Project Annual Report (Jan 1, 2023, to December 31, 2023)

Project funded by North Central Soybean Research Program

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

Summary

Major Accomplishments (Jan 1, 2023 to Dec 31.2023)

- Flower abortion was rated in a diverse panel of soybean lines grown under field conditions in four different soil, climatic, management and environmental conditons.
- Successful in transition the imaging approach from greenhouse conditions and from static images to videos, *which are both unique and novel in the soybean research domain*
- A variety of platforms were developed for imaging in the field to determine the best approach for phenotyping flower and pod counts
- Developed and trained effective deep-learning (machine learning) models for temporal and spatial tracking soybean flower and pods in plants grown under field conditions
- Key genes were shortlisted, conducted gene-based haplotype analysis, and identified significant haploblocks, to help understand the genetic factors influencing flower abortion in soybeans.

Challenges

- Manual flower counting over multiple times (every 4 days) on large diversity panel posed limitations due to the need for a large workforce at each location
- Detecting and counting of flowers posed a challenge due to the small flower size and the soybean plant architecture, further complicated by different levels of lodging across locations
- The imaging platform required customization based on specific location conditions and available resources.
- Image storage is a limitation, with most videos housed by respective locations. In year 2, we hope to address this for long term storage of videos.
- Training the model for accurate flower and pod counting throughout the flowering period after addressing different levels of occlusions presents another hurdle for discern precise flower abortion patterns.
- Lack of contrasting lines with differential level of flower abortion under field conditions has limited our capacity to identify key transcripts and nodal genes that control flower abortion in soybeans

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 1)

1. Explore the genetic diversity in flower abortion under different soil moisture and climatic conditions using a large diversity panel.
2. Develop an image-based field phenotyping system and deep-learning tool to precisely document temporal dynamics in flower abortion and pod retention in genetically diverse soybeans.
3. Discover environmentally stable and region-specific genomic regions controlling flower abortion in diverse soil types, moisture, and climatic conditions.

Objective 1 - Explore the genetic diversity in flower abortion under different soil moisture and climatic conditions using a large diversity panel.

A diverse set of 350 soybean lines were seed increased in winter nursery at Costa Rica, with 310 lines from the USDA soybean germplasm collection showing ideal germination and plant stand in the seed multiplication field (Figure 1). Emphasizing genetic diversity within maturity groups III and IV, 228 lines were selected for distribution across four experimental locations: Texas Tech University, University of Missouri, Kansas State University, and University of Tennessee.

Figure 1. Winter nursery seed increase for the project materials at Costa Rica.

Each experimental site covered approximately three acres; Texas Tech University utilized sub-surface drip irrigation (SDI), while the remaining locations operated under rain-fed conditions. *Planting dates varied*: Texas Tech University planted the crop on June 16th, University of Missouri on May 24th, Kansas State University on May 25th, and University of Tennessee on June 7th. Soybean crop management followed location-specific recommendations. At Texas Tech University, cameras (Go Pro Hero 11) mounted on a tractor (Figure 2A) facilitated imaging for assessing image quality from different angles. As soybean plants entered the reproductive stage, manual flower counting and imaging was performed in the 228 lines. Systematic adjustments to camera angles, lens types, and numbers were made on the tractor to optimize imaging at that time. Approaching harvest, a final round of imaging on dried plants was conducted focusing on pods to refine the machine learning model for pod counting, subsequently, manual harvesting of 3 feet per genotype per rep was collected for plot yield data. To ensure representative data, the tagged plant used for flower counting, along with three additional plants per plot, were selected for other yield-related parameters.

The University of Missouri, Kansas State University, and the University of Tennessee encountered challenges in securing adequate work force for temporal flower counting across the 228 lines. Consequently, a core set of 30 lines was selected based on genetic diversity for manual flower counting and imaging. The manual flower counting protocol aligned with the procedures established at Texas Tech and was consistently applied across all locations.

