Final report (January 2025 to December 2025)

Project funded by Multiregional Soybean Checkoff Program and the United Soybean Board

Project title - Field phenotyping using machine learning tools integrated with genetic mapping to address heat and drought induced flower abortion in soybean.

Participating institutions – Texas Tech University, Kansas State University, University of Missouri, and University of Tennessee.

Goals & Objectives

Long-term Goal – Develop soybean cultivars with 20 to 30% lower flower abortion under favorable to challenging environmental conditions, leading to about 10-15% increase in yield potential.

Objectives (Year 3)

- A novel image-based machine learning tool for quantifying flower abortion with minimal to no manual counting in soybeans grown in diverse environmental conditions.
- Investigate physiological effects of drought stress on contrasting lines in both controlled and field environments to assess tolerance to adverse conditions (new activity).
- Develop recombinant inbred line (RIL) populations using contrasting (low and high) flower abortion lines identified from different environmental conditions.
- Identify key hub genes that regulate flower abortion using contrasting lines and functionally characterize using CRISPR/Cas9-mediated knockout (KO) technology.

Objective 1 - A novel image-based machine learning tool for quantifying flower abortion with minimal to no manual counting in soybeans grown in diverse environmental conditions.

Genotypes of high (CL0J17-3-6-8 and PI567638) and low (IA3023 and PI506862) flower abortion were selected from 2023/2024, plus two cultivars as checks (CTVA38, CTVA40), for all locations trials in 2025, planting happened on the following dates:

- Texas Tech University: June 2nd
- Kansas State University: June 9th
- University of Tennessee: June 3rd
- University of Missouri: June 24th (delayed by rain events)

All locations label plot using QR codes for identification (Figure 1). The video imaging was performed on both sides of the row and in the two-middle rows of each plot with four rows. Texas Tech University had an irrigation treatment in the field to simulate well-watered and drought conditions during flowering (80% and 30% Evapotranspiration). The other locations experiments were rainfed only.

Flower counts were done every three days marking new and olds flowers (Figure 2), using oil-paint permanent marker. All plants in 2 feet of one middle row were used in each plot for manual flower count.

Videos were taken from all locations using four GoPro 11 Cameras (Figure 3) positioned in different angles on the same day of the flower and pod counts. The videos are in the process of downloading to a shared flower to be used in 2026 for further analysis.

Experimental design:

- *Texas Tech University*:

6 genotypes x 2 irrigation regimes (80%ET and 30% ET), three replicates.

- *Kansas State University, University of Tennessee, and University of Missouri*:

6 genotypes, three replicates of four rows, with each row of 8 feet length.

Figure 1. Plot identified with QR code labels at Texas Tech University farm

Flower count model - Texas Tech University

Model performance across locations

Figure 2. Soybean flower classification for manual and machine learning counts.

Our current detector for soybean flowers is built on Faster R-CNN in Detectron2 with a ResNet 50 backbone and a feature pyramid network. The model was originally trained on annotated frames from a single field site and season, i.e., late-summer soybean season of 2024 at Kansas State University. On images obtained at Kansas, the model behaves reasonably well, giving a recall of about 87 percent and a precision of about 90 percent. The high precision shows that, in this source domain, the model produces very few false positives.

When the same model is deployed on videos from other locations, performance drops noticeably. This is not surprising, but it is useful to separate the two distinct failure modes. In our tests on new sites, recall was not affected very much. Missed flowers are not the dominant problem. The detector usually picks up most visible flowers, even at new locations.

The *major issue* is false positives as we move away from the training domain. The detector marks spots and holes on leaves, stems, bright soil patches, weeds, and various background textures as flowers. Soil color, plant spacing, weed pressure, and growth stage all vary from site to site. Sometimes there is more visible sky or more bare soil. All of these create new backgrounds that the model has never seen. Differences in the camera positioning also matter. Different devices have different color responses, compression, and sharpening. Lens focal length and distortion

changes the apparent shape and size of flowers and leaves. The mounting height and tilt of the camera changes the viewing angle. Hence, the detector has to recognize flowers from slightly different perspectives, often with more occlusion than in the original site.

What was done when accuracy dropped at a new location

a. Hard negative mining

Because the errors in new locations are dominated by false positives, hard negative mining is often the most practical way to improve behavior at new sites.

In our setting, a hard negative is a region of the image that the current model believes is a flower with reasonably high confidence, but that is actually background. These regions are very informative because they sit exactly at the boundary where the model is confused.

The practical procedure we followed is outlined here - We first ran the existing model on a sample of frames from the new location, using the same confidence threshold that we use at the training site. We saved all detections as crops, with metadata that records which video, frame, and bounding box each crop came from. These crops were clustered into clear true positives and clear false positives. Ambiguous cases were left to deal with later. The clear false positives form a bank of hard negatives.

