

Predicting Genetic Variance from Genomewide Marker Effects in Maize

Emmanuel Adeyemo

M.S. Student Advised by Rex Bernardo

Maize (*Zea mays* L.) is the most produced grain in the world and is grown in more than 160 countries. The United States produces close to half of the world's harvest, while other top producing countries include China, Brazil, France and South Africa. Maize contains about 72% starch, 10% protein and 4% oil, provides many B vitamins and serves as a good source of fiber. These attributes have made maize a staple food in many regions of the world, constituting about one-third of the calorie intake in Latin America, the Caribbean and sub-Saharan part of Africa. In most developed countries like the United States, maize is mostly used as feed for livestock, and a large proportion is also used for industrial purposes such as syrup and ethanol production (Ranum et al., 2014)

In a typical maize breeding program, many breeding populations are developed each year. Most of the breeding populations are discarded after preliminary evaluations. A metric that can predict the potential of a population, before the population is made, would increase the efficiency of breeding programs. The *usefulness criterion*, U_p , was proposed by Schnell and Utz (1975), as a guide to identifying promising crosses. The U_p , which is the expected genotypic mean of superior progenies within a cross is a function of the mean and genetic variance of that population. The U_p increases when the mean or genetic variance is increased or both the mean and genetic variance increased.

For U_p to be predicted before the breeding population is made, the mean and genetic variance needs to be predicted first. Predicting the population mean is straightforward; quantitative genetics theory and empirical results have shown that the mean of a breeding population can be predicted from the mean performance of the parents used to form the population (Melchinger et al., 1998, Utz et al., 2001,). Breeders have found this to be the method of choice as phenotypic data needed to make the predictions are available from previous tests.

While the mean is easily predicted, predicting the genetic variance in a breeding population is much more difficult. Several approaches have been used to predict genetic variance, including coefficient of coancestry and marker-based genetic distance between the two parents used to develop the breeding population (Melchinger et al., 1998, Bohn et al., 1999, Utz et al., 2001, Hung et al., 2012). Overall, these criteria have been ineffective for predicting the genetic variance. A key limitation of these methods is that the prediction applies to all traits meaning that the pair of parents with the largest coefficient of Coancestry or the highest genetic distance are predicted to have the largest genetic variance for multiple traits such as yield, maturity, flowering date, etc. Breeders know from experience that a given cross does not necessarily show the largest genetic variance for all the traits.

A procedure effective for predicting the genetic variance in each cross would likely require modeling the segregation of progenies within each cross. Bernardo, (2014); proposed that if single nucleotide polymorphism (SNP) data as well as the phenotypic data of the parents are available, it

is possible to (i) estimate the trait effects associated with each marker, (ii) create virtual populations that correspond to simulated progenies within a cross, (iii) and estimate U_p of each cross from the virtual populations. To date, however, the usefulness of virtual populations for estimating U_p has not been validated from controlled studies.

Objective: My M.S. thesis research aims to determine if the mean, genetic variance and U_p predicted from genomewide marker effects correlates with empirical mean, genetic variance and U_p estimated from field data.

Experimental procedure: A subset of eight crosses were selected from the historical Minnesota and non-Minnesota inbreds studied by Schaefer and Bernardo (2013). The training population, used to obtain the genomewide marker effects in the Schaefer and Bernardo (2013) study, comprised 284 lines that were genotyped at 28,626 SNP markers. Marker effects were estimated for all markers by ridge regression – best linear unbiased prediction in the rrBLUP package in R. The predicted mean, genetic variance and U_p for the eight selected crosses were obtained beforehand from the analysis done by Schaefer and Bernardo (2013). The eight populations were chosen so that the mean and genetic variance differed among the populations for different traits.

In 2017, the eight populations were evaluated for plant and ear height, and growing degree days to silking at St Paul, Waseca and Lamberton. Six out of the eight populations consisted of 144 F3 lines while the remaining two populations consisted of 120 F3 lines, for a total of 1104 F3 lines. The lines were evaluated in a randomized complete block design with replications at each location. Due to the large number, the 1104 lines were divided into 24 sets, each set comprising six lines from the six populations that contained 144 lines and five lines from the two populations that contained 120 lines. This resulted in each set having a balanced representation of lines from each of the eight populations and elimination of confounded differences between blocks and populations. The plots were 4m long and 0.76m apart. The plant density was 66,790 plants ha⁻¹.

Preliminary results: The correlation between the predicted mean (from virtual populations) and the observed mean (from field data) was 0.91 for plant height, 0.85 for ear height and 0.80 for silking date. For plant height, the range of values was 50.0m in the shortest population to 68.3m in the tallest population. This indicated differences in genetic variance among the eight populations. Detailed analysis of the genetic variance and U_p will be conducted in the upcoming months. If the approach for predicting genetic variance and U_p is successful, breeders will be able to concentrate their limited resources on the most promising populations and improve efficiency of breeding programs.

References:

1. Bernardo, R. 2014. Genomewide Selection of Parental Inbreds: Classes of Loci and Virtual Biparental Populations. *Crop Sci.* 54:2586-2595. doi:10.2135/cropsci2014.01.0088
2. Bohn, M., H. F. Utz, and A. E. Melchinger. 1999. Genetic Similarities among Winter Wheat Cultivars Determined on the Basis of RFLPs, AFLPs, and SSRs and Their Use for Predicting Progeny Variance. *Crop Sci.* 39:228-237. doi:10.2135/cropsci1999.0011183X003900010035x

3. Hung, H.-Y., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., ... Holland, J. B. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity*, *108*(5), 490–499. <http://doi.org/10.1038/hdy.2011.103>
4. Melchinger, A. E., Gumber, R. K., Leipert, R. B., Vuylsteke, M., & Kuiper, M. (1998). Prediction of testcross means and variances among F3 progenies of F1 crosses from testcross means and genetic distances of their parents in maize. *Theoretical and Applied Genetics*, *96*(3), 503–512. <https://doi.org/10.1007/s001220050767>
5. Ranum, P., Peña-Rosas, J. P. and Garcia-Casal, M. N. (2014), Global maize production, utilization, and consumption. *Ann. N.Y. Acad. Sci.*, *1312*: 105–112. doi:10.1111/nyas.12396
6. Schaefer, C. M., & Bernardo, R. (2013). Genomewide association mapping of flowering time, kernel composition, and disease resistance in historical minnesota maize inbreds. *Crop Science*, *53*(6), 2518-2529. DOI: [10.2135/cropsci2013.02.0121](https://doi.org/10.2135/cropsci2013.02.0121)
7. Schnell, F.W., and H.F. Utz. 1975. F1-leistung und elternwahl euphyder zuchtung von selbstbefruchtern In: Bericht uber die Arbeitstagung der Vereinigung Osterreichischer Pflanzenzucher. BAL Gumpenstein, Gumpenstein, Austria. p. 243–248
8. US Department of Agriculture National Agricultural Statistics Service
9. Utz, H.F., M. Bohn, and A.E. Melchinger. 2001. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. *Crop Sci.* *41*:1470–1478. doi:10.2135/cropsci2001.4151470x