Project report (Fourth quarter October 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

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)

• Explore the genetic diversity in flower abortion under different soil moisture and climatic conditions using a large diversity panel

• 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

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

Progress achieved

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

Note – All data presented on flower abortion is preliminary as the teams are yet to have thorough discussions on the approach taken to ensure the percentage abortion presented is confirmed.
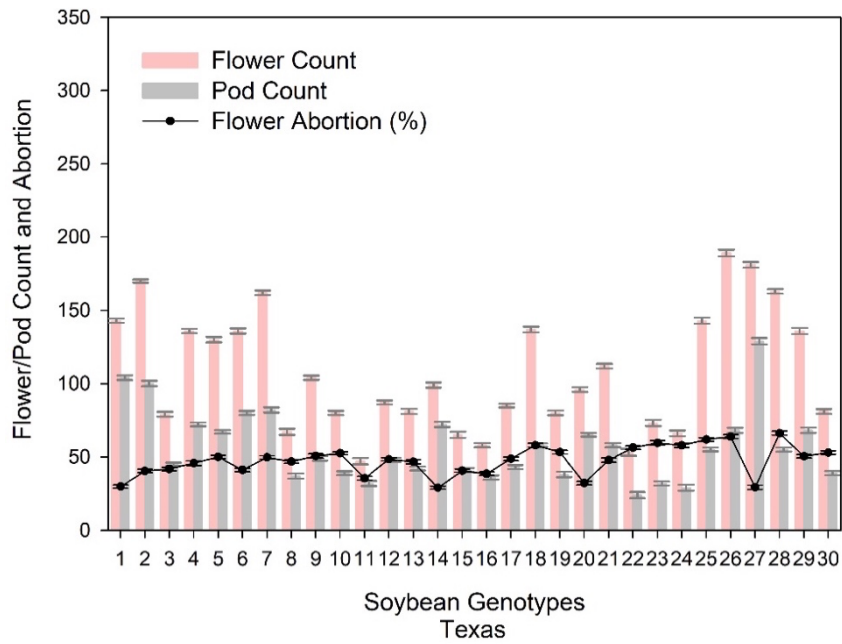
Texas Tech University
Soybean harvesting started on September 22nd and concluded on October 18th (Figure 1). Figures 2 and 3 showcase the data gathered from the field for flower and pod counts. Other measurements including yield, pods per node, number of seeds per plant, and 1000 seed weight, are currently in progress and will be reported upon completion.
In Figure 2, the data illustrates that among the 30 genotypes, the average flower abortion rate is approximately 47%. Meanwhile, Figure 3 presents the percentage variation in abortion across 161 genotypes exclusively analyzed at Texas Tech University. This range spans from 20% to 80%
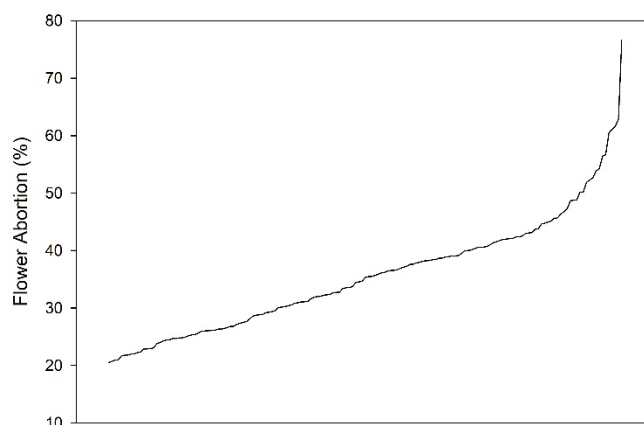
among the different genotypes. The temperatures and precipitation of the experimental farm during the trail is presented in Figure 4.
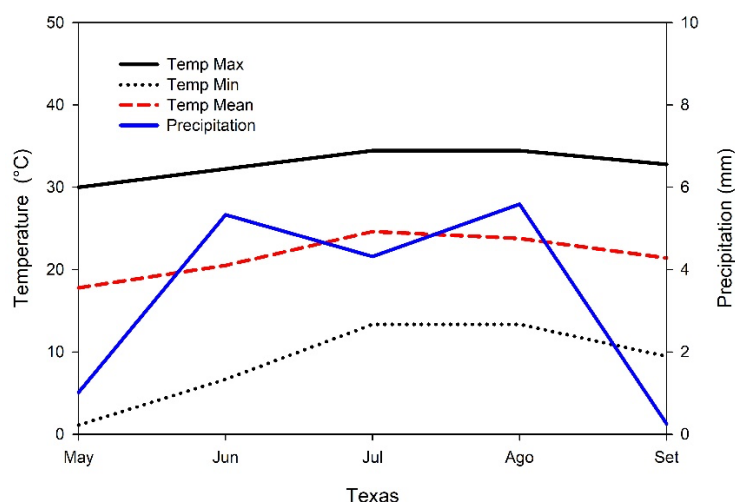

