November 5, 2021 Project report.

SOYGEN 2: Increasing soybean genetic gain for yield and seed composition by developing tools, know-how and community among public breeders in the north central US

Investigator Contact Information:

Leah McHale (PI), Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH 43206, 614-292-9003, mchale.21@osu.edu

Pengyin Chen, Division of Plant Sciences, University of Missouri, Columbia, MO 65211, 573-379-5431, chenpe@missouri.edu

Brian Diers, Department of Crop Sciences, University of Illinois, Urbana, IL 61801, 217-265-4062, bdiers@illinois.edu

George Graef, Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588, 402-472-1537, ggraef1@unl.edu

Matthew Hudson, Department of Crop Sciences, University of Illinois, Urbana, IL 61801, 217-244-8096, mhudson@illinois.edu

David Hyten, Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588, 402-472-3255, david.hyten@unl.edu

Aaron Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, 612-625-6754, lore0149@umn.edu

Katy Martin Rainey, Department of Agronomy, Purdue University, West Lafayette, IN 47907, 765-494-1212, krainey@purdue.edu

Nicolas Frederico Martin, Department of Crop Sciences, University of Illinois, Urbana, IL 61801, 217-300-3016, nfmartin@illinois.edu

Andrew Scaboo, Division of Plant Sciences, University of Missouri, Columbia, MO 65211, 573-882-3462, ScabooA@missouri.edu

William Schapaugh, Department of Agronomy, Kansas State University, Manhattan, KS 66506, 785-532-7242, wts@ksu.edu

Asheesh Singh, Department of Agronomy, Iowa State University, Ames, IA 50011, 515-294-7920, singhak@iastate.edu

Dechun Wang, Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI 48824, 517-353-0219, wangdech@msu.edu

Collaborator Contact Information:

Rex Nelson, USDA-ARS, Ames, IA 50011, 515-294-1297, rex.nelson@usda.gov

Objective 1: Elevating collaborative field trials

1c. Key performance indicators

(3) Collection of genotypic data from the Soy6KSNP chipfor UT and SCN regional trial entries.

We collected 6K genotype data on all 2020 UT lines. The 2020 SCN UT lines will be planted in the field along with all 2021 UT and SCN UT lines for tissue collection and genotyping. All

materials from 2021 UT and 2020 SCN UT was sampled and DNA isolation will commence shortly.

(4) Weather data will be collected for the majority of the future NUST field environments.

Weather datasets were collected in the site years corresponding to NUST field trials from using the geographic coordinates of the field trials linked with the DAYMET weather data. This information along with field trial phenotypic information will be used to compare the year to year site trialing similarity.

(5) The data from the NUST will be analyzed to determine the usefulness of test locations in predicting the performance of the experimental lines.

Objective 2: Development of a genomic breeding facilitation suite

2c. Key performance indicators

(1) Genotyping of 10,000 breeding lines using targeted GBS approach on 1k SNPs during first year of project.

We have received 9620 DNA samples to run with the 1k SNP set. Thus far, 6764 have been genotyped. The remaining samples are in the process of being sequenced.

(2) Beta version of R script to impute underlying whole-genome haplotypes developed.

The scripts were completed and are being tested in the Lorenz laboratory. We have been working to improve their accuracy and iterating new versions to make the scripts more useful in different use cases.

(4) Genomic data management system and allied analysis tools for adoption by soybean breeding community identified.

We installed a genome-wide marker database called GIGWA (https://gigwa.southgreen.fr/gigwa/) and deposited our current genome-wide marker data into this, including all the genotype data collected on the UT as part of this project. A workflow of software tools and scripts was initiated to seamlessly combine data held in this database with phenotypic data and genomic prediction models to ease the use of genomic selection in a practical breeding context.

Collaborator Rex Nelson (soybase.org) is working to implent a version of BreedBase for soybean breeders. An overview of the software package was givent to all PIs on the project who unanimously agreed to its utility. The BreedBase team has agreed to allow an instance in their cloud account for our work, which will make installation and implementation significantly simpler.

Objective 3: Evaluation of soybean breeding methods that increase gain

2

3c. Key performance indicators

(1) Grow single rep progeny row and preliminary yield trials and test two different methods of spatial adjustments (Task 1).

Code and full totorials for the selection process were shared with the entire research group during the last reporting period.

(3) Genotyping of advanced lines, development, and cross-validation of breeding program specific models (Task 2).

More than 6000 advanced breeding lines have been genotyped for the development of genomic selection models (see Objective 2).

(4) Genotyping of 2500 F4 lines in two years for each participating breeding program (Task 3).

DNA has been submitted for genotyping (+1000; Objective 2) while more are in process.

(8) Generate crosses for 5 cross combinations based on breeder selections and 5 cross combinations based on genomic mating selections for protein and yield (Task 4).

We used genomic prediction to predict the mean, variance, yld-pro correlation, and superior progeny mean of all possible crosses among 2019 and 2020 UT lines. We made this information available to all SOYGEN2 breeders for their consideration in terms of 2021 crosses. Crosses with model predicted parents and breeder selected parents were carried out during summer 2021 for 8 breeding programs.

(9) Advance generation by single seed descent for generated crosses in (8) and perform preliminary yield trials with protein data collected by NIRS on F3 or F4 derived lines in FY22 (Task 4).

Due to inability to MTA from the USDA for many of the cultivars used in the pedigrees of these lines, we were only able to complete a single cross combination: LG09-8165 x LG11-5120. Markers were developed for four loci predicted to be selected for yield. F2 lines have been genotyped and harvested. Seed will be sent to a winter nursery in Puerto Rico where F2:3 families will be produced for homozygous alleles and F3 inbred lines will be produced to further the generation of near isogenic lines derived from heterogenous inbred families.

(10) Perform crosses, genotyping, and line advancement according to rapid cycling breeding scheme (FY20-22) (Task 5).

Crosses were made in Nebraska summer 2020 and sent F1 seeds to Puerto Rico to grow F1 plants from October '20 to January '21. Intermating among F1 plants were attempted, but virus issues in Puerto Rico caused issues and we were not able to obtain all of the F1 x F1 crosses. Instead, F2 seeds were harvested from all of the confirmed F1 plants and are now crossing among F1:2 lines for the second intermating.

Objective 4: Characterization and use of the USDA Soybean Germplasm Collection, a foundation for future success

4c. Key performance indicators

(1) Compile and annotate the data for the validation study.

As we did for yield and agronomic traits previously, we conducted a genome-wide association analysis for each of the seed composition traits using the multi-year, multi-location phenotype data collected as part of this project, along with the existing 50K genotype data from the collection. The association analyses were conducted by sampling group (CLU, RAN, SSD) and over all lines together. Results for some of the seed composition traits are shown in Figures 4 to 6. We are continuing with analysis and interpretation of these results, identifying significant SNPs and underlying genes for each of the traits.