Figure 3. Soybean flower classification for manual and machine learning counts.

The simplest and most direct approach was then to go back to the original frames that produced these crops. The frames were annotated based on clustered true positives and bounding boxes were removed for the false positives. Frames that were pure background from the model point of view was left to remain as background only.

We then fine tuned the detector using these corrected frames from the new location, with a low learning rate. This teaches the network that these specific structures are not flowers, without erasing what the model already knows. As the annotation is done automatically based on the unsupervised clustering, the human effort is manageable.

Alternatively, we can construct images that contain only background content. For example, we can tile the hard negative crops on a blank canvas to create synthetic images that consist entirely of things the model tends to confuse with flowers. These images are then included in training with labels that tell the network there are no objects present. This is a classic hard negative mining idea that fits naturally in Faster R-CNN, which already learns to separate background from object regions.

b. Domain adaptation

In parallel with this hard negative mining and clustering work, we have also explored more formal domain adaptation. In particular, we tested a scale aware domain adaptive variant of Faster R-CNN with separate domain classifiers at the image and instance level and additional losses for scale specific alignment.

These experiments are useful and remain a candidate for future development, especially if we want a single model that works across many sites with minimal manual work. So far, however, they have not given clear gains in this soybean project.

The method is natural in settings where the main change is in appearance of the same underlying classes. For example, a detector trained on cars in clear weather can be adapted to cars in fog or rain, where the objects are the same but the context of the scene has shifted. In our case the situation is different. New locations introduce entirely new structures that were not present in the source domain, and many of these structures become systematic false positives. For that kind of error, teaching the model about hard negatives has turned out to be a more direct and predictable tool than the domain adaptation approach we tested. In 2026, we will work on fine-tune the model to improve performance for different locations.

Pod count model - Kansas State University

Model development for pod detection and tracking using the 2024 imagery data

High-throughput plant phenotyping remains a major bottleneck in modern crop improvement programs. While genotyping has advanced rapidly, field-based phenotyping particularly for yield-related traits relies heavily on manual, labor-intensive, and subjective measurement processes. Pod number in soybean (*Glycine max*), a trait strongly associated with yield potential, flower abortion, and reproductive development, is especially challenging to measure efficiently at scale.

Computer vision and deep learning approaches have proven effective in controlled environments, but their performance substantially declines under real-world field conditions due to occlusion, cluttered backgrounds, and lighting variation. Existing soybean pod counting research focuses mainly on static images, lacking support for temporal analysis in videos.

To bridge these gaps, in this project, we developed SoybeanPod-MOTS, the first multi-object tracking and segmentation (MOTS) dataset for soybean pods captured directly in field conditions across multiple locations, genotypes, and reproductive stages. The dataset supports pod detection, segmentation, identity-level tracking, and non-destructive pod counting.

Using the SoybeanPod-MOTS dataset, we evaluated several modern, state-of-the-art object detection models and tracking-by-detection algorithms to identify and track soybean pods across diverse field conditions. The primary goal was to determine which combinations of detectors and trackers produce the most reliable counts and identity associations in videos that vary widely in lighting, background contrast, pod color, and crop stress.

Specifically, three modern versions of the YOLO family were assessed: YOLOv8, YOLOv9, and YOLOv11. These models were selected due to their real-time performance and strong accuracy on agricultural imagery. Each model was tested for both bounding-box detection and segmentation to understand how well they localize pods with irregular shapes or overlapping structures.

Across all evaluation metrics, YOLOv8 consistently outperformed YOLOv9 and YOLOv11. It delivered the highest accuracy in both detection and segmentation tasks. YOLOv8 achieved an AP@50 of 79.30 and a mean Average Precision (mAP) of 69.50 for detection. In the segmentation task, which is more challenging due to pod size variation and cluttered backgrounds, YOLOv8 recorded an AP@50 of 54.50. These results reflect YOLOv8's balanced capacity for feature extraction, generalization across genotypes, and robustness in videos with variability in pod color

and environmental conditions. Although YOLOv9 and YOLOv11 demonstrated competitive performance in some scenarios, their overall accuracy was more sensitive to moisture stress, lighting inconsistencies, and background interference.

In terms of tracking, five tracking-by-detection algorithms were tested: BoostTrack, BoT-SORT, StrongSORT, ByteTrack, and OC-SORT. These trackers represent diverse approaches in appearance modeling, motion prediction, and detection association strategies. The results of these tracking approaches in terms of standard tracking metrics are shown in table 1. The up and down arrows indicate if higher or smaller values are better.

Table 1. Results of the tracking model for pod count.

