Texas Corn Producers Board

Project Final Technical Report, December 14, 2017

(January 1 – June 28, 2017 and June 29 – December 14, 2017)

Title: Enhanced Corn Breeding for Grain Yield Using the Genes Controlling the Targeted Traits

Project Number: M17012827

Period of Performance: March 1, 2017 – December 15, 2017

Hongbin Zhang

Principal Investigator: Dr. Hongbin Zhang, Professor of Plant Genetics, Genomics and Molecular Breeding, Department of Soil and Crop Sciences (SCSC), Texas A&M AgriLife Research, College Station

Co-Principal Investigator(s): Dr. Wenwei Xu, Professor of Corn Breeding, SCSC, Texas A&M AgriLife Research, Lubbock

Summary:

Current corn breeding that is based on the phenotypic performance of agronomic traits, such as grain yield, has significantly contributed to Texas corn production in the past years. This breeding method is known as the phenotypic selection (PS) method. However, it has been currently subjected to a great challenge: how to break the current stagnation of corn grain yield, and continuously and substantially improve corn grain yield, thus promoting corn production. This project is to address this issue through development of gene-based breeding (GBB), a novel technology that is extremely powerful and efficient for enhanced and accelerated corn inbred and F_1 hybrid breeding by directly and effectively using the genes controlling the targeted trait(s), with an emphasis on grain yield.

Corn breeding often includes two steps, inbred breeding and F_1 hybrid selection. In this period of the project, we first developed a GBB system that is designed to breed for desirable corn inbreds (Inbred GBB) by directly manipulating the genes controlling grain yield (*ZmINGY*). Using the Inbred GBB, we were able to accurately predict the grain yields of inbred lines; therefore, an inbred containing desirable alleles could be directly selected, without need of field trials as usually done for the current corn inbred breeding. Then, we developed a GBB system for superior F_1 hybrid grain yield breeding (F_1 hybrid GBB) by directly manipulating the genes controlling F_1 hybrid grain yield (*ZmF1GY*). Using the F_1 hybrid GBB, we were able to accurately predict the grain yields of F_1 hybrids from their inbred parents; therefore, an F_1 hybrid having the highest grain yield could be directly and readily selected from all possible crosses of a set of inbred lines before making the crosses and performing associated field trials, as done for the current F_1 hybrid breeding. In comparison, although it is theoretically possible, it is practically very difficult, if not infeasible, to make all 4,950 possible F_1 hybrid crosses for a set of 100 inbred lines and perform replicated field trials to identify the best F_1 hybrid with the highest grain yield using the current corn breeding method. Because GBB provides the most accurate and comprehensive information about the genes controlling the targeted traits, their favorable alleles and effects on the targeted trait(s), their heterotic genotypes and their most efficient networks, breeders can accurately select germplasm lines containing favorable alleles desirable for their breeding goals as breeding crossing parents, design crosses of the most desirable, and select those best progeny containing the largest numbers of favorable alleles, the largest numbers of heterotic genotypes (for heterosis) and the most efficient gene x gene interaction networks for the targeted trait(s). Therefore, GBB allows effectively and continuously accumulating the favorable alleles and heterotic genotypes from different germplasm lines into the current superior cultivars, thus providing an extremely powerful and efficient tool for breeders to continue substantial crop improvement and to further enhance crop production. The GBB systems developed in this project promise to revolutionize the current corn breeding methods, thus likely promoting a green revolution in corn production. Moreover, this project has helped develop three related corn research proposals (USDA National Institute of Food and Agriculture, Texas A&M AgriLife Crop Improvement and Illumina), of which one has been awarded and two are currently in pending. These related projects will significantly facilitate Texas corn breeding research & development to realize the green revolution in corn production. In addition, we have identified several F_1 hybrids that had grain yields 10% higher than those of the checks at College Station, Lubbock and/or Halfway, TX, which could be potentially used for commercial corn production in Texas.

