North Central Soybean Research Program

Initiation of a genomic selection pipeline for public soybean breeders in the North Central region

Aaron Lorenz (Project Leader, University of Nebraska), William Beavis (Iowa State University), Asheesh Singh (Iowa State University), William Schapaugh (Kansas State University), Dechun Wang (Michigan State University), Leah McHale (Ohio State University), Patrick Brown, Matthew Hudson, Alex Lipka, and Randall Nelson (University of Illinois-Urbana-Champaign), Brian Diers (University of Illinois-Carbondale), Henry Nguyen (University of Missouri), Andrew Scaboo, (University of Missouri), Grover Shannon (University of Missouri), George Graef (University of Nebraska), Katy M Rainey (Virginia Tech). Contact: lore0149@umn.edu

Final Report, October 2016

Increases in soybean yield through breeding are slower than producers expect. There are several possible reasons for the reduced rate of gain in soybean grain yield, including limited genetic variation in the commercially used gene pool, amount of time required for each breeding cycle, size of the breeding populations, and accuracy of evaluations. Advances in genomics have made whole-genome genotyping less expensive than multi-location yield testing. A powerful approach to make use of this genomic information for selective breeding is through a method called genomic prediction and selection. Large datasets of genomic and phenotypic information are required to maximize the effectiveness of genomic prediction.

Fortunately a wealth of data already exists within the public soybean community that could be used to initiate a genomic prediction pipeline to assist soybean breeders to more effectively select for yield and introgress diversity into their breeding program. This pipeline will evolve into a service for soybean breeders that will help address all of the major restraints to soybean breeding progress.

Project Objectives

1. Develop a regional genomic selection database for soybean breeders in the North Central Region.

- Compile existing phenotypic, genotypic, pedigree, and environmental data (weather, soil) from various projects on yield and diversity conducted in the North Central Region.
- Genotype all available soybean lines grown in the USDA Northern Uniform Tests beginning in 2004 and selected lines from regional tests organized by the USB Diversity Project.
- Unify genotypic data collected from the multiple platforms, using a single flexible data management system, capable of adapting to any genotyping platform.

- Develop and validate a unified genomic prediction method and algorithm that optimally uses data from diverse sources.
- Establish a genomic prediction service pipeline for soybean breeders

2. Identify large-effect alleles from exotic sources that improve yield potential of elite U.S. soybean lines
- Perform a genome-wide association analysis.
- Identify exotic genes affecting yield based on genomic sequencing of lines selected for yield from exotic germplasm source

Results

Goal 1. Develop a regional genomic selection database for soybean breeders in the North Central Region.

All known genotypic/phenotypic data compiled and statistical analysis/quality control underway. Eighty percent of weather/soil data compiled.

Phenotypic data has been fully compiled and integrated into a relational database. Genotypic data remains pending due to delays in obtaining seed from breeding programs. Available environmental data has been evaluated for quality and collection has begun. We believe the temperature data is reliable, but the precipitation data is less reliable, which will possible cause problems with modeling. We will continue to work at this to find better sources of data.

Eighty percent of historical collection genotyped with GBS. One hundred percent of historical collection (1/3) genotyped with 6K array. Prelim entries for field season 2016 genotyped with 6K array.

Of the 3,784 lines grown in URTs from 2004-2016, 22% (818 lines) were proprietary or seed no longer exists, 41% (1,545 lines) have not been located by breeding programs (no response or pending search), and 37% (1,421 lines) have been located and are in the process of being genotyped. The lines included in this set are nearly complete for years 2014-2016, with fewer historical lines located. 375 entries from the 2015 URT were genotyped with GBS. The sequence depth obtained was deemed insufficient quality, and a decision was made to shift available funds to genotyping with the 6K array. Of the 1,421 lines located, 960 are in process for genotyping with 6K chip at Michigan State, and should be completed in early November 2016. An additional set of 6K chips to assay 752 lines were ordered at Minnesota to genotype the remaining 461 lines and any additional lines that may be located. These lines will be genotyped by December 2016.

Merging GBS data and 6K will be tested on data available to date.
Because the GBS data collection was abandoned as described above, this goal as stated is no longer specifically necessary. Nevertheless, it would be advantageous to be able to impute higher density SNPs from re-sequencing work conducted by the Nguyen Lab at U. of Missouri. The Nguyen Lab has discovered ~5 million SNPs

through re-sequencing ~400 soybean accessions. About 5 million SNPs have been projected onto the entire USDA Soybean Germplasm Collection by leveraging the available 50K SNP data on the collection. Accuracy of projection was 80% or higher based on masking studies. This new resource holds potential to increase power and resolution of QTL mapping using data from the collection. We will use these same techniques to impute the 5 million SNPs onto the URT lines genotyped with the 6K SNP array as part of this project.

1-4   Eighty percent of the genomic prediction training set completed and initial methods being tested on data from the 2016 Uniform Tests.
Due to delays in obtaining seed from historical lines, genotyping and subsequent prediction modeling has been delayed. Genotyping is currently underway and a full training set should be available by the end of 2016.

Goal 2: Identify large-effect alleles from exotic sources that improve yield potential of elite U.S. soybean lines.

First year of yield, agronomic, and quality data collected.
A set of 981 diverse, re-sequenced accessions were grown by 12 collaborators. The earliest maturing accessions were grown by MN, and the latest maturing accessions were grown by GA, with many states growing intermediate maturity group accessions in between. Notes on flowering time, days to maturity, plant height, shattering scores, lodging scores have been collected. Harvest data collection is currently underway at a select number of sites. At the very least, this will be a good year for seed increases, and observations on lodging and shattering will indicate which accessions can be evaluated for yield in 2017.

Sequencing of 11 lines completed.