Approach	IDP ↑	IDR ↑	IDF1 ↑	MOTA ↑	HOTA ↑	IDSW ↓
BoostTrack	68.58	25.603	37.286	17.786	28.861	**717**
BoT-SORT	57.53	35.875	44.192	16.145	34.376	1175
StrongSORT	85.964	**76.313**	**80.852**	**77.607**	**65.646**	2546
ByteTrack	**86.927**	74.884	80.457	76.299	64.092	2701
OC-SORT	86.905	72.007	78.758	75.082	63.782	2123

The best overall tracking performance was achieved by StrongSORT, which produced the highest identity-focused metrics: IDF1 = 80.85, MOTA = 77.61, and HOTA = 65.65. StrongSORT integrates appearance embeddings with robust motion modeling, allowing it to maintain pod identities effectively even under occlusions, crossing trajectories, and rapid camera movement. This makes it particularly suitable for field conditions where pods frequently overlap or sway due to wind.

In terms of precision, ByteTrack was the strongest performer, achieving an IDP of 86.93, along with competitive MOTA and IDF1 values. Its simple yet effective association strategy excels when detections are clean and consistent. ByteTrack offers lightweight computation with strong precision, making it well-suited for large-scale processing where speed is essential.

OC-SORT delivered balanced results across identity, precision, and robustness metrics. Its motion-centric updates help maintain performance when pod movement is irregular or when appearance cues are less reliable due to pod color similarity.

In summary, YOLOv8 paired with StrongSORT provided the most reliable combination for pod detection and tracking across varied field environments. YOLOv8 offered the strongest detection foundation, while StrongSORT provided stable identity tracking under challenging conditions. However, ByteTrack remained a strong alternative when computational efficiency and high precision are priorities.

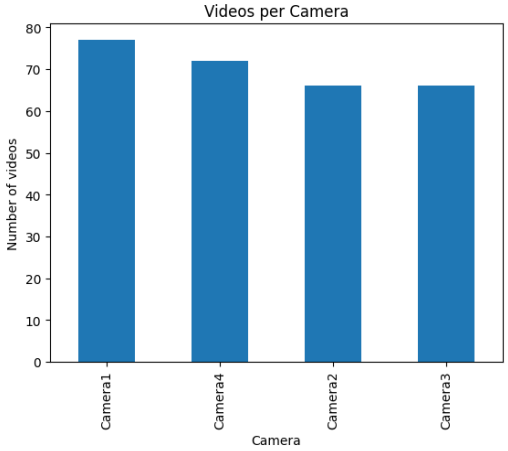

Figure 4. Summary of 2025 videos collected for preliminary results.

Preliminary results of the current models using 2025 imagery data

Using the best models trained on the 2024 data, we have carried out some preliminary data analysis using the imagery from year 2025. The following graph (Figure 4) provides a summary of the 2025 video dataset by camera (note: four cameras were used in 2025).

We tracked the videos captured with the four cameras independently. The following graph shows a boxplot summary of the number of pods tracked in the video collection, by camera. As shown, cameras 1 and 3 generally capture a larger number of pods as compared to cameras 2 and 4. Aggregation of the counts per plant will be performed in the near future, as described in more detailed below, and the resulting counts will be compared with ground truth counts per plant as collected manually in the field.

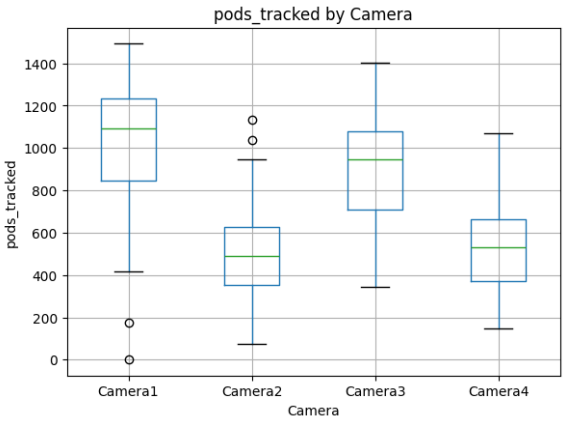

Figure 5. Boxplot summary of number of pods.

A sample pod-annotated video is linked below.

Limitations and Planned Improvements for 2026

A key limitation of the current phenotyping system is the restricted use of video data. In 2024, the data collection platforms recorded four simultaneous camera views, but the present models rely on only one of these views, and imagery was captured from only one side of each row. Consequently, the dataset used for model training and evaluation represents only a fraction of the true pod distribution in the field, leading to incomplete total pod counts.

To overcome this limitation, we plan to incorporate all four camera views into the processing pipeline, enabling reconstruction of a more complete view of each plant side. This multi-view integration is expected to reduce occlusions and improve the accuracy of pod detection and tracking.

