

United States Department of Agriculture
Project Initiation

Project Data for Accession Number 1034401

Project Title

Mitigating The Risk Of Abiotic Stress Damage In Blueberry Production Through Pangenome Graph-Based Genotyping And Phenomics Approaches

Project Details

Sponsoring Institution National Institute of Food and Agriculture	Project Status ACTIVE
Funding Source NIFA Non Formula	Grant Year 2026
Grant No. 2026-67013-45981	Cumulative Award Amt. \$617,000.00
Proposal No. 2024-12870	Multistate No. (N/A)
Project Start Date Apr 01, 2026	Project End Date Mar 31, 2029
Program Code [A1141] Plant Health and Production and Plant Products: Plant Breeding for Agricultural Production	

Project Director

Chu, Y.

Recipient Organization

UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC.
310 E CAMPUS RD RM 409
Athens, GEORGIA 306021589

Performing Department

(N/A)

Non Technical Summary

Blueberry production in the US is threatened by spring freeze damage and summer heat stress. Addressing these issues aligns with the AFRI legislative priority to "Genetically dissect and introduce desirable traits to improve plant's tolerance to temperature extremes". Warm spring temperatures promote early blooms which elevates the risk of early freeze damage; delaying flowering time will reduce the risk. Conversely, when the blueberry fruiting season enters the sizzling summer days, fruit quality is compromised due to heat stress. Shortening the fruiting period will allow the fruits to mature sooner and minimize economic losses to heat stress. Our blueberry breeding population is suitable for genetic dissection of these two critical phenological traits because it exhibits wide ranges of variation in both blooming time and fruiting period. In addition, high correlations between two years of visual ratings indicate strong genetic inheritance of both traits. The objective for this proposal is to identify genetic regions and genetic markers controlling

United States Department of Agriculture
Project Initiation

blueberry flowering time and fruiting period through a pangenome graph-based genotyping and phenomics approach. The long-term goal of this project is to develop resilient blueberry cultivars and enable data-driven farm management. An expanded pangenome graph, genetic markers and phenomics data complementary to the VacCAP project will be made publicly available. These advanced genotyping and phenotyping resources will accelerate trait pyramiding for the blueberry breeding community. Furthermore, a mobile phone App informing blueberry phenological development will be available to growers to enable data-driven farm management decisions.

Animal Health Component

Animal Health Component: 0%

Research Effort Categories

Basic	50%
Applied	50%
Developmental	0%

Classification

Knowledge Area (KA)	Subject of Investigation (SOI)	Field of Science (FOS)	Percent
205	1120	1081	100%

Knowledge Area

[205] Plant Management Systems

Subject Of Investigation

[1120] Blueberry

Field Of Science

[1081] Breeding

Keywords

Blueberry, heat stress, high throughput phenotyping, freeze damage, graph-based genotyping

Goals / Objectives

The long-term goal of this project is to mitigate the risk of spring freeze and summer heat stress damage in blueberries through genomic breeding and improve farm management in response to erratic weather conditions. The specific objectives are 1. Characterize phenological development of the blueberry breeding population using a high throughput robotic phenotyping platform and deep learning. 2. Construct an expanded pangenome graph and curate genetic variants from our breeding population through pangenome graph-based genotyping. 3. Discover genomic regions conditioning blueberry flowering time and fruiting period through Genome Wide Association Studies analysis.

