relationships between rice traits and genetic markers by:

- Modeling phenotypic data to impute and extract Best Linear Unbiased Predictors (BLUPs) for trait predictions
- Utilizing Directed Graphs (DAGs) to move from correlation to causation, enabling identification of genes influencing traits using SNP markers tagging the rice genome

These mathematical approaches enhance our understanding of genetic and environmental interactions, guiding crop improvement and breeding strategies.

## From Correlation to Causation: Pearson Graphs and Directed Acyclic Graphs (DAGs)

### Pearson Correlation Analysis:

- Measures linear associations between traits and SNP markers:

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

- Limitation: Cannot determine causality or directionality.

### Directed Acyclic Graphs (DAGs):

- Represent causal relationships as a graph  $G = (V, E)$ , where:  $V$  = variables (traits, markers);  $E$  = directed edges (causal links).
- Use conditional independence to infer causality and analyze multi-omic interactions.

### Bayesian Networks:

- Extend DAGs with probabilistic relationships:

$$P(V) = \prod_i P(V_i | \text{Parents}(V_i))$$

### Key Insights:

- Example: Correlation shows traits  $A$  and  $B$  are linked; DAGs identify causal direction  $A \rightarrow B$ .
- DAGs enhance understanding of multi-trait interactions for breeding applications.

## Linear Mixed Modeling

**Purpose:** Linear mixed models extend simple linear models to account for:

$$y = X\beta + Zu + \epsilon$$

- $X\beta$ : Fixed effects, capturing systematic influences like environmental conditions.
- $Zu$ : Random effects, modeling variation due to genetic or experimental factors.
- $\epsilon$ : Residual error, accounting for unexplained variation.

**Why Use BLUPs?** BLUPs (Best Linear Unbiased Predictors) provide:

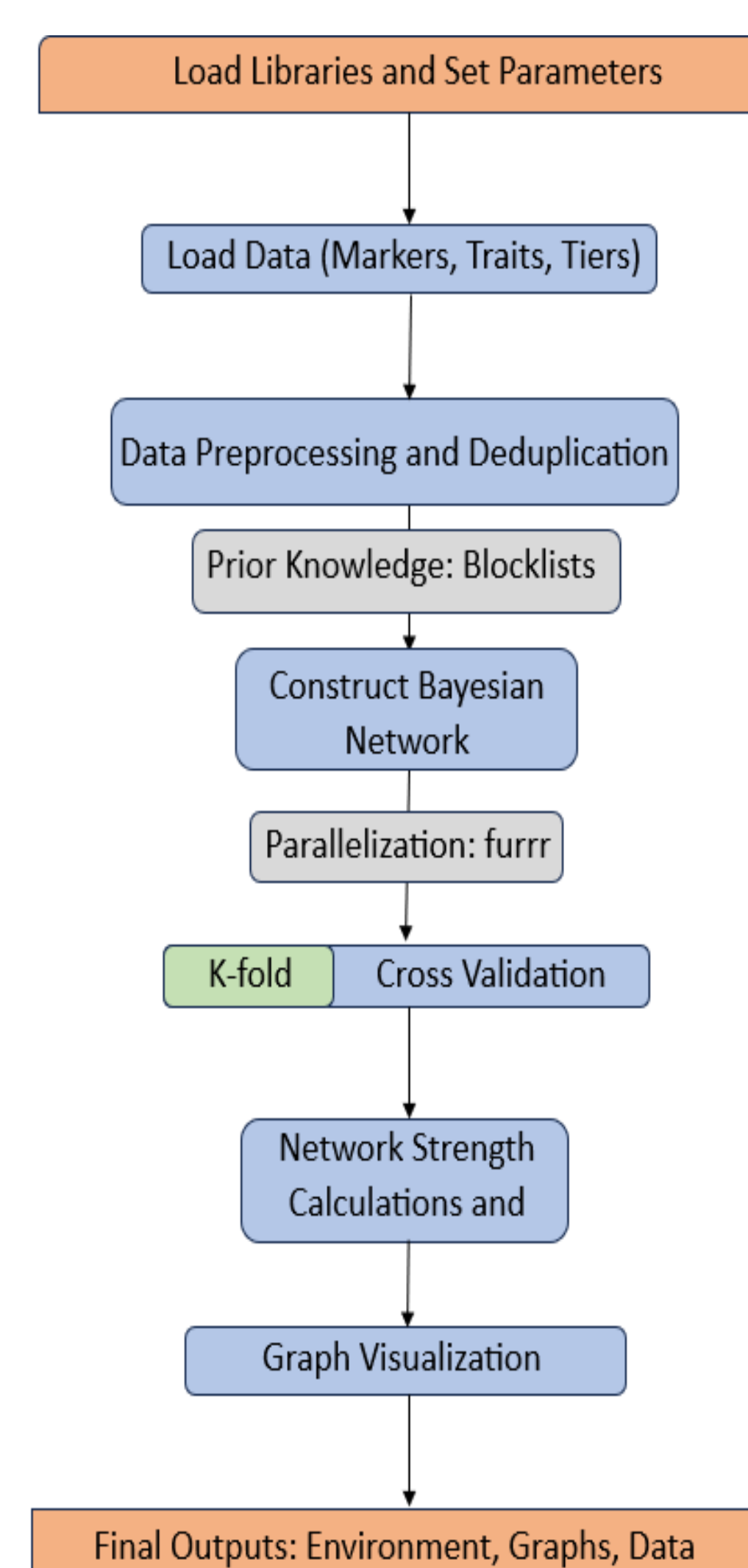
- Unbiased trait estimates by accounting for random and fixed effects simultaneously.
- Cross-environment robustness, enabling comparisons across diverse conditions.
- Environmental covariates, such as temperature and soil quality, were included in the fixed effects to isolate genetic contributions