Furthermore, the 2025 data collection efforts included imaging from both sides of the plants, providing substantially richer coverage. By merging the 2024 single-side videos with the 2025 dual-side dataset, we aim to enhance the robustness and completeness of pod count estimation for the 2026 system.

However, these improvements introduce additional technical challenges. Chief among them is the need to correctly associate pods across multiple camera views—either from different cameras on the same side or from opposite sides of the plant—so that each pod is linked to a consistent identity when appropriate. Addressing these alignments and correspondence issues will be essential for achieving reliable total pod counts in the next iteration of the phenotyping pipeline.

Objective 2 - Investigate the physiological effects of drought stress on contrasting lines for flower abortion under controlled conditions.

This study aimed to assess the flowers abortion among six soybean lines including PI506862, PI567638, IA3023, CL0J7-3-6-8, PI548318, and PI80837 under progressive water-deficit stress (dry-down phase) and subsequent re-watering recovery (recovery phase) in a greenhouse setting at the West Tennessee Research and Education Center, University of Tennessee. On April 10, 2025, seeds were sown in a 1:1 mixture of sand and Lexington silt loam at a 2 cm depth, then thinned to one plant per pot at 13 days after planting (DAP). Fertilizers were applied at 12 DAP (0.075% V/V liquid, 0–10–10) and 24 DAP (0.06% W/V water-soluble, 24–8–16). Plants were maintained under a 14-hour light/10-hour dark cycle and received 200 mL of water daily during the pre-treatment phase. At 28 DAP, when plants had 4–5 trifoliate leaves, the dry-down (DD) phase began; pots were saturated, drained to their pot capacity, enclosed in 15-L plastic bags (Figure 6) to eliminate the evaporation, and fitted with watering tubes for controlled watering and monitoring plants transpiration/water loss. Based on daily transpiration rate (TR), four pots per genotype were designated as well-watered (WW)/control, and six as drought treatment (DD). The DD plants were watered only if TR exceeded 80 g/day, following Shekoofa et al. (2013). Stress progression was monitored using normalized transpiration rate (NTR), with NTR <0.10 indicating the endpoint of available soil water. Recovery occurred on days 36–41 after (depending on genotype) planting by re-watering DD pots with 350 mL to pot capacity. Data were collected daily including number of flowers, flower rate per day, and wilting score (0–5) during stress and recovery phases. The number of nodes and pods was recorded at the end of the dry-down phase and again at experiment termination (June 18, 2025). Data have been processed and interpreted, and the graphs along with key outcomes are incorporated in the current report.

Figure 6. The University of Tennessee, 2025 soybean flowers abortion's greenhouse setup at WTREC. Top: A week after planting and below: Later after the drought treatment was applied. Photo credit: The Shekoofa's lab

Results

Wilting score

Wilting response varied among the six soybean lines during the dry-down and recovery phases (Figure 7). PI548318 and PI567638 exhibited the highest wilting scores during both phases, indicating greater susceptibility to water-deficit stress. In contrast, IA3023 and CL0J7-3-6-8 maintained lower wilting scores, suggesting better stress tolerance and faster recovery after re-watering. PI506862 and PI80837 showed mild wilting under both phases, reflecting moderate stress adaptation.

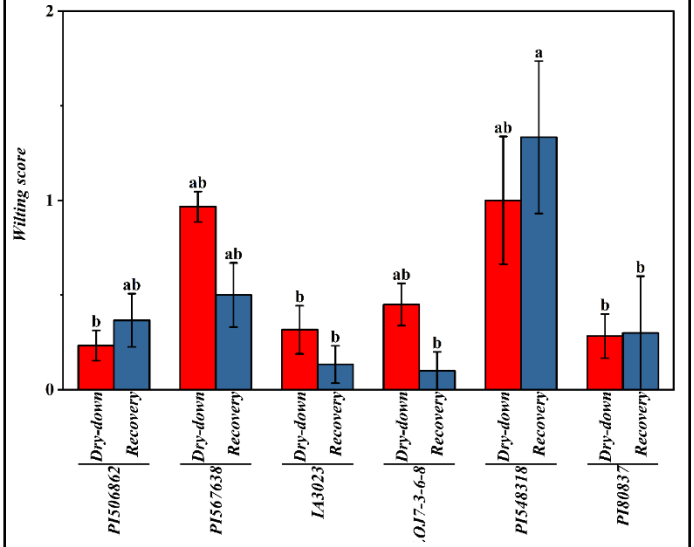

Figure 7. Wilting score of six soybean lines under dry-down and recovery phases in 2025 greenhouse experiment.

Number of nodes

Node number differed slightly among the soybean lines during dry-down and recovery phases (Figure 8). Most genotypes showed a slight increase in node formation after re-watering. CL0J7-3-6-8 produced the highest number of nodes (≈10–11) after recovery, followed by PI548318, IA3023, and PI506862, which maintained moderate node development (7–9). PI80837 showed relatively fewer nodes (5–6) across both phases, while PI567638 had the lowest node number during recovery, indicating limited vegetative recovery potential after water deficit stress.