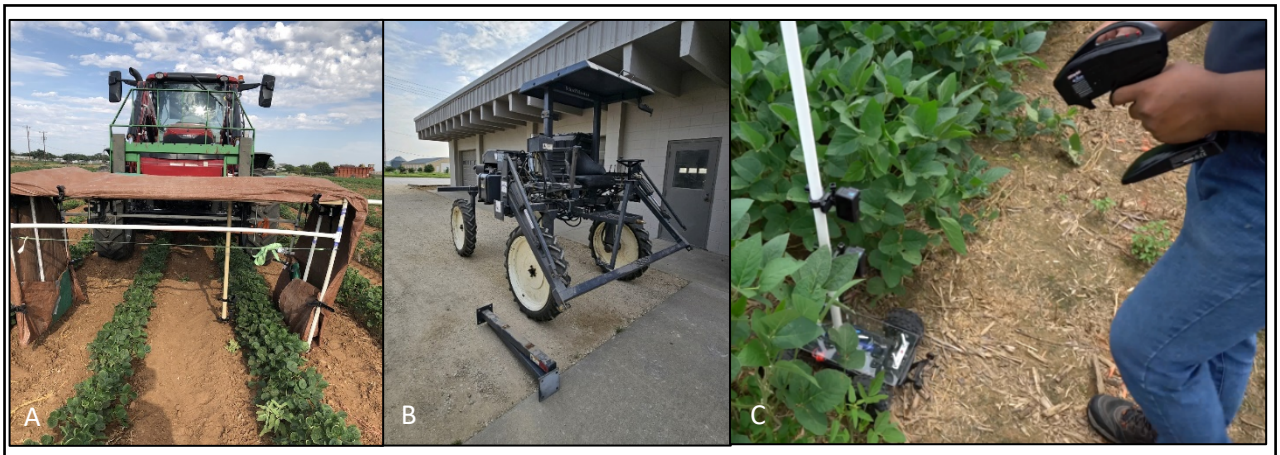

Figure 2. Platforms used at Texas Tech University (A), Kansas State University (B), and University of Tennessee (C).

At the University of Missouri, imaging procedures involved the use of a bicycle, and adjustments were implemented to enhance imaging quality, as necessary. The seed harvest began in September, with the harvested seeds designated for next year's planting. Sufficient seed has been produced at the University of Missouri for supporting Year 2 trails across all four locations.

Kansas State University utilized a modified spray vehicle (Figure 2B) with adjusted wheel spacing to straddle 10' wide plots, serving as the imaging system's mounting platform. This modification ensured optimal coverage and accessibility for capturing high-quality images of soybean plants.

The University of Tennessee employed GoPro Hero11 cameras on a Traxxas Hoss® 4x4 VXL (Figure 2C) conveyor for soybean plant imaging.

Towards the end of the season, soybean harvesting at Texas Tech University took place from September 22nd to October 18th, revealing an average *flower abortion rate of approximately 47% among 30 genotypes (Figure 4)*. Ongoing measurements, including yield per plot and on area basis, pods per node, number of seeds per plant, weight of 1000 seeds, and plant height, are still in progress. The flower abortion variation across 161 genotypes, from the 228 initial ones, was recorded only at Texas Tech University, spanning between 20% to 80%, as illustrated in Figure 5. These findings corroborate the existing literature regarding the variability in abortion rates observed in soybean plants.

At the University of Missouri, the average *flower abortion rate was approximately 50%, ranging between 37% and 62%. (Figure 6).* Harvesting of diverse soybean lines has been concluded, and seed threshing to estimate yield is in progress.

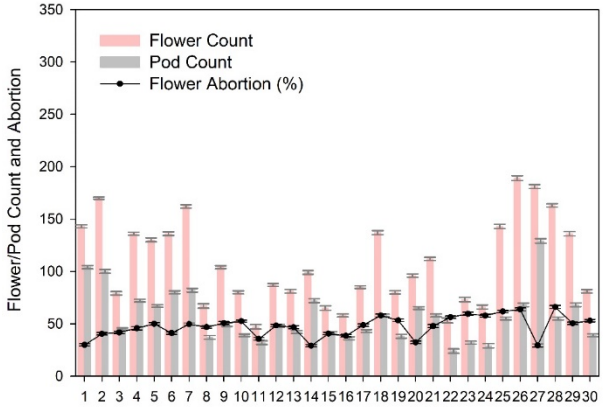

Figure 4. Flower abortion from 30 genotypes grown at Texas Tech University experimental farm.