Figure 1. Soybean harvesting at Texas Tech University experimental farm.

Figure 2. Flower abortion results from 30 genotypes grown at Texas Tech University experimental farm.

Figure 3. Flower abortion variation of extra 161 genotypes of soybean entries grown at Texas Tech University experimental farm.
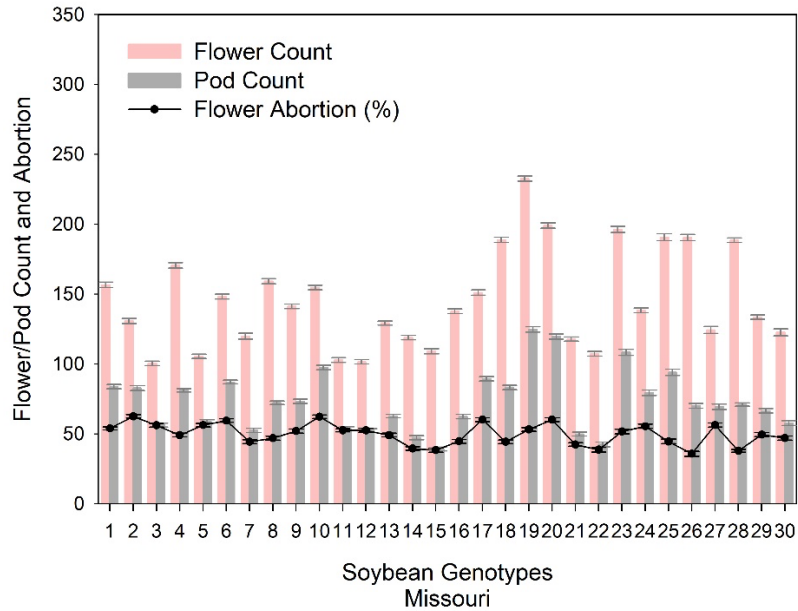

Figure 4. Means of Temperature (Maximum, Minimum and Mean) and Precipitation at Texas Tech University experimental farm in 2023.

University of Missouri

Using phylogenetic analysis a core set of 30 lines was selected that represented the genetic diversity to perform manual flower/pod counting and imaging. We counted flowers over 9 times with intervals of 3 to 4 days during the flowering stage. We performed the final pod counting in October to estimate the flower abortion rates of selected 30 lines. Preliminary data on the flower and pod count and related abortion percentage is presented in Figure 5. The average flower abortion rate at Missouri is approximately 50%, ranging between 37% and 62%. Video imaging at all 9 times was taken during flowering of these 30 lines and shared with the group to optimize the ML-based automatic flower counting platform.

Harvesting of the entire diverse panel of 280 lines from the field (center 2 rows). The harvested plants will be threshed to estimate yield of these lines. The yield data will be used to correlate with flower abortion rates. The temperatures and precipitation of the experimental farm during the trial is presented in Figure 6.

