Molecular Plant
Review Article

CellPress
PARTNER JOURNAL

# Genome-wide Association Studies in Maize: Praise and Stargaze

Yingjie Xiao[1,4], Haijun Liu[1,4], Liuji Wu[2,4], Marilyn Warburton[3] and Jianbing Yan[1,*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[2]Synergetic Innovation Center of Henan Grain Crops, Henan Agricultural University, Zhengzhou 450002, China

[3]United States of Department of Agriculture, Agricultural Research Service, Corn Host Plant Resistance Research Unit, Box 9555, MS 39762, Mississippi, USA

[4]These authors contributed equally to this article.

***Correspondence: Jianbing Yan (yjianbing@mail.hzau.edu.cn)**

http://dx.doi.org/10.1016/j.molp.2016.12.008

## ABSTRACT

**Genome-wide association study (GWAS) has become a widely accepted strategy for decoding genotype-phenotype associations in many species thanks to advances in next-generation sequencing (NGS) technologies. Maize is an ideal crop for GWAS and significant progress has been made in the last decade. This review summarizes current GWAS efforts in maize functional genomics research and discusses future prospects in the omics era. The general goal of GWAS is to link genotypic variations to corresponding differences in phenotype using the most appropriate statistical model in a given population. The current review also presents perspectives for optimizing GWAS design and analysis. GWAS analysis of data from RNA, protein, and metabolite-based omics studies is discussed, along with new models and new population designs that will identify causes of phenotypic variation that have been hidden to date. The joint and continuous efforts of the whole community will enhance our understanding of maize quantitative traits and boost crop molecular breeding designs.**

**Key words:** GWAS, omics, mixed model, population design, functional genomics, *Zea mays*

**Xiao Y., Liu H., Wu L., Warburton M., and Yan J.** (2017). Genome-wide Association Studies in Maize: Praise and Stargaze. Mol. Plant. **10**, 359–374.

## INTRODUCTION

Maize (*Zea mays* L.) originated in the Balsas River basin of southwestern Mexico approximately 9000 years ago (Matsuoka et al., 2002). It has since spread geographically and economically, becoming one of the most important crops globally for food, feed, and fuel. Maize grain production has increased more than eight-fold in the past century to a current annual global production of one billion tons (http://faostat.fao.org/, 2013). However, continuously diversifying demands for maize production has led to the continuous need for genetic improvement of various agriculturally and economically important traits.

The most economically important traits are usually inherited in a quantitative manner, and the genetic basis is attributed to polygenes and interaction effects between genes and/or genes and the environment. Linkage mapping based on a segregating population from a cross between two parents displaying maximally different phenotypes is a well-known approach to locate quantitative trait loci (QTL). They are statistically inferred, generally via linear regression and maximum likelihood estimate methods (Zeng, 1994), and based on a genetic linkage map (Lander and Botstein, 1989). Only a few QTLs are generally detected via linkage mapping in each experiment. Further fine mapping of QTL to a more narrowly precise genetic position and cloning of the underlying gene, as large secondary populations are generally required to achieve sufficient map resolution (Dinka et al., 2007), are particularly resource- and time-consuming processes. The large and complex maize genome, more than 85% of which consists of repetitive sequences (Schnable et al., 2009) further slows progress in QTL fine mapping and cloning.

Genome-wide association study (GWAS) using diverse populations provides another strategy to effectively fine map QTL due to a large number of historical recombination events that lead to the rapid decay of linkage disequilibrium (LD) (Flint-Garcia et al., 2003). This association mapping strategy was firstly applied in plants in the beginning of the 21st century as a candidate gene association study in maize (Thornsberry et al., 2001); however, the first association study at genome-wide scale was reported in maize, in 2008, in which 8590 loci in 553 elite