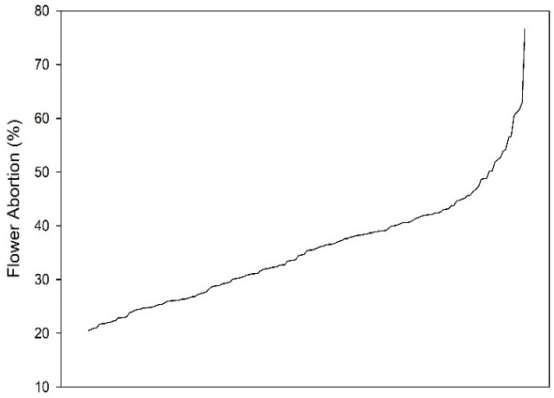

Figure 5. Flower abortion in other 161 genotypes of soybean grown at Texas Tech University

University of Tennessee harvested soybean plots starting on September 14th. *The average flower abortion rate at Tennessee is approximately 17%, ranging up to 29%, as shown in Figure 7.*

Kansas State University completed flowering, and pod counts on a core set of 30 genotypes, The average *flower abortion rate was approximately 40%, ranging between 22% and 70%.* Data on plant maturity, lodging, height, and seed yield are being evaluated. Threshing and cleaning of harvested seed, along with video analysis, are ongoing to assess the relationship between field flower counts and video-tracking flowers.

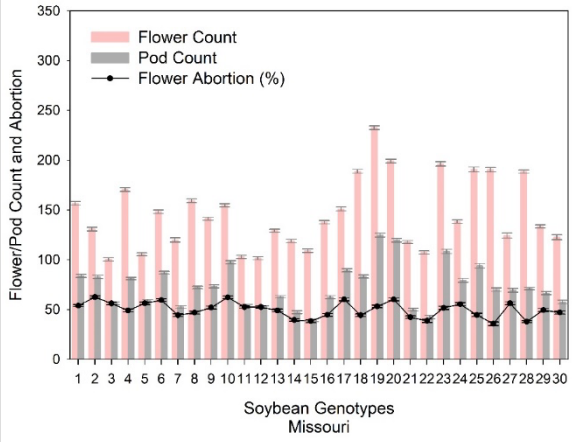

Figure 6. Flower abortion results from 30 genotypes grown at the University of Missouri

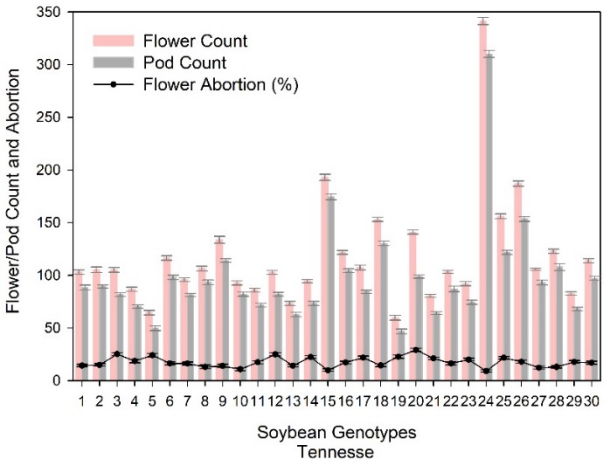

Figure 7. Flower count, pod count, and flower abortion rate of 30 genotypes measured at the University of Tennessee's West Tennessee Research and Education Center

Objective 2 - Develop an image-based field phenotyping system and deep-learning tools to precisely document temporal dynamics in flower abortion and pod retention in genetically diverse soybeans.

Ahead of the field season, our team maximized the use of greenhouse by grown soybean plants to develop a robust machine learning tool (Figure 8A & B). This helped to target the primary objective, to detect flower numbers and the rate of abortion under field conditions.

