Genetic diversity of some quality protein maize lines(genotypes)as revealed by molecular markers

M.Yusuf

Department of Plant Science, Faculty of Agriculture/Institute for Agricultural Research,Ahmadu Bello University, Zaria, Nigeria

Introduction: Unmasking of genetic diversity in maize breeding populations can greatly assist in designing appropriate breeding strategies. Six quality protein maize inbred lines (parents) together with their respective progenies were characterized using molecular (RAPD) markers primarily to determine the genetic diversity within the population and to establish the genetic relationship between the parents and their progenies at molecular level. The primer; OPERON- AF 13 gave the highest number of DNA polymorphic bands (9), suggesting that it could be used as an effective marker in more detailed genetic studies involving these lines and possibly other maize lines. The results of the Dendrogram revealed the relationships between the parents and their respective single cross hybrids with 33.33% of the progenies resembling more like the female parents and 60.67% resembling more like the male parents respectively, this may indicate less maternal effect within the maize population under study which is often desirable in breeding work.

A set of 346 diverse hybrid lines were grown as a structured genome-wide association study (GWAS) to assess aflatoxin resistance, drought tolerance, and other agronomic traits such as yield in 2011 and 2012 in College Station, TX (Barrero-Farfan et al., 2015). These lines originated from a subset of the USDA—Goodman maize association panel (Flint-Garcia et al., 2005), as well as the southern subtropical Williams/Warburton panel (Warburton et al., 2013). This panel was crossed to two isogenic lines of Tx714, a high-yielding, southern United States bred stiff-stalk line that is more than 95% identical to its stiff-stalk relative, B73 (Romay et al., 2013). The Tx714 isogenic hybrids differed only for which of two maize lipoxygenase genes were mutated (De La Fuente et al., 2013; Park, Kunze, Ni, Feussner, & Kolomiets, 2010). Each isogenic hybrid was grown under one or two experimental conditions (well-watered [WW] and limited irrigation, or water stress [WS]) with two replicates in a randomized complete block design, where seed was available. A full description of the hybrids and experimental design is in Barrero-Farfan et al. (2015). Since isogenic hybrids did not differ in grain yield, the data from both were combined and treated as the same hybrid. This population was used for all model training and for validation.

Other elite hybrid breeding trials in the Texas A&M breeding program were also used as independent validation sets for evaluating prediction robustness. These trials, grown in 2011, consisted of breeding relevant hybrids with no known or expected relatedness to the original material (GWAS hybrids) presented above, most being sub-tropical derived lines crossed to U.S. commercial stiff stalk hybrids (Murray et al., 2019). In addition to grain yield, these tests were grown to assess aflatoxin (AF) resistance in the hybrids. Material from within the program represented four tests, and are referred to in this study as 1AF, 2AF, 3AF, and 4AF. Two other tests, which included breeding material from other programs, were also assessed in 2011. The Southeast regional aflatoxin trials (SERAT; Wahl et al., 2017), and the germplasm enhancement of maize (GEM) lines.

Three of these validation tests (3AF, SERAT, and GEM) were grown in College Station, TX, under similar conditions, 1AF and 2AF were grown in both Weslaco and College Station, TX, and 4AF was grown in Corpus Christi, TX. All together, these datasets combined represent 200 pedigrees across three Texas locations from 2011 (679 samples total), most of which were breeding material of commercial hybrid checks. Grain yield and NIRS were collected for each breeding test on a plot basis. This set of tests was used as a practical validation for how broadly the trained models could be used in a breeding program.

Results: The results are presented in two main sections. The first section presents the predictions under the PLSR and LM methods. The second section describes the predictions based on the phenomic selection models (NIRS BLUP and functional regression), for single environment and G × E multi-environment approaches.

Cross validated yield predictions from models using PLSR within the same years demonstrated a high Pearson's correlation to maize grain yield on a plot basis in the testing data (r = .84; Table **1**, PLSR 1). The model was successful in using spectra alone to predict the yield of a sample on a plot basis, evaluated by error in prediction, RMSEP. With only spectral data, the PLSR model using 15 PLS components predicted yield with a RMSEP of 163.67 g m^{-2}.

To investigate if kernel compositional traits were correlated with yield, NIRS composition predictions were used to predict yield using simple LM. A PLSR model as above was run but using only the 2,155 samples with composition predictions to build the training and testing sets (Table **1**; PLSR 2); this had an r of .82. Looking at individual components, crude protein was the only measured component with a strong correlation with yield (r = .58) (Table **1**; LM 1). Both starch and fat had very low correlations ($r \leq .17$, Table **1**; LM 2–3). Combining all compositional predictions into a model to predict yield improved predictions over protein alone, but only slightly, yielding an r of .64 (Table **1**; LM 4).

Next, PLSR scenarios were developed to predict yield on samples from a year unknown to the model, on known or unknown hybrids (see Table **2**). Results showed that a model built on 2011 predicted unknown hybrids in 2011 better than a 2012 model predicting unknown hybrids in 2012 (Table **2**; PLSR 3–4). Results also showed that a model trained on 2011 predicted 2012 well, and better than the reverse (PLSR 5 and 8); even though the predicted hybrids were known to the model in both cases. In predictions between years, the best performing model was trained on all available samples from 2011 to predict 2012 (PLSR 6). The PLSR 6 had more training data than PLSR 5, as this training data represented all available 2011 data (including the USDA and other breeding populations from 2011). We also investigated the ability of a model to predict only unknown hybrids from a new year (PLSR 7), which was comparable to when most of the hybrids were already known from a previous year (PLSR 5) cross-validation under CV0, CV1, and CV2 schemes.

Conclusion: This study supports previous evidence from wheat that the use of grain spectral data is useful for predicting grain yield, but for the first time, shows that this approach is valid in maize. For this reason, these results offer promise for studies in other crops, and will likely guide users to choose a statistical analysis method that best suits their goals. Partial least squares regression is routinely used with NIRS and showed promise for predicting a non-compositional trait on a plot basis, including on unrelated material. However, functional regression generally performed better, within a set of similar germplasm and may offer even more value for breeders once protocols are established.

There is reason to suggest that building NIRS prediction models including multiple years will further strengthen the prediction accuracy. Future experiments or applications of these methods should aim for more than 2 yr of data to predict subsequent years, especially focused on higher-stress environments. Low cost of implementation and evidence provided in this work suggest that breeders interested in this technology should scan as much relevant genetic material as possible when building calibrations, and continually add and update the models just as in genomic selection.

Keywords: Genetic diversity, Quality protein maize inbred lines, Molecular markers, Dendrograms.