Figure 5. Flower abortion results from 30 genotypes grown at Texas Tech University experimental farm.
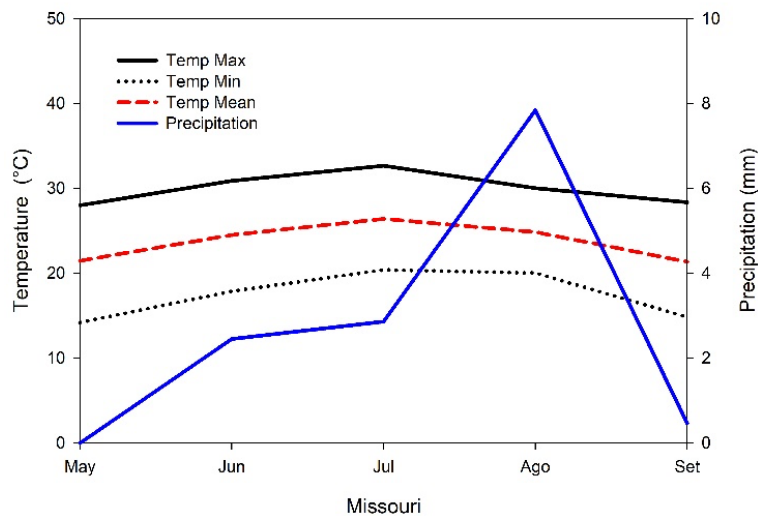
Figure 6. Means of Temperature (Maximum, Minimum and Mean) and Precipitation of Texas Tech University experimental farm in 2023.
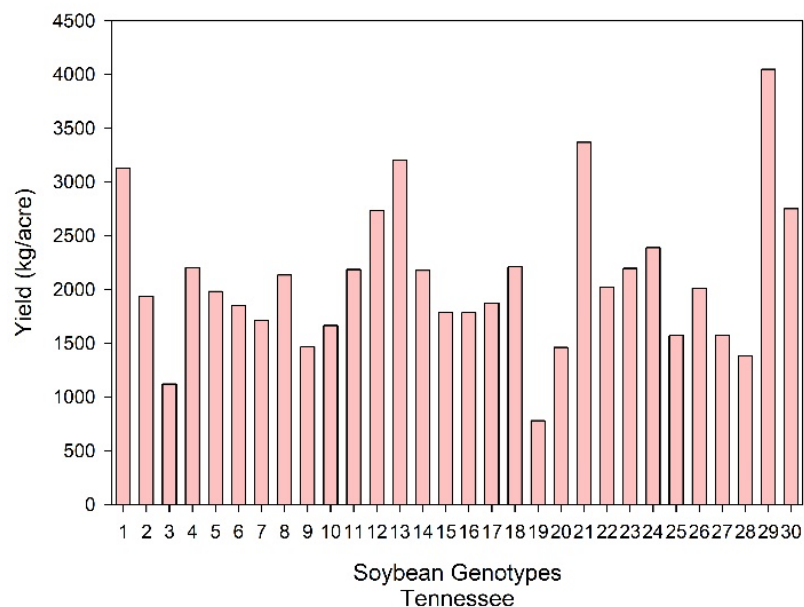
University of Tennessee

Harvesting of soybean plots at the University of Tennessee at the West TN Res. and Edu. Center (WTREC) started on September 14th (Figure 7). Plants were harvested manually from the two-row plots within 1 m^2 and the total seed weight was recorded to calculate the yield in kg/acre (Figure 8). The tagged plant that was used for manual counting of flowers and pod along with other 4 plants within the same row were collected using a burlap fabric roll. Morphological characters such as plant height and number of branches as well as yield component parameters including number of pods plant^{-1} and number of seeds plant^{-1} were recorded before threshing. All harvested plants were threshed using the USDA single plant thresher then the collected seeds per plot were weighed to account for yield. All 90 plots were harvested and data collection is near completion. Figure 9 shows the rate of flower abortion of the soybean genotypes.

**Figure 7.** Soybean harvesting and threshing at the University of Tennessee's West Tennessee Research and Education Center.

The abortion rate among the 30 genotypes was upto 29% The temperature and precipitation at the experimental farm during the trial is presented in Figure 10.