We established an imaging protocol for greenhouse plants, capturing images from multiple views with high resolution (e.g., 4K x 6K). This protocol ensured that even the smallest flowers comprised a minimum of 30 pixels. Our two-stage strategy involved subsampling the acquired image for node detection and subsequently cropping the original image for flower detection.

Node-Detection Network: As an initial approach to detecting nodes, we have employed the Faster R-CNN architecture. We started by pre-training our model with a dataset provided by the studies in 2023 that focused on detecting nodes on Eggplant, Chili, and Tomato plants.

Figure 8. Greenhouse testing to develop machine learning models for flower detection (A) and obtaining ideal camera parameters (B).

The existing model's inference on this dataset indicated that the model's ability to generalize is reasonably good as it was able to locate most of the visible nodes in the new images. However, the model also outputs several more False-Positives and False-Negatives.

Flower detection: The same imaging protocol previously mentioned was used for the flower detection network. Like the node detection network, the flower detection network was also based on the Faster R-CNN architecture. Specifically, we used the Faster R-CNN implementation available in Detectron2 (a library containing state-of-the-art detection and segmentation algorithms made publicly available by Facebook AI Research). We trained an initial model based on a dataset published by Zhu et al. (2022).

Moreover, we constructed a dataset of 154 images that were captured from our greenhouse soybean plants before March 1st. These were subsequently annotated and divided into training and test sets for model development. Notably, during our annotation process, we separated the nodes into two distinct categories: nodes with flowers and nodes without flowers

We then proceeded to fine-tune the pre-trained model for our application by training it on our training set. As a result, the model exhibited improved performance on our test set, particularly in densely populated scenes.

Using both node detection and flower detection (Figure 9) for counting soybean flowers showed great potential to be used, however, it required excessive image processing and storage. For that, it was decided to use only flower detection moving forward on the field trails, as the model demonstrated the ability to effectively detect flowers without the need for node detection.

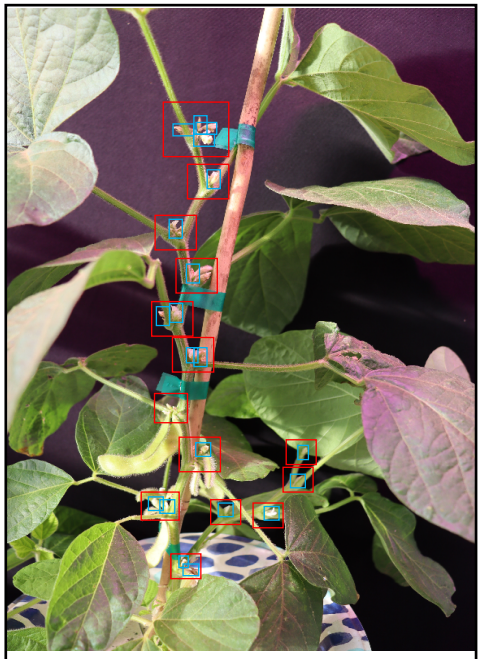

Figure 9. Node (red) and flower (blue) detection in Soybean.

A GoPro Hero 11 camera 27-megapixel was evaluated due to ease of use and image collection. A protocol for image quality collection was developed based on the GoPro camera parameters and was shared with all collaborators. The imaging system was tested in a greenhouse (Figure 8B) and its ability to capture and record high-quality images at 60 frames per second was verified. Furthermore, the captured images were used as input to the flower detection model with successful outcome.

Improving the precision of the flower detection model.

We have fine-tuned the original Faster R-CNN flower detection model to improve its predictions. Specifically, the model was fine-tuned with a variety of images, some taken in a more controlled environments and others resembling images taken in-the-field (Figure 10); some more focused, and others somewhat blurred; or images taken with different imaging systems/cameras producing different resolutions and quality. The Average Precision for detections whose bounding boxes overlaps by at least 50% with the ground truth bounding boxes (denoted as AP50) was 79.53 on the test images.