**Implementation:** Each dataset required a unique mixed model to address its specific structure, including variations in environmental conditions and study designs. The **lme4** package in R was employed to:

- Fit models efficiently across large-scale datasets.
- Extract BLUPs for traits, serving as key inputs for subsequent DAG construction and analysis.

## Bayesian Network

- Structure Learning:** Constructs a DAG using the hill-climbing algorithm with blacklist constraints derived from tiered trait relationships.
- Parameter Estimation:** Calculates edge strengths via  $k$ -fold cross-validation, assessing conditional probabilities  $P(V_i | \text{Parents}(V_i))$ .
- Marker-Interaction Control:** Ensures markers cannot predict one another unless explicitly allowed in script.
- Edge Appearances:** Edge colors determined by:
  - Correlation thresholds ( $\rho$ ) for SNP-trait and trait-trait links.
  - Population-specific differences (e.g., Lemont vs. Teqing).
- Integration:** Combines phenotypic and SNP marker data for multi-omic causal analysis.



## Results

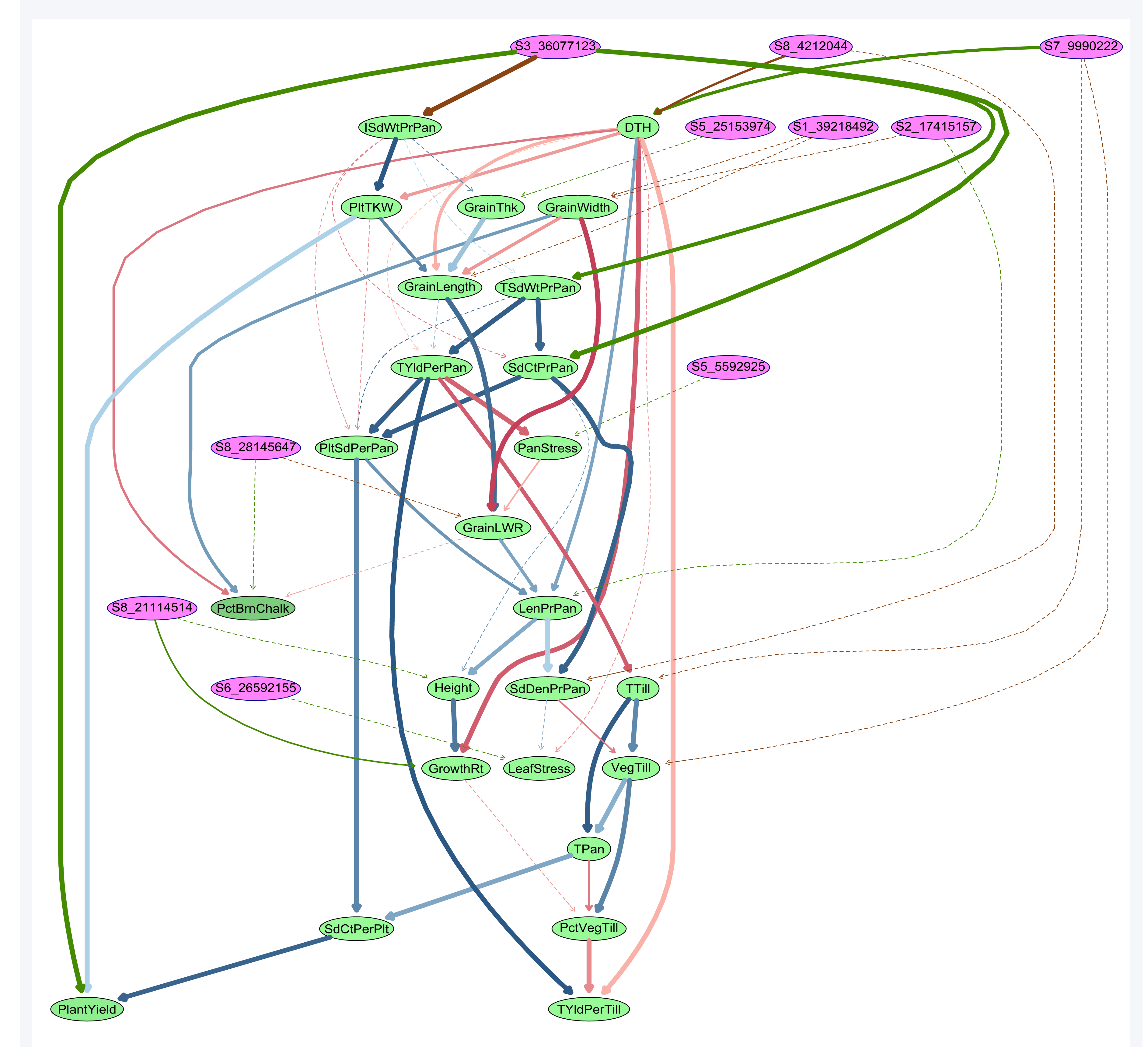


Figure 1. DAG for the Agronomic dataset with included SNP markers.

## Conclusion

Our Bayesian network model provides a robust framework for analyzing causal relationships between SNP markers and rice traits. By integrating multi-omic data and leveraging advanced structure learning, this approach enhances our understanding of trait interactions under variable environmental conditions. These insights hold potential for optimizing breeding strategies and improving crop resilience.

## References

- [1] Rajendra Prasad, Yashbir Singh Shivay, and Dinesh Kumar. *Current Status, Challenges, and Opportunities in Rice Production*, pages 1–32. Springer International Publishing, Cham, 2017.
- [2] Marco Scutari, Phil Howell, David J Balding, and Ian Mackay. Multiple Quantitative Trait Analysis Using Bayesian Networks. *Genetics*, 198(1):129–137, 09 2014.
- [3] Lorenzo Valleggi, Marco Scutari, and Federico Mattia Stefanini. Learning bayesian networks with heterogeneous agronomic data sets via mixed-effect models and hierarchical clustering. *Engineering Applications of Artificial Intelligence*, 131:107867, 2024.