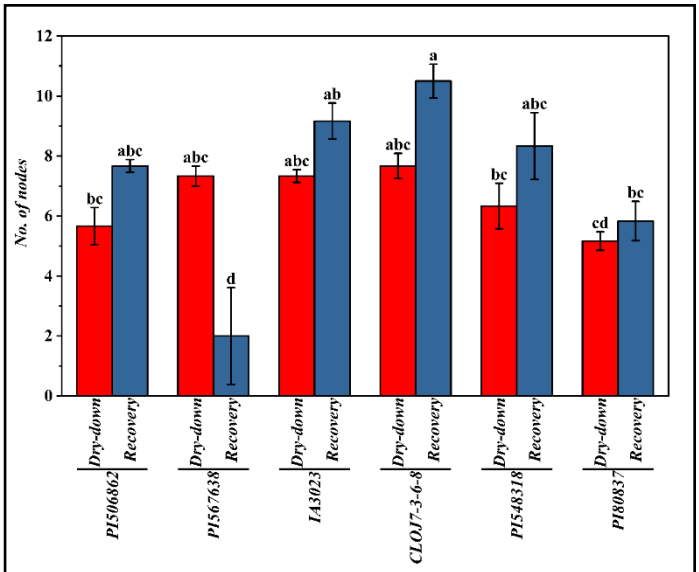

Figure 8. Number of nodes of six soybean lines under dry-down and recovery phases in 2025 greenhouse experiment.

Flower counts

Flower production varied notably among the six soybean lines under dry-down and recovery phases (Figure 9). PI506862 produced a moderate number of flowers (10–20) during the dry-down phase and did not show much recovery after re-watering. In contrast, PI567638 exhibited very few flowers (5–10) during dry-down but produced the highest flower count (120–130) after recovery, reflecting a delayed reproductive response. CL0J7-3-6-8 also showed a strong post-stress

recovery (≈80 flowers). IA3023 and PI548318 maintained moderate flowering during both phases, while PI80837 had consistently low flower production under both stress and recovery conditions.

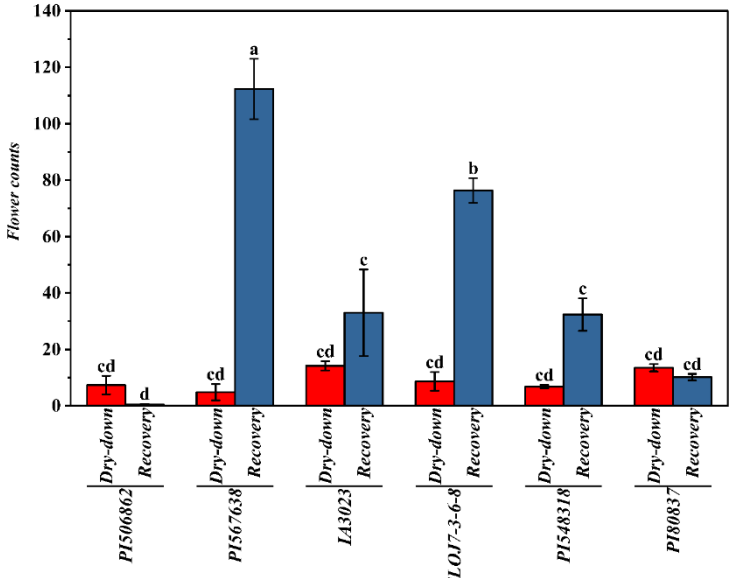

Figure 9. Flower counts of six soybean lines under dry-down and recovery phases in 2025 greenhouse experiment.

Number of pods

Pod production in 2025 differed notably among the six soybean lines (Figure 10). PI548318 and CL0J7-3-6-8 produced the highest number of pods (≈30–32 and 25–27 pods plant^{-1}, respectively), indicating strong pod set and recovery capacity after drought stress followed by PI506862. PI506862 maintained moderate pod production (≈22–24 pods plant^{-1}), while IA3023 showed comparable but slightly lower pod counts. In contrast, PI80837 produced fewer pods (~17–18 pods plant^{-1}), and PI567638, despite vigorous flowering, failed to form pods due to severe flower abortion and poor reproductive recovery under drought stress.

Objective 3 - Develop recombinant inbred line (RIL) populations using contrasting (low and high) flower abortion lines identified from different environmental conditions.