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

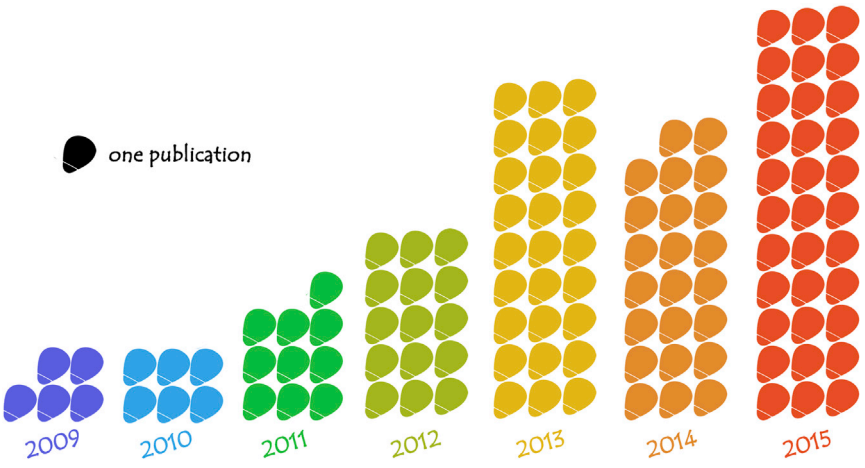


**Figure 1. Increasing Number of Publications on Maize GWAS Since the Release of the B73 Reference Genome.**
The number of publications is represented by the number of maize kernels and the statistics are from the gopubmed tool (http://www.gopubmed.com) by searching against the Medical Subject Headings (MeSH) of "*Zea mays*" and "Genome-wide association study".

maize inbreds was used to explore the genes affecting fatty acid content in kernels (Beló et al., 2008). Currently, GWAS is a routine tool in the study of human disease and other complex traits in many large cohort analyses (Visscher et al., 2012). For maize, since the release of the B73 reference genome (Schnable et al., 2009), GWAS has proliferated substantially (Figure 1), and dozens of agriculturally important traits have been dissected (Table 1). These advances suggest GWAS is a powerful tool to effectively and efficiently identify genome-phenotype associations. In this review, we (1) review maize functional genomics facilitated by GWAS on representative traits and a massive scale; (2) outline progress of new genetic and higher level (over-genetic) variation, method innovations, and population designs that maximize statistical power; and (3) discuss the challenges and opportunities for maize GWAS in the future.

## GWAS: A PROMISING TOOL IN MAIZE FUNCTIONAL GENOMICS

Functional genomics in plants aims to identify the functions of all genes. In the last decade, the explosive development of next-generation sequencing (NGS) technologies, and the release of the maize B73 reference genome, have largely promoted maize genetic research into the genomics era. To date, diverse traits (listed in Table 1), ranging from molecular (including the transcriptome) to cellular (i.e., metabolites), and from the individual morphological scale (agronomic, yield, or reproductive characteristics) to the interaction with different environmental factors (biotic or abiotic stress tolerance), have been comprehensively studied, along with a number of cloned genes and many more proposed gene candidates for corresponding traits, all using a GWAS approach.

### GWAS Helps Understand Genetic Architecture of Complex Quantitative Traits

Very-large-scale GWAS analyses provide new opportunities to understand the genetic architecture of complex quantitative traits. The data generated to date generally demonstrate that a large number of QTL are being identified but each only explains a small part of phenotypic variation for most of the agronomic traits investigated in maize. For example, more than 40 QTL were mapped for flowering time (Buckler et al., 2009) and leaf traits (Tian et al., 2011), but over 90% explained less than 2.5% of trait variance, and the remaining 10% of the QTL explaining under 5%. On the other hand, cellular quality and kernel composition traits appear to have a less complicated genetic architecture, with a smaller number of larger-effect QTL to be identified. In one study, several genes were found to explain more than 10% of the variance for kernel oil concentration (Li et al., 2013). In another study, the associated loci were found that explain over 20% of observed variance in the secondary metabolomic traits of maize kernels, with a median of 7.8% (Wen et al., 2014).

GWAS on transcriptomic variation, also called expression QTL (eQTL) mapping, links genetic markers to expression variation from thousands of genes, and has demonstrated the simplex genetic basis for gene expression traits, as each eQTL often explains a large proportion of phenotypic (expression level) variation. In a recent study, the expression of 14 375 genes was measured, and an average of over 15% of the variance was explained for each trait per eQTL. Particularly, a single eQTL was identified for explaining 87.7% of gene expressional variance (Fu et al., 2013). As the trait under study moves through different omic layers from the phenotype measured in variable environments, to the level of metabolites in single cell types, or to the expression of single genes, it is expected that the number of QTL identified will move from highly quantitative to qualitative or single locus.