Figure 8. Yield (kg/acre) of 30 genotypes reported at the University of Tennessee's West Tennessee Research and Education Center.
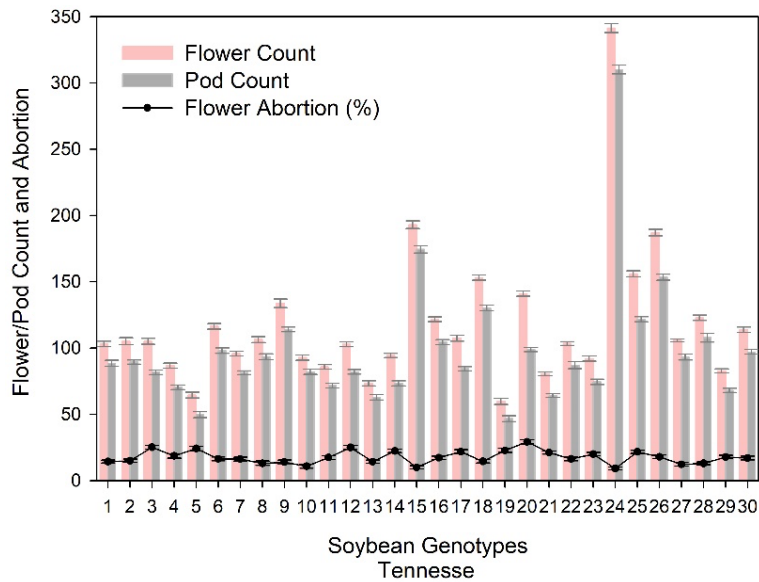
Figure 9. Flower count, pod count, and flower abortion rate of 30 genotypes measured at the University of Tennessee's West Tennessee Research and Education Center.
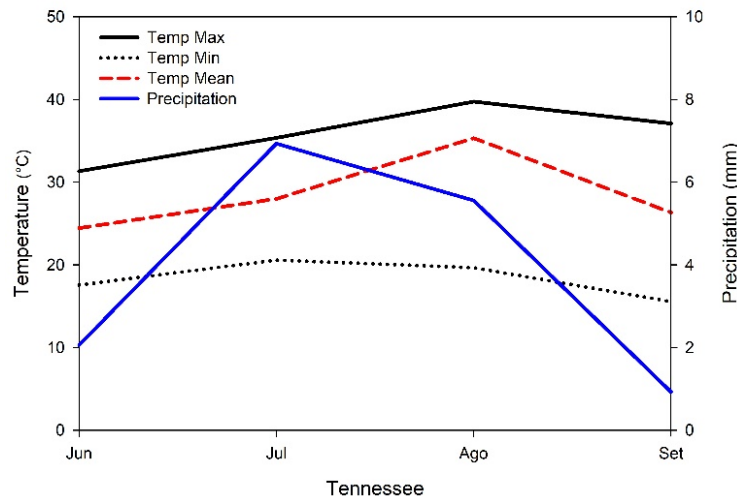
Figure 10. Means of Temperature (Maximum, Minimum and Mean) and Precipitation of University of Tennessee experimental farm in 2023.

Kansas State University

Flower and pod counts were completed this quarter on the core set of 30 genotypes. At the end of the growing season, single plants used for flower and pod counts from this core set were harvested. Number of pods, nodes, and seeds per pod, along with total seeds, 100 seed weight and total seed weight were determined for each of the single plants. Those counts have been completed. The data

is now being evaluated for quality prior to analysis. Based on a preliminary analysis of the flower and pod counts, we observed about a 40% difference in the relative flower abortion in the core set, with abortion ranging between 22% and 70%. This needs to be confirmed and compared to the results at the other locations. Seed yield, plant maturity, lodging and height were taken on the entire panel during September, October and early November. At harvest, these plots were threshed with a stationary thresher. The harvested seed is now being cleaned and weighed to measure final seed yield. Videos taken weekly of the developing plants are now being inventoried and labeled to evaluate the relationship between the flower counts in the field throughout the season, and the detection of the flowers in the videos.

Objective 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.