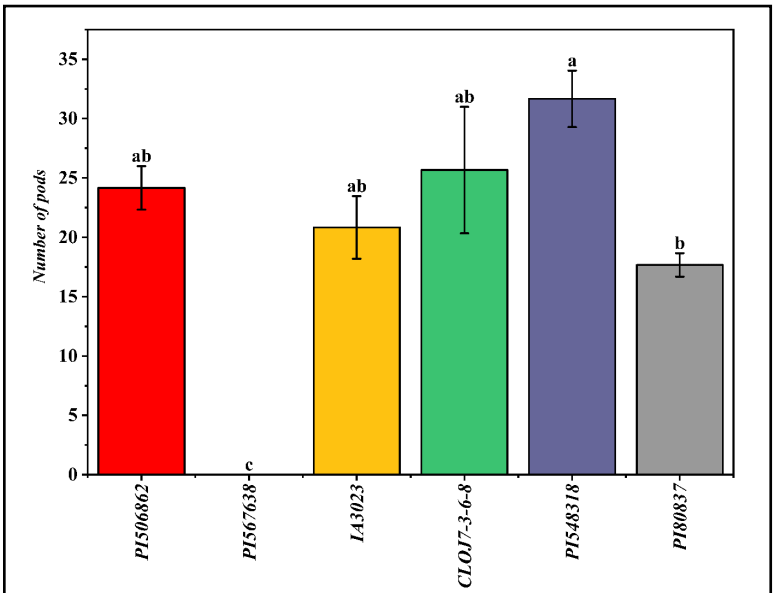

Figure 10. Number of pods among six soybean lines in 2025 greenhouse experiment under drought stress.

Three genetically diverse populations (PI 552538 x PI 556511, IA3023 x PI552538, and HS6-3976 x PI 556511) were created based on performance during the first two years of Phase 1 evaluations. These populations are currently being advanced in the Puerto Rico winter nursery. Additionally, a fourth F1 population (PI 506862 x PI567638, lines used in Obj. 4) will be grown this winter (2025) in Kansas State University greenhouse. Our objective is to obtain F5 progeny by spring 2026, enabling seed increase during the summer of 2026 and coordinated multi-location field evaluations of F4:5 progeny in 2027 across Missouri, Kansas, and Texas.

To expand the genetic resources available for mapping traits related to flower retention and yield potential, we are also initiating the development of four new recombinant inbred line (RIL)

populations (PI437578 x PI643395, PI437578 x PI417479, PI643395 x LG05-4317, LG05-4317 x PI417479, lines selected from Texas under stress condition). The parental lines—selected for strong contrasts in flower-abortion rate based on three years of field and greenhouse phenotyping—have been planted in the Kansas State University fall greenhouse to create four additional populations. Controlled crosses to produce F1 progeny will be conducted in early winter, after which these populations will be advanced rapidly via single-seed descent toward the F5–F6 generations for future evaluation. Table 2 shows the lines selected according to contrasting high and low flower abortion rates collected from project's phase 1 and Table 3 show the eight populations to be advanced to F5-F6 until 2027-2028.

Table 2. Lines selected to develop recombinant inbred line (RIL) populations from three years of flower abortion data collection.

Lines ID	Abortion Level	Growth habit	Maturity Group
PI552538	High	Indeterminate	III
HS6-3976	High	Indeterminate	III
IA3023	Low	Indeterminate	III
PI556511	Low	Indeterminate	III
PI567638	High	Indeterminate	IV
PI506862	Low	Determinate	IV
PI437578	High	Indeterminate	III
LG05-4317	High	Indeterminate	IV
PI417479	Low	Determinate	IV
PI643395	Low	Indeterminate	III

Table 3. Parental Crosses used to develop eight RIL populations.

Female	Male	Flower Color	
		Parents	F1
PI437578	PI643395	W x W	W
PI437578	PI417479	W x P	P
PI643395	LG05-4317	W x W	W
LG05-4317	PI417479	W x P	P
PI552538	PI556511	P x P	P
IA3023	PI552538	W x P	P
HS6-3976	PI556511	W x P	P
PI506862	PI567638	W x W	W

Objective 4 - Identify key hub genes that regulate flower abortion using contrasting lines and functionally characterize using CRISPR/Cas9-mediated knockout (KO) technology.

To elucidate the molecular determinants of flower abortion, we performed a comparative transcriptomic analysis of high and low abortion soybean genotypes across four reproductive stages - bud, closed flower, open flower, dry flower, and the leaf as vegetative tissue for reference. Differential expression and functional enrichment analyses revealed distinct transcriptional landscapes between high and low abortion plants, particularly in genes associated with hormonal

signaling, abscission regulation, cell wall remodelling, and reproductive failure (Figure 11). The high abortion plants showed an elevated expression of ethylene responsive and abscission-associated genes, including *Inflorescence Deficient in Abscission* (*IDA)*, *Inflorescence Deficient in Abscission-Like* (*IDL)*, *Homeobox Protein Knotted-1-Like 6* (*KNAT6)* and *Arabidopsis thaliana Homeobox Gene 1* (*ATH1/BEL-1)*, along with increased activity of pectin-degrading proteins. In contrast, low abortion plants exhibited upregulation of *Mitogen-Activated Protein Kinase Phosphatase 1* (*MKP1)* and *KNAT1*, consistent with repression of MAPK activity and maintenance of boundary integrity.

