10:14 am, 14 Feb 2.

Nebraska Soybean Board
FINAL Research/Extension Education Report Form

Note: Submit this report no later than 60 days after the NSB-funded project officially terminates.

This post-project 60-day time-frame will allow the Lead PI/Extension Educator time to complete any final data analysis and a final technical report, plus the drafting of any articles for submission to scientific journals.
This completed report will be provided to the National Soybean Checkoff Research Database: soybeanresearchdata.com.

Project # and Title:

#1742 Increasing yield and seed composition stability through diverse germplasm and genomic selection

PI / Extension Educator:

Dr. David Hyten

Co-PI's / Co-Extension Educator's:

Dr. George Graef

Project Date ***(Including No-Cost Extension)*****:** 10/1/19 **to** 9/30/22

Total Budget for Project: $ 306,990

1. Briefly State the Rationale for the Research.

Releasing a new cultivar for agricultural production requires that the experimental variety has consistent high yields across a large geographical region. Many experimental varieties are eliminated in advanced yield trials because they may yield at the top of some yield tests but near the bottom of other tests. Enhancing a variety's ability to have stable seed composition for protein and oil across environments will also add value to soybean. The soybean genome has genes that will enhance a variety's ability to have consistent yield and seed composition across diverse environments. This project builds upon the previous yield stability project to discover new yield stability genes in diverse germplasm and to develop genomic selection to select for increased yield and seed composition stability in early generation breeding material. Early selection with a large number of experimental lines will help maximize the potential to enhance yield and seed composition stability for cultivars being developed for Nebraska soybean production areas. Discovering new gene for yield stability not present in current breeding material germplasm and developing an early selection method, will enable us to enhance these key traits for soybean producers.

2. Research Objectives: (copy from original proposal, but keep in a brief bullet format).

This research project's main objective is to develop methods that improve soybeans ability to produce consistent yield and seed protein and oil concentration across different Nebraska fields and across years.

Objective 1: Discover new loci that confer yield stability.

Objective 2: Develop and test genomic selection methods for predicting yield and seed composition stability in early generation breeding material.

3. General Approach Used and (if applicable) the Nebraska Test Location.

Objective 1: We used previously generated yield and genotype data on diverse material and on the data set generated for objective 2 to map GxE loci for yield and for protein and oil. To determine selection being acted upon GxE loci we resequenced 200 elite lines to a 9x coverage and applied the population statistics methods of Tajima's D and diversity measurements. We also compared the populations to wild soybean, landrace and other elite populations using Fst do understand how GxE regions have diverged in this elite population.

Objective 2: We previously collected yield and seed protein and oil composition data on 213 elite breeding lines across 13 environments to discover yield and seed composition vQTL. This data was used as our training population. We selected 213 independent lines to make up the validation population which was yield tested in 11 different environments over three years located in Nebraska, Iowa, and Missouri. The 213 lines were skimmed sequenced and genotyped with the 1k Soy SNP MIPs set. Genomic prediction was performed for multiple stability measurements for yield, protein, and oil.

4. Describe Deliverables & Significance Attained for Each Research Objective.

Objective 1: There were no major yield stability genes discovered. The number of genes and low effect of the genes again points to genomic selection being the most effective way to integrate yield stability genes into elite germplasm. Of the QTL that have been discovered for yield stability or GxE effects, we have observed that the discovered QTL were positive in at least one environment while negative in another. This led to the hypothesis that these QTL might not be under positive selection by breeders. We tested this hypothesis and found that loci significant for GxE effects have higher diversity than loci that contribute mostly additive effects. This higher diversity is indicative of balancing selection. This lack of positive selection of GxE effects means that current genomic selection models are not accounting for these GxE loci and the models need to be adjusted to take into account that the alleles of GxE loci need to be selected so that they don’t favor one set of environments over the others.

Objective 2: Genomic prediction for yield stability has yielded useful results, as the area is relatively unexplored. We assessed 13 stability measures (5 AMMI, 3 parametric dynamic, 5 non-parametric dynamic), two different marker selection methods (high-effect GWAS SNPs vs. random SNPs), SNP marker densities ranging from 500 to 15,000 markers, three different prediction scenarios (across population TP > VP, TP cross-validation, VP cross-validation), and three different types of prediction accuracy.

We found that across population prediction (applied scenario) and cross-validation (within TP and within VP) yielded similar prediction accuracies. Contrary to our expectation, this indicates that predicting across breeding populations separated by 3 years in the same program does not result in a noticeable decrease in prediction accuracies compared to prediction within a population. The most applicable measure of accuracy we used is rank coincidence. This measure, not commonly used in genomic prediction studies, indicates the proportion of individuals in the same half of both predicted stability ranks and observed stability ranks. Overall, prediction accuracies were low for AMMI and parametric dynamic stability with rank coincidence between 0.45 and 0.55, or similar to expected accuracy when randomly selecting. This was expected, as yield stability is a complex trait and these two stability types are more complex in their calculation. With non-parametric (rank-based) stability measures, prediction

4. Describe Deliverables & Significance Attained for Each Research Objective. (continued)

accuracy across populations was consistently higher. When predicting these measures across populations, we found that rank coincidence values ranged from 0.61 to 0.65. This means that roughly 65% of individuals predicted to be in the top half of stability in a population are in the top half of observed stability. If we assume random selection coincidence to be 0.5, this is roughly a 30% increase in rank coincidence accuracy compared to random. Genomic prediction studies normally focus on the correlation between predicted and observed values. For a trait like yield stability, rank coincidence has potential to increase selection accuracy and provide a valuable tool for breeders with concern for yield stability in the early stages of yield trials. We also found that SNPs selected as highest-effect from a genome-wide association analysis (GWAS) were much more effective in predicting yield stability than SNPs selected at random. Furthermore, prediction accuracies began to plateau at 1,500 markers, indicating negligible increase in prediction accuracy when 2,000+ markers were used for prediction. These are controllable aspects of genomic prediction that give a starting point for application of genomic prediction for yield stability.

Genomic prediction for yield stability has potential to increase efficiency in breeding programs. Yield stability measurement requires data from multiple years and numerous locations to observe in the field, and genomic prediction can predict a phenotype before ever being observed in the field. Overall, the power of this study lies in the use of two real-world breeding populations to test the application of methods in an applied breeding setting. This range of inference is uncommon for genomic prediction studies, especially when yield data across many environments and multiple years are needed for stability assessment of two separate populations. We found that better-than-random accuracies can be achieved for genomic selection of yield stability in an applied breeding scenario. Changes in methods regarding prediction models and the number of environments for training and validation sets, may give rise to further increases in prediction accuracy and applicability.

5. List where the Project Research Results/Findings were Publicized.

We plan to submit two papers for journal peer review under the genomic prediction study and one paper describing the selection forces acting upon GxE QTL.

Note: The Final Report comprised of the above listed items must be kept to THREE PAGES.

A Technical Report of no more than TEN PAGES (preferably fewer) can be appended to this report.

Submit the form with the following file name format: #XXX_FINAL_Project Title_LastName

Please submit this completed form with attached files to the Agriculture Research Division, jmcmahon10@unl.edu, based on the reporting schedule given to you.

If you have any questions, please call Jen McMahon at the Agricultural Research Division (402) 472-7082.

Please click to attach technical reports, etc. Attach files Please check your information before submitting the form. Submit by Email

Please note: Attach files button may not work in some versions of Acrobat Reader. You may need to save a copy of this form and then attach files to the copy.