Texas Tech University
In this quarter, our focus has been on advancing the development of a customized Multi-Object Tracking (MOT) algorithm specifically tailored for counting soybean flowers in the field, aligning with the overarching objective of creating an image-based field phenotyping system.

1. Tracking Dataset Preparation, Continued

As mentioned in the previous report, preparing a dataset for the evaluation of MOT algorithms is quite challenging and time-consuming. This intricate task involves identifying and tracking individual flowers across an extensive sequence of consecutive frames, a laborious process exacerbated by the presence of long-term occlusions. This endeavor requires precision and attention to detail, as each flower needs to be tracked separately.

Our tracking dataset has expanded from one video (211 frames) to five videos (1,382 frames), comprising 22,606 individual flower annotations. These videos, hailing from diverse locations such as Kansas, Tennessee, Missouri, and Texas, enrich our dataset with varied environmental conditions and soybean varieties. This growth enhances the representativeness and applicability of our dataset for robust algorithm evaluation and refinement.

It's important to note that well-constructed MOT datasets are inherently scarce, given the complexities involved. Crafting a MOT dataset tailored specifically for Soybean Flower counting adds an extra layer of rarity. Even in its current state, we consider our dataset highly valuable, recognizing its uniqueness within the research landscape.

Our commitment to dataset expansion remains steadfast. The addition of more diverse videos is key to enhancing the accuracy of our algorithm evaluations, and we will continue this effort in the upcoming phases of our research.

2. Tracking and Counting Evaluation Method

We've delved into diverse evaluation methods for our tracking and counting algorithms. Specifically, we've implemented two approaches:

1. A dedicated approach for assessing the accuracy of counting flowers.
2. An evaluation method gauging the quality of flower tracking across frames. This method incorporates various metrics, with a particular emphasis on achieving high tracking accuracy.

Inspired by the widely recognized [HOTA](#) paper by Luiten (2020), this approach is instrumental in refining our counting performance.

## *3.	Tracking for Counting Algorithm Development*

Expanding on our algorithm development, we've incorporated two additional state-of-the-art tracking algorithms, ByteTrack and DeepSORT, alongside our previous implementations of SORT, OC-SORT, and OC-SORT with Byte. The inclusion of DeepSORT is particularly noteworthy, introducing deep learning and a specialized neural network for tracking. While still in the active exploration phase, our preliminary findings indicate an unexpected trend— the integration of deep learning appears to be adversely affecting the performance of our tracking algorithm. These initial results carry significant implications and shed light on the current state of tracking algorithms. Our commitment to rigorously evaluating and investigating these outcomes remains paramount, guiding our next steps in algorithm refinement and optimization.

In summary, our investigation into tracking algorithms for counting soybean flowers is groundbreaking. While previous studies focus on detecting soybean flowers from still images, our work stands out by offering a novel solution to use detection algorithms for large-scale flower counting. This unique contribution addresses a gap in existing research, providing a practical approach to advance soybean phenotyping.

Kansas State University

Recognizing the intricate shape of soybean pods, we have opted for an instance segmentation method as opposed to bounding box object detection, for the task of identifying and counting the pods. The instance segmentation approach enables us to obtain precise segmentation masks of the pods, ensuring a more accurate representation of their complex structures.

Addressing the challenge of limited labeled data, we have devised a strategy of utilizing images extracted from field videos captured at various stages of the soybean growing phase. These images are subsequently annotated using AnyLabeling, a powerful tool driven by the Segment Anything Model (SAM) developed by Meta. This innovative tool allows us to generate precise masks by leveraging weakly supervised prompts, mitigating the need for laborious labeling of segmentation masks.

Through the utilization of AnyLabeling, we have successfully annotated approximately 300 frames of images extracted from videos recorded across all four locations. This annotation process not only facilitates the accurate identification of soybean pods, but also contributes to the enrichment of our dataset with diverse and representative samples. The incorporation of weakly supervised prompts, coupled with the efficiency of SAM, empowers our annotation process, ensuring that the generated masks accurately delineate soybean pods in their varying stages of development. This meticulous annotation approach enhances the robustness and reliability of our dataset, laying the foundation for more accurate and comprehensive analyses of soybean phenotypic traits.