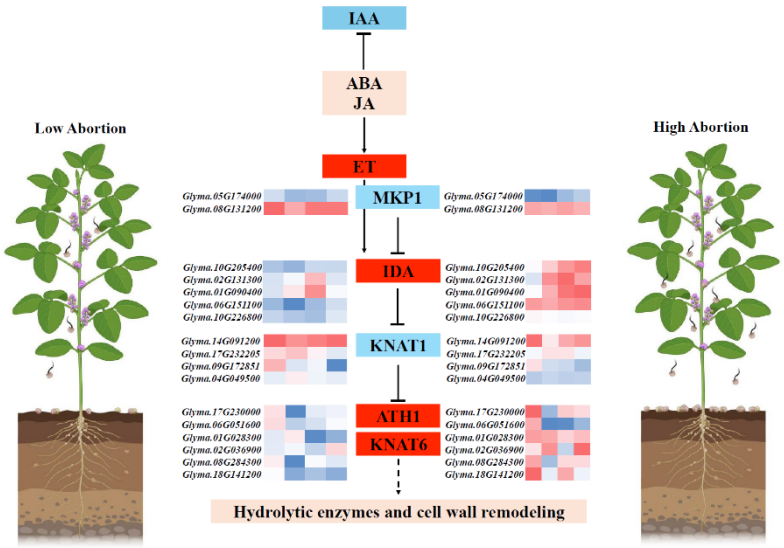

Figure 11. Proposed regulatory model underlying differential flower abortion in soybean genotypes. Schematic representation depicting the hormonal and genetic regulation associated with floral abscission. The left panel represents the low flower abortion (LA) genotype, while the right panel represents the high flower abortion (HA) genotype. Differential expression of hormone-related pathways, auxin (IAA), abscisic acid (ABA), jasmonic acid (JA), and ethylene (ET) and key regulators including MKP1, IDA, KNAT1, ATH1, and KNAT6 is illustrated with corresponding heatmap insets (red: high; blue: low). In the HA genotype, enhanced ethylene and abscission-related gene activity promotes floral organ separation and increased abortion, whereas the LA genotype maintains higher auxin signaling and suppressed ethylene responses, favoring flower retention. Downstream activation of hydrolytic enzymes and cell wall remodeling genes mediates organ detachment at the abscission zone, contributing to the observed phenotypic differences between genotypes.

The final phase of floral organ separation in *Arabidopsis* is governed by a conserved signaling network that integrates the *IDA-HAE/HSL2-MAPK* module with downstream cell wall modifying enzymes. In our study, several orthologs representing this core abscission module and its regulators were differentially expressed between the HA (High abortion) and LA (Low Abortion) genotypes, suggesting their potential roles in modulating floral retention versus shedding. The MAP kinase phosphatase 1 (*MKP1*) acts as a key negative regulator establishing a signaling threshold prior to activation of abscission cascade. Functionally, *MKP1* dephosphorylates *MPK3/6* thereby dissipating basal MAPK activity that might otherwise trigger premature cell wall hydrolysis and organ detachment. In our dataset, one of the soybean orthologs (*Glyma.08G131200*) of *AtMKP1* was upregulated and had

comparatively higher transcripts in LA, preventing premature activation of abscission responses, thereby enhancing floral retention.

Inflorescence Deficient in Abscission (*IDA*) and its close homologs, the IDA-like (*IDL*) peptides bind to a receptor complex on the plasma membrane formed by the leucine-rich repeat receptor-like kinases (*LRR-RLKs*) HAESA (*HAE*) and HAESA-like2 (*HSL2*). They in turn heterodimerize with co-receptors from the somatic embryogenesis receptor-like kinase (*SERK*) family to form an active receptor complex. IDA binding activates the receptor complex which triggers a MAP kinase cascade, leading to the transcriptional induction of cell wall hydrolase genes. In the current study, IDA and IDL peptides were upregulated in the HA genotype, suggesting that these plants may activate abscission signaling earlier or more strongly, priming the abscission zone for separation. The LA genotype, on the other hand, had lower *IDA* and *IDL* expression and higher levels of *MKP1*, a negative MAPK regulator, consistent with attenuated pathway activation. Even though *HAE/HSL2* receptors were not differentially expressed in the floral tissues, the elevated *IDA* levels in HA likely drive stronger or prolonged activation of this pathway in abscission-zone cells, leading to enhanced cell wall hydrolysis and flower drop.