### GWAS on Molecular Phenotypes Yields Genome Annotation and Insights into Mechanisms

An increasing number of publicly available GWAS results provide the opportunity to narrow associations to single, well-annotated candidate genes and to understand genome structure and constitution associated with each trait of interest. Early attempts to calculate the distribution pattern of associated loci at the genome-wide level found that genic and nearly genic regions (as opposed to intergenic regions) contribute most to maize trait variation, especially in the 5′UTR (untranslated region) regions (Li et al., 2012b). Furthermore, non-synonymous mutated single-nucleotide polymorphisms (SNPs) are the most functionally enriched, together with large copy number variants (CNVs), while intergenic regions show significant depletion for functional SNPs (Wallace et al., 2014). These systematic studies suggest that gene regulation in expression level should play a key role in phenotypic diversity. Under this hypothesis, the expression landscape of immature maize kernels has been extensively explored (Fu et al., 2013; Liu et al., 2016a), and the same

| Trait category | Phenotype | Population[a] | Sample size | No. marker[b] | Reference |
|---|---|---|---|---|---|
| Molecular and cellular | Gene expression | IAP | 368 | 557K | Fu et al., 2013 |
| | | IAP | 368 | 1.25M | Liu et al., 2016a |
| | Secondary metabolome | IAP | 368 | 557K | Wen et al., 2014 |
| | Oil concentration | IAP | 508 | 557K | Li et al., 2013 |
| | | USNAM + IAP | 4699 + 282 | 1.6M/52K | Cook et al., 2012 |
| | Carotenoid | IAP | 380 | 476K | Suwarno et al., 2015 |
| | | IAP | 201 | 284K | Owens et al., 2014 |
| | Tocopherol | IAP | 513 | 56K | Li et al., 2012a |
| | | IAP | 252 | 294K | Lipka et al., 2013 |
| | Carbon and nitrogen metabolism | IAP | 263 | 56K | Liu et al., 2016b |
| | | USNAM | 4699 | 1.6M | Zhang et al., 2015 |
| | | USNAM + IAP | 4699 + 282 | 1.6M/52K | Cook et al., 2012 |
| | Amino acids | IAP | 289 | 56K | Riedelsheimer et al., 2012 |
| | | USNAM + IAP | 4699 + 282 | 1.6M/52K | Cook et al., 2012 |
| | Leaf lipidome | IAP | 289 | 56K | Riedelsheimer et al., 2013 |
| | Leaf metabolome | IAP | 289 | 56K | Riedelsheimer et al., 2012 |
| | Drought-related metabolites | IAP | 318 | 157K | Zhang et al., 2016 |
| | Iron homeostasis | IAP | 267 | 438K | Benke et al., 2015 |
| Developmental and agronomic | Forage quality | IAP | 368 | 557K | Wang et al., 2016a |
| | Shoot apical meristem | IAP | 384 | 1.2M | Leiboff et al., 2015 |
| | Flowering time | IAP | 1487 | 8.2K | Van Inghelandt et al., 2012 |
| | | IAP | 368 | 557K | Yang et al., 2013 |
| | | IAP | 513 | 557K | Yang et al., 2014b |
| | | IAP | 346 | 60K | Farfan et al., 2015 |
| | | USNAM | 5000 + 281 | 1.1K | Buckler et al., 2009 |
| | | USNAM + IAP | 5000 + 281 | 1.1K | Hung et al., 2012 |
| | | USNAM + CNNAM + Ames | 4763 + 1971 + 1745 | 950K | Li et al., 2016c |
| | | MAGIC | 529 | 54K | Dell'Acqua et al., 2015 |
| | Plant height related | IAP | 284 | 41K | Weng et al., 2011 |
| | | IAP | 289 | 56K | Riedelsheimer et al., 2012 |
| | | IAP | 513 | 557K | Yang et al., 2014b |
| | | IAP | 258 | 224K | Li et al., 2016b |
| | | IAP | 346 | 60K | Farfan et al., 2015 |
| | | USNAM | 4892 | 1.6M | Peiffer et al., 2014 |
| | | MAGIC | 529 | 54K | Dell'Acqua et al., 2015 |
| | Leaf architecture | USNAM | 4892 | 1.6M | Tian et al., 2011 |
| | | IAP500 | 513 | 557K | Yang et al., 2014b |
| | | USNAM + NCRPIS | 4892 + 2572 | 1.6M/405K | Xue et al., 2016 |
| | Husk traits | IAP | 508 | 557K | Cui et al., 2016 |
| | Tassel architecture | IAP | 513 | 557K | Yang et al., 2014b |
| | | USNAM | 4892 | 1.6M | Brown et al., 2011 |
| | | USNAM + CNNAM + IAP | 4623 + 1972 + 945 | 500K/500K/44K | Wu et al., 2016 |

