Nebraska Soybean Board
Year-End Research Findings Report

Please use this form to summarize the practical benefits of your research project and what has been accomplished. Your answers need to convey why the project is important and how the results impact soybean production.

Project Title: *Implementing and Vetting Genomic Selection in Soybean*

Contractor & Principal Investigator: *George Graef*

Please check/fill in appropriate box: ☐ ***Continuation research project***
■ ***Year 2 of 3 research project (for example: Year 1 of 2)***

1. ***What was the focus of the research project or educational activity?***

This research project evaluates different aspects of genomic selection, specifically related to soybean, and aims to identify ways to optimize implementation in a soybean breeding program to improve genetic gains for yield and other important traits.

We are evaluating primarily three main aspects of genomic selection related to its implementation in the Nebraska soybean breeding program: (1) Training population size and constitution relative to prediction accuracy in the target population, (2) Comparison of phenotype selections, genotype selections, and random selections in two large soybean populations (>800 lines) to evaluate effects of training population design on prediction accuracy; evaluate different prediction models, prediction of yield and prediction of other traits like seed composition, maturity, height, lodging; incorporation of genotype-environment interactions in the models; (3) Comparison of lines derived from intermating among parental lines selected based on phenotype information vs. lines derived from parents selected based on genomic predictions.

2. ***What are the major findings of the research or educational activity?***

The main points from our work to date indicate that (1) The highest prediction accuracy we observed in this study is about 0.5, (2) Prediction accuracy can be improved by having greater genetic relationship between the training population and target population, (3) A larger training population does not necessarily improve prediction accuracy.

Some questions still to answer with incoming and future data and analyses include effects of heterogeneous variances among populations, inclusion of genotype-environment interaction effects in the models, and prediction of multiple traits.

3. ***Briefly summarize, in lay terms, the impact your findings have had, or will have, on improving the productivity of soybeans.***

The potential is that we will increase the rate of genetic gain for yield in soybean. That means that we will make faster progress in developing higher yielding soybean varieties for producers. At least three aspects of genomic selection help to achieve that acceleration in genetic gain: (1) **Cost**: The cost of a single yield test plot from start to finish is about $25. Before being selected as a parent in a breeding program, a line has at least 12 replications of yield data from multiple environments = $400 cost per line. A single sample of DNA analysis currently costs ~ $35. So our initial calibration of yield information with genotype information involves costs for both the phenotyping and the genotyping, but application of that calibrated information – applying the selection model to new genotype information – has only the genotype cost associated with it, or maybe the genotype costs along with much reduced phenotyping. So now instead of $400 per line, we could spend $35 per line and evaluate 11 or 12 times as many lines to find superior genotypes. Then we can concentrate further yield testing on those best genotype-selected lines. (2) **Time:** There are different ways to implement genomic prediction and selection in a breeding program, but one of the main impacts will be on reducing cycle time. So the number of years between a cross and identification of superior lines can be reduced by one or two years. A 2-year reduction of a 6-year breeding cycle is a 33% faster cycle time. (3) **Increase selection intensity and accuracy:** With good genotype information and a good model, we can combine genotype and phenotype information to improve our ability to identify potentially superior lines earlier in the program. So we can then concentrate our yield testing efforts on those best lines. In addition, the cost of genotyping will decrease, so the number of genotypes we can evaluate from a population or over populations will increase for a given fixed cost. That increases our ability to evaluate more recombinants from crosses and improves our potential to identify more superior genotypes than we would compared with the smaller population sizes used in replicated, multi-location plot testing.

4. ***Describe how your findings have been distributed to (a) farmers and (b) public researchers. List specific publications, websites, press releases. etc.***

Our initial publication (Jarquin et al., 2014, BMC Genomics 15:740) was the first to evaluate use of genotyping-by-sequencing data for genomic prediction in soybean breeding populations. This study follows up on that to implement genomic selection in the breeding program. We also have presented posters at national and regional meetings and are working with other regional and national projects on implementation of genomic selection in soybean breeding.

5. ***Did the checkoff funding for your project leverage any additional state or Federal funding? Please list sources and dollars approved.***

We developed a large, multi-institutional grant proposal that will extend some of the findings from this and other genomic selection projects to the broader public soybean breeding efforts in the north central region. The project is led by Dr. Aaron Lorenz at the University of Minnesota, funded by NCSRP.