Adding pods to the flower model

As a new approach Kansas State University team have enriched our model with the ability to detect pods. We have adapted the previous Faster R-CNN model to detect pods (in addition to flowers) by fine-tuning it with 2693 annotated pod images. We used the Faster R-CNN implementation available in Detectron2.

Texas Tech University team has been working on dataset preparation for the flower detection model and implementing an algorithm for flower counting. These are essential to the foundation for the successful development of our phenotyping system.

Accurate flower counting in captured videos is essential for Objective #2. To achieve this, we focused on tracking detected flowers across frames to prevent overcounting. We evaluated the three tracking algorithms with various parameter combinations. Surprisingly, our findings indicate

that the choice of algorithm is not the critical factor for accurate tracking and counting of flowers. All three algorithms yielded accurate results when specific parameter settings were used.

Figure 10. Field images used to increase the precision of the flower detection model

Our focus at Texas Tech University has been on advancing the development of a customized Multi-Object Tracking (MOT) algorithm tailored for counting soybean flowers in the field, aligning with the overarching objective of creating an image-based field phenotyping system. The tracking dataset has expanded to five videos Quarter4_Tracking_1.mp4 comprising 22,606 individual flower annotations, enriching the dataset with varied environmental conditions and soybean varieties. Evaluation methods include assessing the accuracy of counting flowers and gauging the quality of flower tracking across frames. Algorithm development incorporates two additional state-of-the-art tracking algorithms, ByteTrack and DeepSORT. Preliminary findings suggest that the integration of deep learning may adversely affect tracking algorithm performance, guiding next steps in algorithm refinement and optimization.

At Kansas State University, recognizing the intricate shape of soybean pods, an instance segmentation method was employed for identifying and counting pods, providing precise segmentation masks for accurate representation (Figure 11). To address limited labeled data, images extracted from field videos were annotated using AnyLabeling, a tool driven by the Segment Anything Model (SAM). Approximately 300 frames were annotated and split into three

subsets for model training, development, and evaluation. The current trained model shows performance in terms of average precision (AP), average precision at 50% IoU (AP50), and average precision at 75% IoU (AP75) 42.373, 63.557, and 44.903, respectively. The annotation approach enhances the robustness and reliability of the dataset for more accurate analyses of soybean phenotypic traits. In addition to pod segmentation, we have also worked on tracking the soybean pods. We are currently using multi-object trackers and fine-tuning the trackers, so that we can use them on field level vídeos (Figure 12).

Figure 11. Samples of annotated frames of the Soybean pod in the field, before and after annotated.

Figure 12. Soybean pod tracking imaging.

Objective 3 - Discover environmentally stable and region-specific genomic regions controlling flower abortion in diverse soil types, moisture, and climatic conditions.

In our exploration of flower abscission in soybeans, we surveyed the key determinant genes involved in flower and flower organ abscission in Arabidopsis and identified orthologs in soybean genome. Most genes expressed in abscission layer in the model organisms are associated with hormone biosynthesis/transport and nutrient uptake. We have selected a subset of these genes (mainly transcription factors) involved in hormone regulation. We conducted a gene-based haplotype analysis to select the group of lines and correlated the large effect variants with the phenotypic data. The confounding effect (if any) of flowering QTLs was compared for the selected genes.

We have shortlisted 6 genes (Blade on Petiole (BOP), KNAT (KNOX genes), BREVIPEDICELLUS 1 (BP1), INFLORESCENCE DEFICIENT IN ABSCISSION (IDA), HAE/HSL (leucine-rich repeat receptor like kinase), and DNA BINDING WITH ONE FINGER 4.7 (DOF4.7)) in Arabidopsis which correspond to 27 orthologous genes in soybean involved in floral organ abscission. In addition, we shortlisted additional genes which have been reported to also play a role in floral organ abscission in addition to their known function- ASYMMETRIC LEAVES1 (AS1), AGAMOUS-like 15 (AGL15), and FOREVER YOUNG FLOWER (FYF). The mutant alleles of these genes have shown significant

effect on several stages of floral organ abscission. Lastly, the maturity locus E1-E4 plays a significant role in the regulation of flowering in soybeans. The J locus, ortholog of AtELF3 (EARLY FLOWERING 3), is under the influence of E1. The functional analysis of mutant alleles for these genes showed an early flowering phenotype. The haplotype analysis for these genes is currently in progress, the analysis shows that some higher maturity group (MG) lines used in the current project (MG III, IV) retain one or more of the variant alleles. From this analysis, a group of lines correlating large effect variants associated with flowering traits (floral initiation and flower abortion) was selected to identify causal genomic regions and thereby underlying genes.