In 2018, we will, as originally proposed in this project, first screen all 4,095 possible F_1 hybrids of our 91 inbred lines and all 1,711 possible F_1 hybrids of the 59 inbred lines developed by the Texas A&M AgriLife Research Corn Breeding Program using the F_1 hybrid GBB system developed in this project to identify the best F_1 hybrid(s) with the highest grain yield among all the 5,806 F_1 hybrids. Then, we will further verify the selected best F_1 hybrid(s) through a replicated field trial and develop 1 – 2 superior F_1 hybrids that could be potentially released for commercial corn production in Texas and probably, nationwide. These continued experiments are essential to further verify the GBB systems developed in this period of the project and to introduce the GBB systems into the Texas Corn Breeding Program for enhanced and accelerated inbred and F_1 hybrid corn breeding, thus dramatically benefiting Texas corn producers in corn production.

I. Project Research Objectives:

This proposal addresses the **Priority Area of Corn Breeding for Yield,** with aims to introduce and apply gene-based breeding (GBB), a new breeding method recently developed by our group and shown to be extremely efficient particularly with complex multi-gene traits such as grain yield, to the Texas Corn Breeding program. In particular, the **following research objectives** will be accomplished:

1) To compare the GBB method against the current genomic selection (GS) and conventional PS methods for inbred and hybrid corn grain yield breeding; and
2) To introduce and apply the GBB method to enhanced inbred and hybrid corn breeding in the Texas Corn Breeding program.

This proposed research, once accomplished, will revolutionarily enhance the ability and efficiency of corn breeders to design and develop superior corn inbreds and hybrids, thus increasing corn production and benefiting Texas corn producers.

II. Research Progress Report:

Objective 1 - *To compare the GBB method against the current GS and conventional PS methods for inbred and hybrid corn grain yield breeding* (Year 1)

January 1, 2017 – June 28, 2017:

The accurate prediction of the targeted trait is the core of application of the molecular technologies in enhanced and accelerated plant breeding. We have accomplished the comparison of the gene-based breeding (GBB) method that we developed for enhanced and accelerated corn gran yield breeding (Inbred GBB) with the current genomic selection (GS) and the conventional phenotypic selection (PS) for inbred grain yield breeding. We found that when the number of favorable alleles or genotypes of only 27 of the *ZmINGY* (*Zea mays* inbred grain yield) genes that were identified to contain nucleotide mutations was applied for grain yield prediction, a prediction accuracy of grain yield was obtained that is comparable to those thus far achieved by GS using up to hundreds of thousands DNA markers. When the expression profiles of the gene transcripts responsible for grain yield were used for the prediction, a prediction accuracy of $r = 0.85$ (Pearson's correlation coefficient between predicted and field trial observed grain yields) was achieved, approaching the most accurate prediction. This prediction accuracy is higher than those of GS thus far achieved using genome-wide DNA markers by 63%. When two or all of the three datasets of the *ZmINGY* genes were jointly used for the prediction for GBB, the prediction accuracy of grain yield was 100% ensured.

Therefore, we further performed selection of superior lines using the Inbred GBB method. The inbred lines having the highest grain yields were selected at a selection rate of 10% for both GBB and PS. The GBB selection was performed based on the grain yields of the inbred lines predicted using (i) the number of favorable alleles of the 27 SNP/InDel-containing *ZmINGY* genes, (ii) the genotypes of the 27 SNP/InDel-containing *ZmINGY* genes and (iii) the transcript expression of the 150 selected *ZmINGY* genes in each inbred line, separately and in combination. The PS was performed according to the grain yields of the inbred lines measured through replicated field trials. The result showed that when genic datasets i, ii and iii were separately used for the selection with the GBB method, 40.0%, 50.0% and 66.7% of the superior inbred lines selected using the GBB method were shared, respectively, with those selected by the current PS method. These numbers were dozens-fold higher than that (10/91*10/91*100 = 1.2%) selected by chance using both one GBB genic dataset and PS. Moreover, we checked the ranks of the remaining 33.3% of the superior lines selected with the GBB genic dataset iii in those of PS and found that they were all within the top 25% inbred lines that had the highest observed grain yields, thus suggesting the accuracy of selection based on the phenotypes predicted using the GBB method. Furthermore, we conducted selection for superior lines jointly using two or all of the three *ZmINGY* genic datasets with the GBB method. We found that except for the superior lines selected with the combination of i+iii, 66.7% of which were in the top 10% inbred lines selected by PS, those selected with all other combinations, i+ii, ii+iii and i+ii+iii, of the GBB method were all in the top 10% inbred lines selected by the current PS method.

