

Laboratory for High-throughput Gene Discovery and Gene-based Breeding

Department of Soil and Crop Sciences, Texas A&M AgriLife Research and Texas A&M University, College Station, Texas 77843-2474, USA

Contact:

Dr. Hongbin Zhang, Professor of Plant Genetics, Systems Genomics, and Gene-based Breeding, Department of Soil and Crop Sciences, Texas A&M AgriLife Research and Texas A&M University, College Station, TX 77845-2474, USA. Tel.: +1-979-862-2244; E-mail: hbz7049@tamu.edu

Background:

Genes are the primary determinants of performance of all biological traits and processes and the core molecular bases of crop and livestock breeding. Therefore, it is crucial to genome-wide cloning of the genes controlling agronomic traits or biological processes to understand the molecular bases underlying the traits and to develop innovative or advanced tools efficient to improve crops and livestock *continuously* and *substantially*.

Most agronomic traits, such as yield, quality, and resistance to environmental biotic and abiotic stresses, are complex traits each controlled by numerous genes, probably hundreds to more than one thousand genes. It is difficult for breeders, without knowledge of the genes and efficient tools, to incorporate the favorable alleles and heterotic genotypes of such many genes controlling the breeding objective traits continuously and effectively into new varieties to enhance or sustain crop or livestock production, while agricultural production is being subjected to global climate change.

We have conducted invention, research, and development of life science and technologies for nearly 20 years to develop new technologies that can genome-wide, high-throughput, and rapidly clone the genes controlling agronomic traits or biological processes. We invented the *gExpress* technology that enables high throughput cloning of genes controlling traits from any species, including plants, animals, humans, and microbes, regardless of genome size, genome complexity, ploidy level, and availability of molecular or genomic knowledge.

Using the *gExpress* technology, we have successfully cloned 9,051 genes controlling grain yield and grain yield and quality component traits in maize, and 10,954 genes controlling fiber (lint) yield and fiber yield and quality traits in cotton. We confirmed or validated the functions of the genes by genetic transformation, biological effect analysis of their mutations on targeted traits, and their ability and efficiency for phenotype prediction of the targeted traits [1-6] through over 10-year research. The results of these studies consistently showed that the genes cloned with the *gExpress* technology are reliable, as those cloned with the current gene cloning methods that are widely used, but extremely low in throughput, such as map-based cloning, and useful for a variety of advanced biological research and enhanced breeding, including, but not limited to, deciphering of the molecular mechanisms underpinning the traits, development of biomarkers, i.e., SNP markers contained in the genes and having significant biological impacts on targeted traits, and practice of gene-based breeding [1-6]. Importantly, the *gExpress* technology has, for the first time, made it feasible to clone and determine the biological functions of most, if not all, of the genes contained in a species, including plants, animals, humans, and microbes.

Gene-based breeding (GBB) was first proposed in 2012 and presented in 2014 [7-9], based on the following:

- ✓ Genes are the primary determinants of all traits' performance, despite modulated by epigenetic factors and environments; therefore, a variety having a desirable gene content will best perform when grown in an environment to which it is best adapted.
- ✓ Most agronomic traits, such as yield, quality, and biotic/abiotic stress tolerance, are each controlled by numerous gene, which makes it difficult for breeders to incorporate the favorable alleles and heterotic genotypes continuously and efficiently from breeding parents into new varieties for enhanced production.
- ✓ The molecular basis of breeding is to continuously incorporate the favorable alleles and heterotic genotypes of the genes controlling breeding objective traits into new varieties.
- ✓ The current breeding procedures, including for inbred-line variety breeding (e.g., cotton and wheat) and hybrid variety breeding (corn, sorghum, rice, and vegetable crops), all include parent selection, crossing design, and progeny selection, followed by multi-location variety testing to certify new varieties and to identify where a new variety is grown for commercial production.

Gene-based breeding (GBB) is a revolutionary breeding technology and science that allows breeders to accurately, efficiently, and rapidly develop superior varieties for commercial production by design according to their breeding objectives, based on the genes controlling breeding objective traits, particularly their number of favorable alleles or heterosis genotypes (NFAs), genotypes of their SNP biomarkers, and/or their expression activities and networks, through the entire breeding process, including parent selection, crossing design, and progeny selection [1-7]. Since the parent selection, cross design, parent crossing, and progeny selection of GBB are all conducted based on the content of genes controlling breeding objective traits, such as favorable alleles and heterosis genotypes, and where new varieties developed by GBB are grown can be determined by subsequent multi-location variety testing, the entire process of GBB can be performed in greenhouse or phytotron and new varieties can be produced in a factory manner. Since GBB allows accurate selection of the most desirable or complementary breeding parents in gene content to approach breeding objectives, wise designing for parent crossing to maximally combine the favorable alleles and heterosis genotypes of the genes controlling objective traits from parents into progeny, and earlier, accurate, and rapid selection for superior individuals from progeny and develop them into new varieties competitive for production, GBB increases the ability and efficiency of the current breeding methods, such as those assisted by genomic selection and/or marker-assisted selection by dozens folds, shortens and accelerates the current breeding cycle by 3 – 5 years, and reduces the current breeding cost by several folds [2-7]. For instance, a breeder can annually evaluate over 1,000,000 parent combinations with GBB with limited funding (< \$0.30 per combination) for hybrid variety development, likely resulting in over 1,000 new hybrid varieties, if a selection intensity of 0.1% is performed, for multi-location variety testing and commercial production, while it may take a breeder hundreds of years to evaluate the 1,000,000 parent combinations with the current hybrid variety breeding methods and tens of million U.S. dollars.