**Table 1. List of Diverse Traits Dissected via GWAS in Maize.**

*(Continued on next page)*

| Trait category | Phenotype | Population[a] | Sample size | No. marker[b] | Reference |
|---|---|---|---|---|---|
| | Ear height | IAP | 513 | 557K | Yang et al., 2014b |
| | | IAP | 346 | 60K | Farfan et al., 2015 |
| | | USNAM | 4892 | 1.6M | Peiffer et al., 2014 |
| | | USNAM + CNNAM + Ames | 4763 + 1971 + 1745 | 950K | Li et al., 2016b |
| | | MAGIC | 529 | 54K | Dell'Acqua et al., 2015 |
| | Stalk strength | IAP | 368 | 557K | Li et al., 2016a |
| | | USNAM + NCRPIS | 4536 + 2293 | 1.6M/681K | Peiffer et al., 2013 |
| | Root related | IAP | 267 | 438K | Benke et al., 2015 |
| | | Ames | 384 | 681K | Pace et al., 2015 |
| Yield | Ear architecture | IAP | 513 | 557K | Yang et al., 2014b |
| | | IAP | 513 | 49K | Liu et al., 2015b |
| | | IAP | 368 | 557K | Liu et al., 2015c |
| | | USNAM | 4892 | 1.6M | Brown et al., 2011 |
| | | USNAM + NCRPIS | 4892 + 2572 | 1.6M/405K | Xue et al., 2016 |
| | | ROAM | 1887 | 185K | Xiao et al., 2016 |
| | | XP panel | 400 | 940K | Yang et al., 2015 |
| | Grain size | IAP | 513 | 557K | Yang et al., 2014b |
| | | MAGIC | 529 | 54K | Dell'Acqua et al., 2015 |
| | Biomass | IAP | 289 | 56K | Riedelsheimer et al., 2012 |
| Stress resistance | Disease resistance | IAP | 1487 | 8.2K | Van Inghelandt et al., 2012 |
| | | IAP | 527 | 557K | Chen et al., 2015 |
| | | IAP | 1687 | 201K | Zila et al., 2014 |
| | | IAP | 999 | 56K | Ding et al., 2015 |
| | | IAP | 890 | 56K | Mahuku et al., 2016 |
| | | IAP | 818 | 43.4K | Chen et al., 2016 |
| | | IAP | 274 | 246K | Mammadov et al., 2015 |
| | | IAP | 287 | 261K | Tang et al., 2015; Warburton et al., 2015 |
| | | IAP | 380 + 235 | 259K/264K | Gowda et al., 2015 |
| | | IAP | 267 | 47K | Zila et al., 2013 |
| | | IAP | 346 | 60K | Farfan et al., 2015 |
| | | IAP | 267 | 287K | Horn et al., 2014 |
| | | USNAM | 4892 | 1.6M | Poland et al., 2011; Kump et al., 2011 |
| | Insect resistance | IAP | 302 | 246K | Samayoa et al., 2015 |
| | Hypersensitive response | IAP | 231 | 47K | Olukolu et al., 2013 |
| | | USNAM | 3381 | 26.5M | Olukolu et al., 2014 |
| | Drought tolerance | IAP | 80 | 1K | Hao et al., 2011 |
| | | IAP | 368 | 525K | Liu et al., 2013; Mao et al., 2015; Wang et al., 2016a, 2016b |
| | | IAP | 318 | 157K | Zhang et al., 2016 |
| | | IAP | 346 | 60K | Farfan et al., 2015 |
| | | IAP | 240 | 30K | Thirunavukkarasu et al., 2014 |

**Table 1.** *Continued*

*(Continued on next page)*

| Trait category | Phenotype | Population[a] | Sample size | No. marker[b] | Reference |
|---|---|---|---|---|---|
| | Water tolerance | IAP | 350 | 56K | Xue et al., 2013 |
| | Cold tolerance | IAP | 125 | 56K | Huang et al., 2013 |
| | | IAP | 375 | 56K | Strigens et al., 2013 |
| | | Dent + Flint | 306 + 292 | 50K | Revilla et al., 2016 |

**Table 1. *Continued***

[a]IAP, inbred association panel (consists of a set of inbred lines); NAM, nested association mapping (US and CN means USA and China, respectively); ROAM, random-open-parent association mapping; MAGIC, multi-parent advanced generation intercross population; NCRPIS, public collection of 2815 maize inbred accessions from the U.S. Department of Agriculture North Central Region Plant Introduction Station (NCRPIS; Romay et al., 2013); Ames: a subset of 1745 diverse inbred lines from the USDA-ARS NCRPIS; Dent and Flint, two temperate maize subgroups representing breeding germplasm adapted to European agro-climatic conditions; XP panel, extreme-phenotype populations. Plus symbol (+) represents multiple population are integrated in association mapping.

[b]The values separated by slash (/) indicate marker numbers corresponding to multiple populations.

conclusions were reached as previous findings on quantitative traits; i.e., that non-synonymous SNPs are the most significant drivers of expression regulation, with a higher number of SNP-eQTL associations (Liu et al., 2016a).