In addition, we investigated the reproducibility of the GBB selection for superior lines using the transcript expression of the 150 selected *ZmINGY* genes (genic dataset iii) to find why we were not able to find all superior lines selected with one of the three GBB genic datasets in the top 10% inbred lines selected by the current PS method. The result showed that the selection for superior lines with the GBB genic dataset had a selection reproducibility of $r = 0.98 – 0.99$ ($P = 0.0000$), which was far higher than the reproducibility of $r = 0.52$ ($P = 0.0000$) for the current PS method. Therefore, we concluded that the prediction accuracy of $r < 1.00$ achieved above with *ZmINGY* genes and the incomplete consistency of the selection results between GBB and PS were largely due to the observed grain yield variation of PS because it was readily subjected to environmental variation. The GBB method is far more accurate and efficient than the current PS method for inbred corn breeding.

<u>**Publications:**</u> A manuscript from these results has been submitted for publication in a peer-reviewed journal, in which the TCPB is acknowledged: "**This research was supported in part by grants from the Texas Corn Producer Board (TCPB) (M1701827)**".

- Meiping Zhang, Yanru Cui, Yun-Hua Liu1, Wenwei Xu, Sing-Hoi Sze, Shizhong Xu, Hong-Bin Zhang. 2017. Phenotype prediction of complex traits using their contributing genes for enhanced plant improvement through gene-based breeding (submitted, in review).

<u>June 29, 2017 – December 14, 2017:</u>

Almost all, if not all, of the corn cultivars that are currently grown for corn production in the USA are F_1 hybrids. Therefore, we focused on development of GBB for F_1 hybrid corn cultivar breeding (F_1 hybrid GBB) from their inbred parents in this period of the project. We first identified 128 genes that significantly contributed to F_1 hybrid grain yield (hereafter *ZmF1GY*) through systems analysis of the genes expressed in their inbred parents. Then, we tested ability, utility and efficiency of these 128 *ZmF1GY* genes in prediction of grain yields of F_1 hybrids according to the following:

(1) The expression profiles of 111 of the 128 *ZmF1GY* genes in their inbred parents; and
(2) The number of favorable alleles (including heterotic genotypes) of 17 of the 128 *ZmF1GY* genes in the F_1 hybrids, inferred from the genotypes of these genes in their inbred parents.