Using the cloned genes, we have first developed GBB in maize [2,4,5] and cotton [1,4,6], tested its ability, utility, and potential for both inbred-line and hybrid variety breeding, and are developing superior varieties with GBB for maize and cotton. For instance, we are developing

hybrid maize varieties by GBB with approximately 70,000 hybrids, from which studies showed that over 200 superior hybrid maize variety candidates will be identified for multi-location variety testing that have grain yields higher than the locally released hybrid varieties by 10% - 30% and are combined with heat tolerance. When one of the above three genic datasets, NFAs, SNP biomarkers, and gene expressions and networks, was used for progeny phenotype prediction and selection for inbred-line varieties, the prediction accuracy, i.e., Pearson's correlation coefficient (r) between GBB predicted and replicated field trial phenotypes, approached to 0.86 and 67% - 75% of the top 10% superior progeny individuals selected with the phenotypes predicted by GBB were consistent with those selected based on replicated field trials for maize inbred line grain yield [2] or cotton fiber length [3]. When the phenotypes predicted with two or all three of the above genic datasets were jointly used for progeny selection, the top 10% superior progeny individuals selected with the phenotypes predicted by GBB were consistent, at a rate of 100%, with those selected based on replicated field trials for maize inbred line grain yield [2] or cotton fiber length [3]. For hybrid variety breeding, we have developed GBB in maize, through which the performance of F₁ hybrids can be predicted from gene contents of the parents at a prediction accuracy of $r = 0.86$ ($P < 0.0001$), before hybrids are produced [5]. Moreover, we have developed a simple, rapid, and efficient method for both inbred-line and hybrid variety breeding, for which only four simple numbers, "0, 1, 2, and 3", are used for GBB [4] and model training and validation is unnecessary. Importantly, studies in cotton showed that the current best variety or breeding lines contains only 52% of the favorable alleles and heterosis genotypes of the genes controlling an agronomic trait that has been extensively subjected to modern breeding, revealing the potential of continuously improving the current best varieties [6]. It was demonstrated that GBB can potentially further improve current best commercial varieties by 73% - 118% [6], promising to large-scale and continuously improve current crop varieties or livestock strains and, thus, helping feed the world.

Services:

The laboratory provides the following services to *private* and *public organizations* to advance the current research and promote GBB in crops and livestock, and gene-based health, clinics, and medicine in humans.

- ✓ ***Genome-wide high-throughput cloning of genes controlling a biological trait or process.*** The service includes extraction of DNA and RNA from tissues provided; identification of candidate genes controlling targeted traits; and validation of their biological functions, for which 100 – 200 mg of tissues from each genotype or individual and phenotypic or clinic data of the experimental genotypes or individuals are required. We also provide consultation necessary for experimental design.
- ✓ ***Molecular characterization of cloned genes.*** The service includes identification of favorable alleles and heterosis genes; development of biomarkers; estimation of biological impacts of their mutations on targeted traits; and estimation of biological effects of their expression and/or network variation on targeted traits.
- ✓ ***Development of GBB for a species of interest.*** This service includes identification of biomarkers and transcripts that are suitable for GBB for diverse populations and/or genotype-

or environment-specific populations and development and validation of model(s) for GBB. We also develop kits, tools, and knowledge necessary for gene-based health, gene-based clinics, and gene-based medicine for humans.

- ✓ ***GBB in maize and cotton to produce superior varieties for commercial production.*** The service includes extraction of DNA and/or RNA from tissues provided and genotyping of potential breeding parents and progeny with the yield and/or yield and quality component trait genes that we have cloned thus far, varying from a dozen to hundreds of the genes, for the purposes of parent selection, crossing design, and progeny phenotype prediction and selection, including phenotype prediction of F₁ hybrid combinations from parents before hybrids are actually generated.

For the costs of these services and any other questions, please contact Prof. Dr. Hongbin Zhang at e-mail: hbz7049@tamu.edu; Phone: +1-979-862-2244.

For more information:

- [1] Zhang et al. 2020, *Scientific Reports* 10:10074. <https://doi.org/10.1038/s41598-020-66271-8>
- [2] Zhang et al. 2020, *Genomics* 112:225-236. <https://doi.org/10.1016/j.ygeno.2019.02.001>
- [3] Liu et al. 2020, *Frontiers in Plant Science* 11:583277. <https://doi.org/10.3389/fpls.2020.583277>
- [4] Liu et al. 2022, *Plant Science* 316 (2022) 111153. <https://doi.org/10.1016/j.plantsci.2021.111153>
- [5] Zhang et al. 2022, *Plant Science* 324:111424, <https://doi.org/10.1016/j.plantsci.2022.111424>
- [6] Liu et al. 2022, *Plant Science* 321: 111318. <https://doi.org/10.1016/j.plantsci.2022.111318>
- [7] Zhang et al. 2021, CABI, United Kingdom, pp. 380-404. <https://doi.org/10.1079/9781789245431.0022>
- [8] Liu et al. 2014, *International Plant & Animal Genome Conference XXII*, P474 (<http://www.intlpag.org/>)
- [9] Zhang et al. 2014, *International Plant & Animal Genome Conference XXII*, P875 (<http://www.intlpag.org/>).