

## **Predicting Heat Stress Responses in Soybean by Integrating Transcriptomic, Phosphoproteomic, and Physiological Data with Machine Learning**

Rising temperatures threaten soybean (*Glycine max*) productivity, yet the agronomic traits that matter most to growers — yield and seed composition — are only measurable at the end of a plant's life cycle, making rapid screening for thermotolerance impractical at the scale needed for crop improvement. We hypothesize that early molecular and biochemical signatures, specifically the state of protein phosphorylation prior to heat exposure, can predict a plant's longer-term physiological and agronomic response to temperature stress. To test this, we combined controlled growth chamber (Phytotron) and field (Inwood) experiments in soybean genotypes Pana and PI 567357, applying elevated temperature treatments and profiling transcriptome, phosphoproteome, leaf-level photosynthesis and respiration, and yield/seed composition across a dense time course spanning minutes to weeks after heat onset. This multi-tiered dataset allows us to track how rapid signaling events (protein phosphorylation) and transcriptional reprogramming precede and potentially forecast slower-developing physiological changes in photosynthetic capacity ( $V_{cmax}$ ,  $J_{max}$ ,  $\Phi_{PSII}$ ) and, ultimately, seed yield and quality. We are using machine learning to integrate across these molecular, biochemical, physiological, and agronomic layers, aiming to identify early predictive biomarkers of heat tolerance that could substitute for costly, time-intensive agronomic phenotyping. This approach bridges mechanistic plant biology with a scalable, data-driven framework for identifying thermotolerant germplasm, with implications for accelerating breeding efforts as climate variability increasingly threatens soybean production.