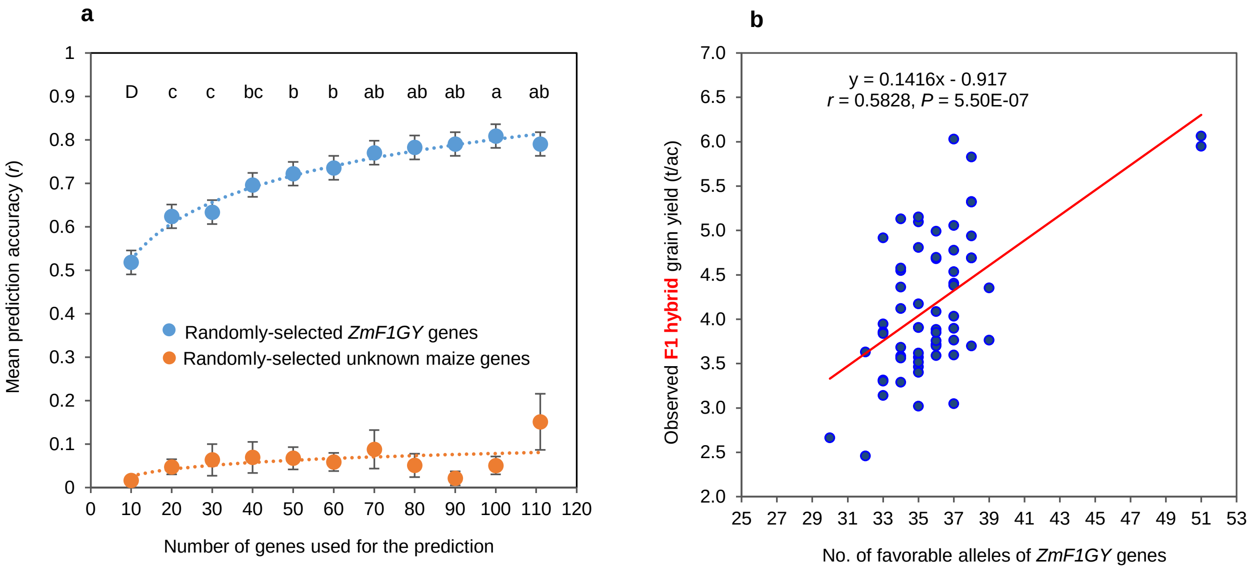

Figure 1 Predicting the grain yields of F_1 hybrids from their inbred parents. (a) Using the expression profiles of different numbers of *ZmF1GY* genes expressed in their inbred parents. The *ZmF1GY* genes and maize unknown genes were randomly selected by bootstrap sampling, with each number having 100 replications. The prediction accuracy (*r*) was the Pearson's correlation coefficient between the predicted and observed field trial grain yields. The mean prediction accuracy (*r*) shown in *y*-axis was the mean of the eight prediction models, with each model/gene number having 100 bootstrap replications. Error bar, standard deviation; small letter, significant at $P \leq 0.05$; capitol letter, significant at $P \leq 0.01$; same letter, not significant at $P \leq 0.05$; different letters, significant at $P \leq 0.05$ or 0.01. (b) Using the number of favorable alleles of 17 *ZmF1GY* genes containing SNPs/InDels that significantly influenced F_1 hybrid grain yields. The genotypes of each *ZmF1GY* gene in the F_1 hybrids were inferred from its genotypes in their inbred parents.

For Dataset (1), eight statistical prediction models were used for the grain yield prediction of F_1 hybrids from their parents. These eight statistical models were Partial Least Square (PLS) regression, Linear Support Vector Machine (SVMlinear), Bayesian Ridge Regression (BRR), Bayes A, Bayes B, Ridge Regression (RR), Random Forest (RF) and Least Absolute Shrinkage and Selector Operator (LASSO). For Dataset (2), the number of favorable alleles of the 17 *ZmFIGY* genes in F_1 hybrids inferred from the genotypes of these genes in their inbred parents was directly used for the grain yield prediction of F_1 hybrids, while their additive, dominant and over-dominant effects were taken into account. Consequently, we were able to predict the grain yields of F_1 hybrids at an accuracy of $r = 0.88$, when the expression profiles of 111 of the 128 *ZmF1GY* genes expressed in the inbred parents were used for F_1 hybrid yield prediction with the eight statistical prediction models (**Figure 1a**). This prediction accuracy has approached the highest accuracy of prediction due to the influence of environmental variation on observed field trial grain yield. Furthermore, we were also able to predict the grain yields of the F_1 hybrids at an accuracy of $r = 0.58$, when the number of favorable alleles of the 17 *ZmF1GY* genes in the F_1 hybrids inferred from the genotypes of the genes in their inbred parents (**Figure 1b**).