Beyond the MAPK cascade and *IDA-HAE/HSL2* signaling core, abscission is tightly coordinated with the developmental patterning machinery that defines organ boundaries and tissue fate. In Arabidopsis, class I KNOTTED-like homeobox (*KNOX*) genes such as *KNAT1* and *KNAT6*, together with the BELL-family transcription factor *ATH1*, constitute a trans criptional axis that regulates boundary specification and spatial restriction of abscission competence. In our soybean dataset, we observed a distinct expression pattern of these homeobox genes between the HA and LA genotypes. *KNAT1* transcripts were elevated in the LA genotype, aligning with its known role in maintaining tissue integrity and repressing cell separation. In contrast, *KNAT6* and *ATH1* genes were upregulated in the HA genotype, suggesting a transcriptional shift toward activation of AZ differentiation. *KNAT6* functions in meristem maintenance but is spatially restricted by *KNAT1/PNY* during reproductive development. Its expression promotes boundary expansion and cell separation, and the removal of *KNAT6* rescues the *knat1* mutant phenotypes (shortened internodes and downward pointing siliques) in *Arabidopsis*. Moreover, *KNAT6* acts downstream of the *IDA-HAE/HSL2* cascade, linking peptide-mediated signaling to transcriptional activation of genes involved in wall hydrolysis and remodeling. Elevated *KNAT6* expression in HA soybean flowers may thus enhance boundary competence and sensitize cells in the floral base to separation cues. Similarly, the *ATH1* homeobox gene, which acts in concert with *KNAT1* and *KNAT6* to define floral organ boundaries, was more strongly expressed in HA. High *ATH1* expression in the HA genotype likely reinforces cell fate transitions within the abscission zone, supporting the upregulation of cell wall modifying genes observed in HA.

Taken together, these patterns indicate that in soybean, the differential regulation of *KNAT1*, *KNAT6* and *ATH1* establishes opposing transcriptional states that influence floral retention. The LA genotype maintains a KNAT1 like regime that preserves structural integrity and suppresses separation, while the HA genotype transitions toward a *KNAT6/ATH1* dominated state that primes abscission-zone differentiation and accelerates the separation process. This interplay mirrors the antagonistic relationships characterized in tomato and *Arabidopsis* and underscores a conserved developmental logic linking meristem-derived patterning genes to abscission signaling.

Major outputs for 2025

Phase 1 of this project has yielded significant progress towards addressing the long-term challenge of flower abortion and our ability to fully capture soybean's yield potential. The outcomes have resulted in **five manuscript submissions**, which are currently under review or revision and listed below:

Paper 1: Plasticity in flower number and abortion shape soybean (*Glycine max* (L.) Merr.) yield under different environmental stresses. (Revised version submitted: ***Journal of Agronomy and Crop Science***)

Outcome: Distinct flowering plasticity strategies were detected, with LG05-4317 and PI506862 identified as promising candidates for breeding high-yielding cultivars with optimized flower abortion.

Paper 2: Automated flower counting method for in-field soybean phenotyping. (Under review: ***Computers and Electronics in Agriculture***)

Outcome: A scalable field-based phenotyping system that combined a mobile imaging platform achieved 85% average precision in distinguishing 'new' and 'old' flowers and captured genotype-specific flowering dynamics, providing a foundation for estimating flower abortion rates.

Paper 3: Soybean Pod-MOTS: A video dataset and benchmark for soybean pod detection, segmentation, tracking and counting in field videos. (Under review: ***Smart Agricultural Technology***)

Outcome: The first multi-object tracking and segmentation (MOTS) from field videos, covering 637,520 pod instances in six genotypes and multiple environments, with 80% accuracy established a benchmark for scalable, automated pod phenotyping, and flower abortion analysis.

Paper 4: A contrasting transcriptomic atlas of high and low flower abortion soybean genotypes reveals coordinated regulation of abscission and reproductive failure (submission: ***Plant Cell Reports***)

Outcome: Hormone signaling, abscission regulation, and cell wall remodeling were key pathways that influenced floral retention. High-abortion lines had elevated expression of ethylene-responsive and abscission-related genes (e.g., IDA, IDL, KNAT6, ATH1), whereas low-abortion lines upregulated MKP1 and KNAT1, suggesting suppression of MAPK activity and maintenance of boundary integrity.

Paper 5: Drought-induced flower abortion and recovery dynamics in soybean: Insights from a controlled greenhouse experiment. (Under review: ***Journal of Agronomy and Crop Science***)

Outcome: PI506862 was identified as a superior line with enhanced drought tolerance and reproductive resilience, making it a valuable candidate for improving flower and pod retention in breeding programs.