In our prior analysis, we meticulously selected six pivotal genes that are well-documented in their roles pertaining to the initiation of the abscission zone (AZ), the facilitation of AZ development through ethylene signaling, the activation of tissue separation mechanisms, and the subsequent deposition of protective layers following organ detachment from the plant. Furthermore, we incorporated genes known to be involved in soybean maturity and flowering processes. To perform a robust gene-based clustering analysis and to identify alleles with substantial effects, we leveraged a state-of-the-art gene-haplotype analysis framework, which was executed on the high-performance computing servers at TTU. In a preliminary study, we executed the gene-based haplotype analysis on a cohort comprising 481 lines as a means of testing our analytical pipeline using major flowering genes. During this analysis, we successfully pinpointed four significant haploblocks, with particular emphasis on haploblocks H1 and H4 (as illustrated in Figure 13A), which exhibited pronounced allelic variations possessing substantial effects on the observed traits. While the haplotype analysis of additional genes remains an ongoing endeavor, our ultimate objective is to compare the lines that overlap with these haploblocks to field data especially flower number and aborted flowers. Following the data collection from all locations will overlay the phenotypic data with haplotype analysis to identify the most diverse accession for further analysis. Previously, we selected key soybean homologs involved in flower abortion including Initiation of abscission zone (AZ) and promotion of AZ by ethylene and performed haplotype analysis. We identified two major and two minor haplotypes for one of the transcription factors (GmRNI) involved in flower organ abscission (Figure 13A). The major haplotype carries two alleles and showed higher allelic diversity in wild accessions (Figure 13B). Most interestingly, this gene expressed during R1 flower stage in multiple soybean accession and suggests a critical role in flower development and probably in floral abscission (Figure 14). Currently we are performing additional analysis to identify allelic variants in a subset of accessions that were selected from Year 1.

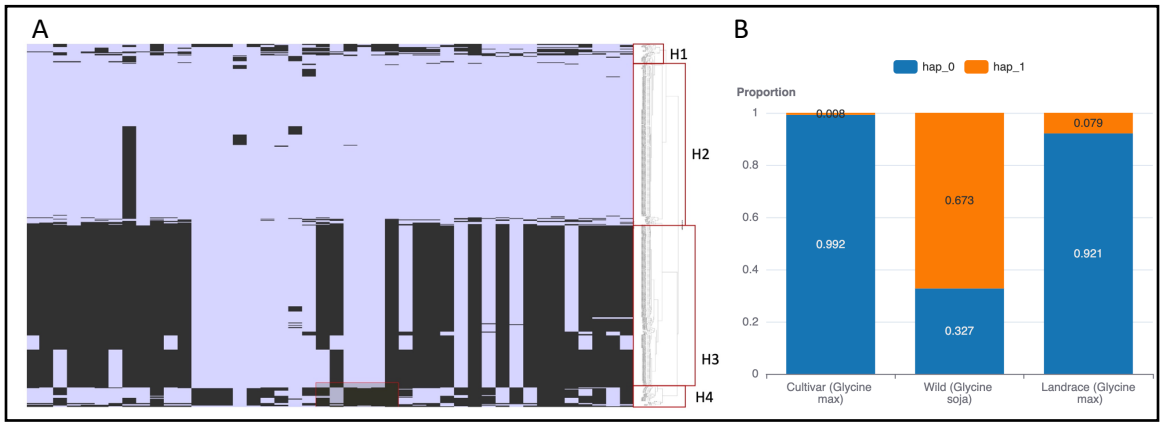

Figure 13. Identification of (A) haplotypes and (B) allelic variants for TF involved in floral organ abscission in soybean.