Therefore, we have successfully developed a GBB system for enhanced and accelerated F_1 hybrid grain yield breeding in maize (F1 hybrid GBB). Using the F_1 hybrid GBB system, breeders could select the inbred parent x inbred parent combination(s) that will yield the F_1 hybrids of the highest grain yields among all possible combinations of all inbred parents used in their F_1 hybrid breeding programs, before making any crosses and perfuming F_1 hybrid field trials. Therefore, the GBB systems not only save the money for making a large number of inbred parent x inbred parent crosses and performing associated field trials of all the crosses, and speed up the breeding process; importantly, they also allow the favorable alleles and heterotic genotypes of the genes controlling F_1 hybrid grain yield to be continuously and efficiently accumulated in inbred parents and maximizing the ability of F_1 hybrid breeding, thus promising to revolutionize the current corn F_1 hybrid breeding method.

Objective 2 - *To introduce and apply the GBB method to enhanced inbred and F_1 hybrid corn breeding in the Texas Corn Breeding Program* (Year 1 and Year 2)

January 1, 2017 – June 28, 2017:

For this objective, we planted two corn breeding populations developed by Dr. Wenwei Xu, Texas A&M AgriLife Research, Lubbock for grain yield field trials. They were ANTIGOT and MP313BTC testcross F_1 populations, with their 59 inbred parents. The ANTIGOT population consisted of 21 experimental hybrids and 4 commercial checks, and the MP313BTC population consisted of 38 experimental hybrids and two commercial checks). The field trials were conducted, with two-row plots and three replicates, at three locations, College Station, Lubbock and Halfway, TX.

June 29, 2017 – December 14, 2017:

We have accomplished the field trials of the ANTIGOT and MP313BTC testcross F_1 populations and their 59 inbred parents that are currently used for F_1 hybrid corn breeding by Dr. Wenwei Xu, Texas A&M AgriLife Research, Lubbock at two locations, College Station and Halfway, TX. We sampled the developing top ear shoots and ear leaves at the 13-leaf stage from every inbred parental line of the testcross F_1 hybrids. The samples were frozen in liquid nitrogen on site and are currently stored in a -80°C freezer for RNA and gene analysis in 2018. We have also collected the grain yield data as well as grain yield component traits from the F_1 hybrids from the field trials at all three locations, College Station, Lubbock and Halfway, TX. The field trial data analysis showed that 16 of the test F_1 hybrids of the MP313BTC population had grain yields 10% higher than those of the two checks in the Lubbock trials, of which 6 and 2 had grain yields 5% higher than those of the two checks in the College Station and Halfway trials, respectively. From the 21 test F_1 hybrids of the ANTIGOT population, we identified nine hybrids that had grain yields 10% higher than those four checks in the Lubbock trials, of which 1 and 8 had grain yields 5% higher than those of the four checks in the College Station and Halfway trials, respectively.

In 2018, we will, as originally proposed, analyze the developing top ear shoots of the inbred parents sampled and stored in a -80^{o}C freezer with the grain yield data of their F_1 hybrids. Particularly, we will conduct the followings:

- Sequence all genes, including those grain yield genes (*ZmINGY* and *ZmF1GY*) that we previously cloned, expressed in the 13-leaf top ear shoots of all 59 inbred parents used in this project from Texas A&M AgriLife Research Corn Breeding Program, Lubbock, TX;
- Predict the grain yields of all possible F_1 hybrids using the F_1 hybrid GBB system that we developed above in Objective 1. These F_1 hybrids include all 4,095 [91*(91-1)/2] possible F_1 hybrids of the 91 inbred lines that we previously developed and sequenced, and all 1,711 [59*(59-1)/2] possible F_1 hybrids of the 59 inbred parents being used in the Texas A&M AgriLife Research Corn Breeding Program, Lubbock;
- Verify the grain yields of the F_1 hybrids that will be predicted to give the highest grain yield, including making crosses to obtain these F_1 hybrids and performing the field trials on these selected F_1 hybrids;
- Compare the superior F_1 hybrids developed by GBB with the superior F1 hybrids developed using the current PS method based on the field trials of the F_1 hybrids; and
- Develop 1 – 2 F_1 hybrids that could be potentially released for commercial production in Texas and probably, nationwide.

