

## Genomewide Selection in Maize, Down the Road

Sofía Brandariz – Advised by Dr. Rex Bernardo

Maize (*Zea mays* L.) was domesticated in Central America and Mexico around 10,000 years ago from teosinte (Doebley et al., 2006). Since its domestication, maize has been widely cultivated and has been under continuous selection during this time. Half of the world's maize is produced in the Americas (FAOSTAT, 2016), and most of the U.S. production occurs in the Midwest “Corn Belt.” In Minnesota, mean grain yields have increased from around 118 bushels per acre in 1995 to 175 bushels per acre in 2005 (Ye, 2008), and to 188 bushels per acre in 2015. Maize in the U.S. is used mainly for feed and for fuel ethanol (USDA, 2016). With a growing need for maize, breeders are looking for ways to increase efficiency in their programs.

Maize breeding usually involves crossing two inbreds (A and B) and evaluating the testcross performance of the progeny from the A/B cross (Hallauer, 1990). Recent advances in genotyping (Davey et al., 2011) have allowed genomewide selection for important traits in maize. Genomewide selection involves the use of training populations. These are populations for which both phenotypic and genotypic data exist and which can be used to predict the phenotypic performance of test populations using only their genotypic data (Meuwissen et al., 2001). Genomewide selection can reduce the amount of phenotyping needed and increase the genetic gain per unit time (Heffner et al., 2010).

Previous studies have shown that genomewide selection is most effective when the training population is representative of the test population (Schulz-Streeck et al., 2012; Riedelsheimer et al., 2013; Jacobson et al., 2014). The General Combining Ability (GCA) model, where an A/\* cross (\* being a parent from the same heterotic group as A) and an \*/B cross are used as training populations to predict an A/B cross, was found to be useful in elite germplasm in maize. On average, genomewide selection led to around 85% of the gains that eventually would have been obtained with phenotypic selection (Jacobson et al., 2014). In addition, the predictions from the GCA model were improved by using imputed Single Nucleotide Polymorphism (SNP) in maize (Jacobson et al., 2015).

Performing routine genomewide selection within a breeding program will lead to two main consequences. First, only a subset of the lines in a biparental cross will be later phenotyped. If that cross is used as part of the training population for predicting the performance of lines in a future cross with either of the parents, the training population would be a selected subset of lines. Second, introducing new lines into the breeding program will lead to having information on only one parent (either A or B) to predict an A/B cross, or having information on related lines from the same heterotic group but on neither parent (Same Background model, \*/\*). Furthermore, breeders have mostly used genomewide selection models that consider the additive effects of individual alleles but not the interactions among them (epistasis), with few exceptions (Jarquín et al., 2014; Jiang and Reif, 2015). These two consequences plus epistasis in the model could affect the accuracy of the predictions.

Routine genomewide selection could also affect the rate of loss of genetic diversity in a breeding program. Genetic diversity is needed to maintain long-term gains in breeding programs (Condón et al., 2009), particularly when faced with increased biotic or abiotic stresses and the effects of climate change (Gepts et al., 2006; Gutiérrez et al., 2009). However, studies on the effects of routine genomewide selection on genetic diversity are limited.

This research considers genomewide selection in 969 maize biparental crosses. Monsanto provided phenotypic and marker data for these populations, which were evaluated for grain yield ( $\text{Mg ha}^{-1}$ ), moisture ( $\text{g H}_2\text{O kg}^{-1}$ ), and test weight ( $\text{kg hL}^{-1}$ ) in multiple locations from

2001 to 2008. We selected 27 F<sub>2</sub> populations as the test populations, based on having at least four A/\* and \*/B crosses, a minimum population size of 50 lines, and an entry-mean heritability significantly greater than zero for each trait. The parents of the populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers, whereas the progeny were genotyped with fewer markers. The SNP data for the progeny were then imputed from the parental SNP data. This research will focus on the effect of the following objectives on prediction accuracy, selection response, and genetic diversity: (i) selection in the training population; (ii) level of similarity between the training and test population; and (iii) epistasis in the prediction model.

Preliminary results for the first objective showed that having phenotypic information on only a selected fraction of the lines significantly reduced ( $p=0.05$ ) the response to selection and prediction accuracy. However, augmenting the training set with the five poorest lines nearly restored the prediction performance. Prior selection in the training population did not increase the genetic similarity of the best lines in the test population. Preliminary results for the second objective showed that for predicting the performance of lines derived from an A/B cross, a training population that included lines with A as one parent (A/\*) and lines with B as one parent (\*/B) led to higher prediction accuracies and responses to selection. Using the Same Background model significantly reduced the predictions performance even with an increased number of training populations compared to the GCA model. Results on epistasis in the prediction model are forthcoming.

## References

- Condón, F., D.C. Rasmusson, E. Schiefelbein, G. Velasquez, and K.P. Smith. 2009. Effect of Advanced Cycle Breeding on Genetic Gain and Phenotypic Diversity in Barley Breeding Germplasm. *Crop Sci.* 49: 1751-1761.
- Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12: 499–510.
- Doebly, J.F., B.S. Gaut, and B.D. Smith. 2006. Review The Molecular Genetics of Crop Domestication. *Cell* 127: 1309–1321.
- FAO. 2016. FAOSTAT. (online). Rome. Consulted Oct, 26<sup>th</sup> 2016. Available at <http://faostat.fao.org>.
- Gepts, P., J. Nelson, C. Qualset, and H. Shands. 2006. Plant Genetic Resources Conservation and Utilization: The Accomplishments and Future of a Societal Insurance Policy. *Crop Sci.* 46: 2278–2292.
- Gutiérrez, L., J.D. Nason, and J.-L. Jannink. 2009. Diversity and Mega-Targets of Selection from the Characterization of a Barley Collection. *Crop Sci.* 49: 483-497.
- Hallauer, A.R. 1990. Methods used in developing maize inbreds. *Maydica* 35: 1–16.
- Heffner, E.L., A.J. Lorenz, J. Jannink, and M.E. Sorrells. 2010. Plant Breeding with Genomic Selection : Gain per Unit Time and Cost. *Crop Sci.* 50: 1681–1690.
- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2014. General Combining Ability Model for Genomewide Selection in a Biparental Cross. *Crop Sci.* 54: 895-905.
- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2015. Marker Imputation Before Genomewide Selection in Biparental Maize Populations. *Plant Genome* 8: 1-9.
- Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15: 740.
- Jiang, Y., and J.C. Reif. 2015. Modeling Epistasis in Genomic Selection. *Genetics* 201: 759–768.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, and A.E. Melchinger. 2013. Genomic Predictability of Interconnected Biparental Maize Populations. *Genet. Soc. Am.* 194: 493–503.
- Schulz-Streeck, T., J.O. Ogutu, Z. Karaman, C. Knaak, and H.P. Piepho. 2012. Genomic Selection using Multiple Populations. *Crop Sci.* 52: 2453–2461.
- Ye, S. 2008. Economic impact of the corn and ethanol industry. Agricultural Marketing Services Division. Minnesota Department of Agriculture.
- U.S. Department of Agriculture. 2016. Consulted Oct, 26<sup>th</sup> 2016. Available at <http://www.usda.gov/>.