Project Methods

United States Department of Agriculture Project Initiation

Customized mobile robot: A customized mobile robot will be developed using modules from the MARS X robot. The robot will feature four independently steerable drive wheels, allowing it to move both laterally and longitudinally. It will have an adjustable frame that can change its height and width, enabling it to adapt to different bush sizes and be easily stored. The robot will drive over blueberry bushes to scan both sides in a single pass. A camera mounting frame will be installed on each side, equipped with a lifting arm powered by a linear actuator to adjust the height of the camera. Six customized stereo cameras will be installed to cover the entire blueberry bush. To control lighting conditions, a canopy made from translucent fabric will be installed on the robot. Real-time image processing will be handled by NVIDIA Jetson AGX Orin. The software of the robot will be developed using the Robot Operating System (ROS) and will utilize RTK-GNSS for localization and autonomous navigation. The detection results will be georeferenced based on the location of the robot. The proposed robot will be equipped with a custom camera sensor. This sensor will be a stereo camera, consisting of two Raspberry Pi HQ camera modules. These modules are advantageous because they have high resolution and support interchangeable lenses. The cameras themselves will be equipped with some ability to process the data internally. This design alleviates load from the robot's main controller and leads to a system that is more expandable. Internally, the camera is equipped with a Raspberry Pi 5 to read data from the cameras and perform stereo matching, as well as an AI accelerator that can be used to perform basic computer vision tasks such as object detection. The Raspberry Pi runs on ROS and communicates with the rest of the robot via an Ethernet connection. The camera is also equipped with power-over-Ethernet, so that only this single cable is required.

Image Processing Pipeline: To detect individual blueberries and/or flowers in each image, YOLOv8 algorithm with enhancements will be developed. The main advantage of YOLO is that it is fast, potentially enabling real-time tracking. For keeping track of detected objects over time, we propose using a version of GCNNMatch++ that is lightly modified to enable efficient real-time performance on low-powered hardware. This is an online, multi-object tracking approach that can associate the detections in multiple frames from the same camera. With a single camera, then, a robust MOT approach is all that is needed to derive accurate counts.

Population imaging and ground truth data collection: After establishing an accurate image analysis pipeline for classifying blueberry phenological stages, the entire population will be phenotyped by collecting both robotic images and visual ratings bi-weekly from February to June for two years.

Mobile application for flower and fruit counting: While ground robots can provide accurate flower and fruit counting, the cost and the technical expertise required for operation may limit accessibility for many farmers and breeders. We propose developing a mobile app that uses images captured by mobile devices to count flowers and fruits. This app will deploy our advanced detection model to identify and count these objects in the images. To speed up the model inference on mobile devices, we will optimize the network using pruning and quantization. The app will feature two primary interfaces: detection and file management. The detection interface allows users to either select an existing image or capture a new one for analysis. After running the detection model, the results will be displayed on the image, along with a count of the detected items. The file management interface can list all the detected files organized by blueberry fields. The App can overlay the images on Google Map using the location of the images to give an intuitive spatial distribution of the detection results of the field.

The expanded pangenome will be constructed using Minigraph-Cactus and refined using an in-house pangenome graph pipeline called khufuPAN. The final output will be a panHap map that can be visualized on a web interface. There is flexibility for genomes to be added or removed from the pangenome. Mapping resolution changes based on variations in the genomes present in the pangenome will be recorded, with the outcome of understanding if additional genomes may significantly affect mapping resolution. The whole genome sequencing of the breeding population will be carried out, and graph-based genotyping will be performed with the Khufu pipeline by mapping the reads to the expanded pangenome graph and calling SNPs, Indels, PAVs and CNVs, and imputing missing markers. Genotypic classes of all markers will be assigned using the updog R-package. P-values of genetic markers associated with flowering time and fruiting period will be assigned using the highly efficient exact method presented by the GEMMA software. The population structure will be determined by the first four principal components of the Principal Component Analysis, while the kinship will be computed using the AGHmatrix R-package using a tetrasomic inheritance model. Marker-trait association analyses will be conducted using a linear mixed model accounting for population structure and relative kinship matrices as implemented in the GWAS poly R-package. Additive and dominance gene action models will be tested in the GWAS analyses. Phenotype data from both visual rating and imaging analysis will be utilized for GWAS

United States Department of Agriculture Project Initiation

analysis. Significant markers will be annotated based on their genomic location and coding effects using the SnpEff software. Candidate gene mining will be performed in the significant genomic window surrounding the QTL. To facilitate the application of significant markers in blueberry breeding programs, we will convert markers within the QTL region to a capture enrichment panel. The process uses Twist 96-plex library preparation in combination with the Twist standard capture protocol. A custom probe set will be designed to include variants within the QTL regions so that all variants can be assessed for each sample in one reaction. Libraries will be sequenced on an Illumina NovaSeq 6000. Analysis of raw sequence reads will be done with custom scripts that assay the SNP panel and estimate allele dosage based on the ratio of alleles at each site.