To explore another important layer shaping genetic diversity in maize, Wen et al. (2014) performed a metabolome-based GWAS in maize kernels to illustrate the whole biochemical landscape, seeking general and specific trends. These molecular-level association studies have taken full advantages of available GWAS data, which help to understand the intrinsic functional genome underlying trait variation. The resulting inferences are capable of guiding new or more in-depth gene identification studies. For example, QTL results from the metabolite study and agronomic traits measured on the same population were jointly analyzed (Wen et al., 2015). This allowed the identification of a major QTL affecting both the metabolic trait (on which it had a big effect) and the agronomic trait (on which it had a minor effect). Using the clue provided by the metabolic trait as a bridge led to the identification of a gene underlying the QTL affecting the agronomic trait, and to a better understanding of the underlying mechanism. When known metabolites are associated with unknown genes or vice versa, good clues are provided for novel annotations of both metabolites and genes (Wen et al., 2014). Complex metabolic networks can be further reconstructed or identified as important for a given trait by combining linkage or association mapping and networks including expression regulatory networks (Wen et al., 2016) and known metabolic pathways (Tang et al., 2015). Combining large datasets gathered using many different protocols, especially emerging omics tools, with high-throughput trait association analyses will thus speed maize functional genomic study.

## GWAS Success in Enhancing Maize Breeding by Identifying Beneficial Alleles

An early successful and practical association example is the pro-vitamin A biofortification of maize. Currently, over 250 000 children suffer from blindness each year due to vitamin A deficiency (VAD), and nearly two billion people, mostly in developing countries, remain at risk for deficiencies for this and other micronutrients (Sherwin et al., 2012; Liu and Yan, 2015). Rare favorable alleles of *LcyE* and *crtRB1* were identified in candidate gene association analyses (Harjes et al., 2008; Yan et al., 2010). By introgressing these rare alleles into elite maize germplasm via molecular marker assisted breeding, maize with improved levels of pro-vitamin A is now consumed by tens of thousands of African children who would benefit immediately, and thus the prevalence of VAD is declining (Fiedler et al., 2014). It is one of the most successful molecular breeding projects of the Global Maize Program of the International Maize and Wheat Improvement Center (CIMMYT) to date (Sherwin et al., 2012; Liu and Yan, 2015).

The optimization of photoperiod sensitivity is one key factor that allows plants to adapt to new environments in different latitudes. The gene *ZmCCT* has a major effect on photoperiod sensitivity and has been clearly dissected in two independent GWAS publications (Hung et al., 2012; Yang et al., 2013). The CACTA-like transposable element in the 5′UTR region of *ZmCCT* is the causal variant for the methylation level of the promoter region. One allele reduces gene expression, which promotes early flowering, allowing maize to grow adaptively in higher latitudes.

Optimizing plant architecture is currently one of the key objectives in maize breeding. This includes optimizing height to reduce lodging and to allow an increase in plant density. *Brachytic2* (Multani et al., 2003) has been identified as a severe, rare, and natural mutant affecting plant height. A rare and natural mutation that moderately reduced plant height has proved to be the causal variant. This is a recent mutation, and is only detected in temperate maize germplasm, which will make it easier to realize its potential for yield improvement in future breeding programs (Xing et al., 2015). A separate study identified the *lg1* and *lg2* loci, which were found to be significantly associated with upper leaf angle, which correlated to an increase in the efficiency of solar radiation capture. This efficiency gave the *lg2* allele potentials to significantly increase grain yield (Tian et al., 2011). Finally, a 3-Kb intergenic sequence within the *KRN4* locus was found to be responsible for maize kernel row number (KRN) variation by regulating the expression of the SBP-box gene *Unbranched3*. The favorable allele of *KRN4* has been proved to be significantly enriched in elite temperate inbreds but not in tropical maize germplasm (Liu et al., 2015c).

Drought tolerance is a particularly complex quantitative trait controlled by many loci with small effects. It is highly influenced

by the environment, and is thus regarded as difficult to dissect using GWAS. A series of researchers simplified the drought phenotype by measuring a component trait, seedling survival rate under water stressed condition, and a series of association studies identified favorable natural variants of a number of genes that could be used for drought tolerance improvement in maize (Liu et al., 2013; Mao et al., 2015; Wang et al., 2016b). Reducing trait complexity by precisely measuring correlated metabolic traits instead of yield itself under drought has also led to successful GWAS dissection of drought-related traits in maize (Setter et al., 2011; Xue et al., 2013; Farfan et al., 2015; Zhang et al., 2016).