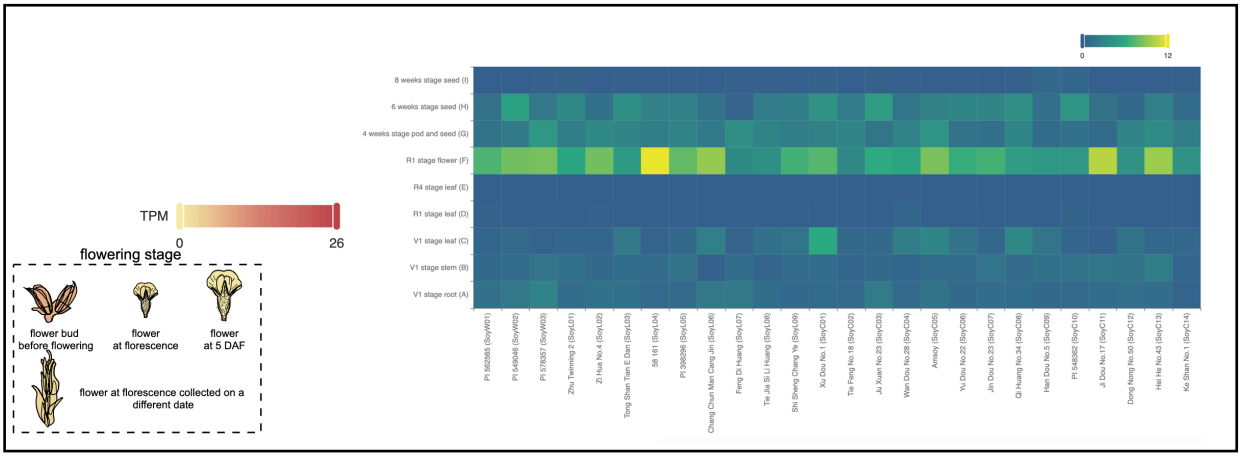

Figure 14. Expression of selected TF in divers soybean accession indicates flower tissue specific expression.

Showcasing the Project and Presenting Results

Texas Tech University

July 2023, Dr. Jagadish aired a radio interview of the Dakota Farm Talk to highlight the project and indicated the benefits that the progress made will have on the US and global soybean industry.

October 29th, 2023, Dr. Espíndola delivered an oral presentation titled "Advancing Phenotyping for Flower Abortion in Soybeans through Image Analysis and Machine Learning" at the 2023 annual meeting of ASA-CSSA-SSSA in St Louis, Missouri.

University of Tennessee

Tennessee team were able to release a podcast on the UTIAg website (available on Spotify for Podcasters) about our soybean flowers abortion project. Find the link here: https://podcasters.spotify.com/pod/show/utiag/episodes/Culture--Agriculture-Ep--4-Research-Could-Improve-Soybean-Yield-e277htq/a-aa5c7ku

Furthermore, they worked with the UTIA communication team to create and release a video about our current research project. To see the video, click here: https://www.youtube.com/watch?v=H5CVeWbiliU

Finally, they put a research abstract together titled “Image-based field high throughput phenotyping for quantifying flower abortion in genetically diverse soybean germplasm” and submitted it to the 2023 ASA-CSSA-SSA annual meeting in St. Louis, Missouri.

In brief, Year 2 work plan

- Grow 50 diverse and elite lines in all four locations, in 4 row plots with 3 replications; irrigated and stress (50% irrigation) at Texas Tech University
- Obtain temporal (4 days once) images and manually count flowers from at least 4 plants per line
- Establish a uniform vehicle carrier for the imaging system for uniform video generation across locations
- Refine both the flower detection and the pod detection models and enhance the predictive capacity
- Candidate lines with increased flower and pod retention identified, specifically for each location and across locations
- Unravel molecular mechanisms (transcriptomics) that control flower abortion from contrasting lines identified from Year 1 field trials, under controlled greenhouse trails
- Identify key transcripts that determine flower abortion in soybeans