The 300 annotated frames were split into three subsets, used for model training (Train), model development/hyper-parameter tuning (Valid) and model evaluation (Test) as shown below:

Train	Valid	Test
200	50	50

The performance of the current trained model is shown below in terms of average precision (AP), average precision at 50% IoU (AP50) and average precision at 75% IoU (AP75):

AP	AP50	AP75
42.373	63.557	44.903

Some samples of annotated frames are shown below (Figure 11), together with the original un-annotated frames.

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

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).

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

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 13). 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
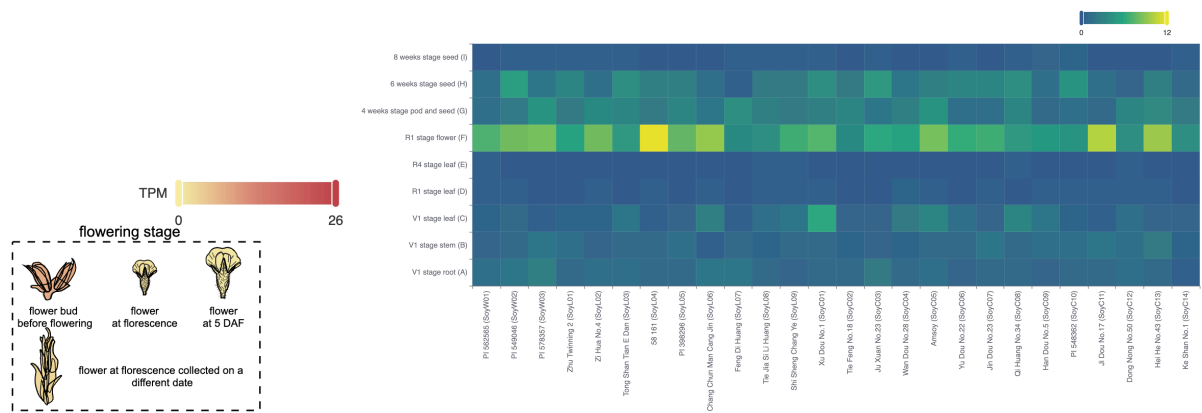

Figure 12. Soybean pod tracking imaging.

identify allelic variants in a subset of accessions that were selected from Year 1.
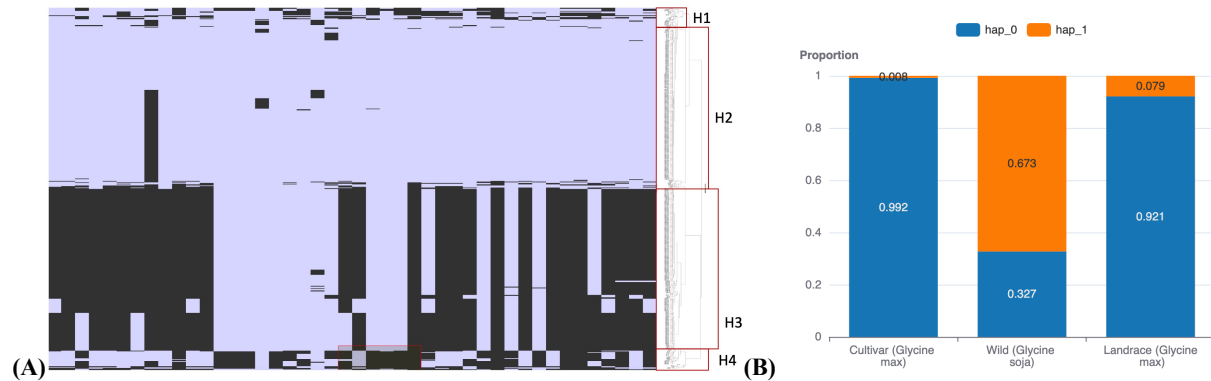
Figure 13. Identification of (A) haplotypes and (B) allelic variants for TF involved in floral organ abscission in soybean.

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