In summary, these and many more recent results summarized in Table 1 have allowed detailed study of specific traits of interest; these studies have proved to be fairly quick and direct because the identification of the best favorable alleles already present in natural populations. In addition, these GWAS studies paved the way for the most efficient exploitation of this natural variation, as seen in the example from the *crtRB1* study of pro-vitamin A. Linkage analysis had identified one QTL covering the *crtRB1* region in two independent segregating populations with similar phenotypic variation explained. Further gene-based association analysis identified six common haplotypes within *crtRB1*, each with different effects. The association information allowed the identification of potential parents containing the exact haplotypes that should be crossed for maximal expression of the trait (Yan et al., 2010). This was workable in spite of the fact that the best combination of parents came from different association panels and breeding pools, and might never have been crossed without the genomic information gained by GWAS, thus proving the ability to provide direct and reliable information for choosing appropriate parents and/or donors for breeding.

## GWAS MAY ENLIGHTEN THE DEBATE OVER "MISSING HERITABILITY"

In the past few decades, GWAS has successfully identified thousands of associated loci in humans, animals, and plants (Visscher et al., 2012; Lipka et al., 2015); this has provided many beneficial clues to improve disease therapies and animal and/or crop breeding. However, only a small portion of phenotypic variation for a trait can be explained in any given GWAS, especially in human studies, raising a long-term debate over the problem of "missing heritability" (Maher, 2008). For example, GWAS has identified dozens of human height associated loci in a large human cohort data set, but they only accounted for a small fraction (<5%) of total heritability (Visscher, 2008). On the other hand, by using all genome-wide SNPs in a study (not just statistically significant SNPs), Yang et al. (2010) collectively increased the estimate of height heritability to ~67% via a classic quantitative genetic approach.

These results provide a reminder that heritability may remain hidden in genomics studies until the proper tools can uncover the missing part. This part is believed to be partly accessible via the study of variants in genes of minor effect in more genetically homogeneous backgrounds, by increasing variants present in populations at low frequency until their effect can be properly measured, or by including new genetic variants untapped in previous studies (Manolio et al., 2009). A major difference between plant and human systems is that controlled crossing experiments are possible in the former but not in the latter; hence, plant systems provide a feasible opportunity to facilitate GWAS in elaborate population or statistical designs to improve mapping power. In the next sections, we summarize three potentially complimentary approaches that, separately or jointly, may contribute to uncovering missing heritability in plants, especially in maize, including novel types of genotypes and phenotypes, statistical method innovations, and new genetic designs.

## VARIATION ABOVE THE GENOMIC LEVEL CAN FUNCTION IN A COMPLEMENTARY MANNER

Structural variations, including CNV, presence/absence variation (PAV; an extreme form of CNV), inversions, and translocations, are prevalent in the maize genome (Springer et al., 2009; Swanson-Wagner et al., 2010) and many studies show they have significant contributions to plant phenotypic variation (Yang et al., 2013; Mao et al., 2015; Liu et al., 2016a). Lu et al. (2015) identified over one million PAVs by mapping 26 million tags from 14 129 inbred lines, and found that this type of variation exhibits enriched associations with a wide range of phenotypic traits. The PAVs of transcribed sequences were largely involved in expression regulation, metabolic fluctuation, and higher levels of phenotypic diversity and heterosis (Jin et al., 2016). These major structural rearrangements, including even larger scale repetitions caused by transposable elements, are known to account for large percentages of the content within the maize genome and may play larger roles in creating phenotypic variation than single-nucleotide variants (SNVs); however, because many of the larger scale changes are in high LD with flanking SNV, it is difficult to assign the phenotypic changes to one or the other exclusively (Lu et al., 2015).

Higher-order or over-genomic variations, such as changes or differences in expression levels, have proved to be great resources as "molecular phenotypes" (Figure 2) (Eichten et al., 2013; He et al., 2013). Therefore, we propose that they could also be regarded as independent "molecular genotypes" that are not simply in LD with genomic variation. Jin et al. (2016) used expression PAV (ePAV; those genes being expressed, or not) as the "genotype" and found this kind of marker playing significant roles in expression regulation, metabolome variation, and morphological trait diversity. By transforming expression level into binary variation, Liu et al. (2015a) used high vs. low expression (relative to the median value) as variation to disclose the contribution of differentially expressed genes to their corresponding cellular and agronomic trait variance. Results indicate that transcriptomic variation is prevalent and works at the regulatory level, and show several advantages that are complementary with SNP-trait association studies: (1) they reflect variation in both genetic and epigenetic regulatory elements; (2) they provide additional evidence to fine map QTL; (3) they help to understand molecular mechanisms and construct regulatory networks. Molecular omic phenotypes above the genome level can now be quickly measured with high-throughput and low-cost platforms, and can be used on either side of the