<u>Publications:</u>

- Zhang MP, Liu YH, Paker J, Murray SC, Xu W, Zhang H-B. Gene-based breeding in maize: Predicting the grain yield of F_1 hybrids from their inbred parents (in preparation).

III. Efforts for Leveraging Resources:

January 1, 2017 – June 28, 2017:

Using the results of Objective 1, we prepared one research proposal for USDA National Institute of Food and Agriculture (NIFA) program and one research proposal for the Texas A&M AgriLife Research Crop Improvement:

- USDA NIFA: Hongbin Zhang, Meiping Zhang, Wenwei Xu and Seth C. Murray, Promoting Gene-based Breeding for Enhanced and Accelerated Grain Yield Improvement in Maize, $500,000, 09/01/2018 – 08/31/2020 (submitted, in review).
- Texas A&M AgriLife Research Crop Improvement: Hongbin Zhang and Meiping Zhang, Packaging the Gene-based Breeding Toolkit for Enhanced and Accelerated Genetic Improvement in Maize, $160,000, 09/01/2018 – 08/31/2019 (funded).

<u>June 29, 2017 – December 14, 2017:</u>

First, one of the two proposals that we submitted in the first half year of 2017 has been awarded with $140,000 for two years:

- Texas A&M AgriLife Research Crop Improvement: Hongbin Zhang and Meiping Zhang, Packaging the Gene-based Breeding Toolkit for Enhanced and Accelerated Genetic Improvement in Maize, $140,000, 09/01/2018 – 08/31/2019 (awarded).

  The objectives of this project are three fold:

  1) Identify the key genes to the 12 individual grain yield and quality component traits, and integrate them into the current GBB toolbox for grain yield breeding by integrative analysis of our genic and phenomic datasets already generated for these traits (FY 18 and FY 19);
  2) Sequence the genes constituting the new toolbox developed in this research in 100 diverse inbred parents of one (USIF) of our inbred/F1 hybrid populations and identify their SNPs and favorable alleles for grain yield and component traits contained in the lines (FY 18 & FY 19);
  3) Test and validate the new GBB toolbox through breeding practices for grain yield and grain yield component traits with GBB using the toolbox (FY 19).

  Therefore, this project will greatly promote the applications of GBB in maize breeding. Given that GBB is designed to breed for superior cultivars by directly and efficiently using and manipulating the genes controlling the targeted traits, especially (i) their number and effects (additive, dominant and over-dominant) of favorable alleles, (ii) their genic SNP/InDel genotypes and (iii) their expression profiles and networks, GBB allows effectively and continuously accumulating the favorable alleles and heterotic genotypes from different germplasm lines into the current superior cultivars, thus providing an extremely powerful and efficient tool for breeders to continue crop substantial improvement.

  The other proposal that we submitted in the first half year of 2017 is still in pending as of December 14, 2017.

- USDA NIFA: Hongbin Zhang, Meiping Zhang, Wenwei Xu and Seth C. Murray, Promoting Gene-based Breeding for Enhanced and Accelerated Grain Yield Improvement in Maize, $500,000, 09/01/2018 – 08/31/2020 (in pending).

  Furthermore, we developed another proposal for the Illumina Greater Good Initiative Grant (GGIG) Program 2018, based on the results resulting from this TCPB project in the second half year of 2017:

- Illumina GGIG: Hongbin Zhang, Helping Feed the World through Gene-based Breeding (GBB), a Novel and Extremely Powerful and Efficient Molecular Method for Enhanced and Accelerated Breeding in Crops. $200,000 (in kind), 04/01/2018 - 03/31/2019 (in pending).

  In this proposal, we propose to "sequence the 13-leaf developing ear transcriptomes of a maize population consisting of 227 inbred lines widely used in the worldwide maize breeding programs and their 300 F_1 hybrids using the Illumina NovaSeq 6000."