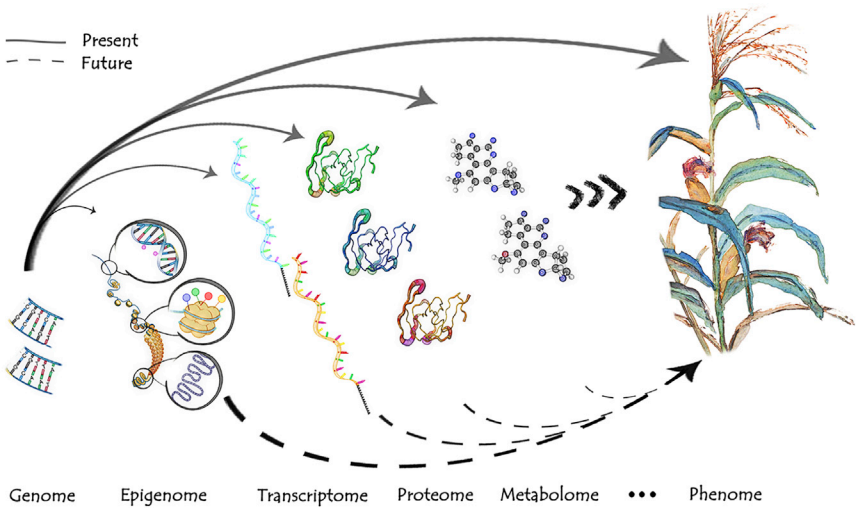


**Figure 2. Schematic of the Status of Maize GWAS Presently and in the Future.**
Different levels of variation are indicated with specific description expressed at the bottom; solid arrows demonstrate current GWAS efforts, which consider multi-layer data as "phenotypes". Integration of datasets collected on corresponding intermediate levels, such as RNA variation, metabolite or proteins, and component traits, would help a lot to disclose functional pathways underlying complex traits. The dashed arrows imply the expectation of modeling multi-level variation separately as "genotypes" or jointly as "covariates", to facilitate identification of biological causes for given traits, which will be the focus of research in coming years, as the OWAS (omic-wide association study) era opens.

mapping equation (i.e., as the "phenotype" or as the "genotype"; Figure 2); this could fill in missing information when one or more variables were considered complementarily.

The epigenomic code provides another critical layer in deciding the fate of a cell or an organism. Epigenetic changes can be independent from genomic variation and respond (in vivo and in vitro) to environment changes, accompanying phenotypic variability. Recent advances in population epigenomics in a handful of species have focused on decoding the genetic basis, still regarding the epigenetic modification as "traits" (Schmitz et al., 2013; Kawakatsu et al., 2016). Regarding the epigenome as the "genotypic" side of the equation would be valuable, similar to the transcriptome discussed above, and, together with other omics studies of the proteome and metabolome, this kind of over-genomic variation could be applied as molecular markers to dissect downstream phenotypes (Figure 2). Because the epigenome and transcriptome are closer to their phenotypic consequences than the genome does, they also hold the promise of retrieving missing heritability.

The identification of variations in these new "genotypes" at the population level for species with large genomes such as maize, currently is still nontrivial, as it does not consider tissue and developmental stage differences. Furthermore, as different as genomic variation with binary or limited numbers of alleles can be, variation for new omics traits are quantitative and range mutable, thus making conventional and straightforward mapping methods unsuitable for new omics data. Li et al. (2015) suggested a low-cost capture-based bisulfite approach for robustly and reliably analyzing DNA methylation for pre-defined genomic regions. McGeachie et al. (2014) proposed a conditional Gaussian Bayesian method to deduce possible links between discrete and continuous values under a network strategy, which may be promising for association mapping as well. A linear regression model accounting for population structure and relatedness matrix when analyzing higher level omics data, quantitative GWAS (qGWAS), has recently been proposed to solve the continuous genotype issue, and has been used to explore the regulatory network by treating expression level of genes as both "genotypes" and "phenotypes" (Wen et al., 2016). Future developments in sequencing technologies should be considered concomitantly with new analytical methods focused on higher level variation to more fully illuminate the hidden mysteries at the lowest omics levels (Figure 2).

## INNOVATIVE GWAS METHODS IMPROVE MAPPING POWER

To quantify how genes or QTL influence phenotypic variation, early GWAS simply regressed the marker variable against the phenotype, and additionally used control background noise with covariates to rule out false associations due solely to population structure in natural populations of plants (Pritchard et al., 2000; Price et al., 2006). However, in a natural population, the individuals always share complex ancestral relatedness, usually due to extensive intercrossing efforts in crop breeding (Myles et al., 2009). The one-dimension scale of inferred subpopulations for GWAS is generally insufficient to control the covariance relationships between individual pairs in multi-dimensional scales, and the extent of bias depends on the trait per se. In animal-breeding programs, the mixed-model approach is routinely used to select favorable individuals in a breeding population based on breeding values, estimated using a pedigree-based genetic relatedness matrix (GRM) between individual pairs of animals (Henderson, 1975, 1984). Although seemingly straightforward (Zhang et al., 2005), it is infeasible to directly apply the animal mixed model to plant GWAS, because crop breeding pedigree records are sometimes inaccurate and often completely unknown. A unified mixed-model method that substitutes the unavailable pedigree-based GRM with a marker-based GRM was an efficient solution to simultaneously account for population structure and the variance-covariance matrix in GWAS (Yu and Buckler, 2006). In the following section, we call this the standard Mixed Linear Model (MLM) method (Yu et al., 2006) and we summarize key points and motivations of recent improvements in GWAS methods (see also Table 2).

### Variation of Mixed Models for Improving GWAS Power

Statistically speaking, it is straightforward to improve power via increasing sample size. However, the standard MLM method is

| Year | Method | Positive semidefinite matrix requirement[a] | Strategy for increasing computational speed | | | Computational speed | Statistical power[d] |
|---|---|---|---|---|---|---|---|
| | | | Approximate/ Two-step approach[b] | Matrix optimization[c] | Low-rank matrix | | |
| 2006 | Standard MLM | | | | | Low | High |
| 2007 | GRAMMAR | | + | | | Very fast | Intermediate |
| 2008 | EMMA | + | | + | | Intermediate | High |
| 2010 | EMMAX | + | + | + | | Fast | High/Intermediate |
| 2010 | P3D and CMLM | | + | | + | Fast | High/Intermediate |
| 2011 | FaST-LMM | + | | + | | Fast | High |
| 2012 | GEMMA | + | | + | | Fast | High |
| 2012 | FaST-LMM-Select | + | | + | + | Very fast | High |
| 2014 | ECMLM | | + | | + | Intermediate | High/Intermediate |
| 2014 | SUPER | + | | + | + | Fast | High |

**Table 2. Performance Comparison of Different Methods in Mixed Linear Model GWAS.**

[a]Positive semidefinite matrix allows to be spectrally decomposed, a key step in EMMA algorithm (Kang et al., 2008), which significantly speeds up estimation of variance components in mixed model.

[b]Approximate approach includes residual and P3D approaches, commonly based on polygenic model that assumes that the effect of any locus on the trait is very small.

[c]Matrix optimization indicates the spectral decomposition technique first proposed in the EMMA algorithm (Kang et al., 2008).

[d]The relative power comparisons are summarized empirically, but further systematic simulations are needed.

inefficient for large data sets (those with thousands of individuals), because of the computational burden necessary for numerical optimization (Smith, 1990; Johnson and Thompson, 1995). In a first attempt to improve the speed of GWAS calculations, the efficient mixed-model association (EMMA) method simplifies matrix operations via spectral decomposition (Kang et al., 2008). However, this "exact method" of solving mixed-model equations with whole-genome markers in an iterative fashion is of limited value when testing millions of markers to identify all recombination events. Other diverse new methods with different assumptions have been proposed for analysis with continuously increasing sample size and marker density, and details of these new methods are briefly introduced below and summarized in Table 2.

A good trade-off between computational speed and statistical power can be made by only estimating model parameters once and then constantly testing markers iteratively using an approximate method, typically including the P3D (population parameter previously determined) and the residual approach. The P3D approach, including P3D (Zhang et al., 2010) and the EMMA expedited (EMMAX) algorithm (Kang et al., 2010), provides similar benefits to the residual approach (such as genome-wide rapid association using mixed model and regression [GRAMMAR]; Aulchenko et al., 2007); however, the P3D and residual approaches are technically different. The residual approach fits the residuals from a reduced mixed model as the dependent variable to test marker significance using a linear regression, whereas P3D fits the original phenotype as the dependent variable and tests marker significance under a mixed model with fixed variance parameters derived from the reduced model.

Assuming polygenic inheritance, approximate methods have proved to be an efficient solution to speed up GWAS for large-scale data sets aiming to increase mapping power. In truth, however, it is not possible to know a priori how accurate the approximate methods will be (Yang et al., 2014a). Two independent algorithms were proposed to improve the speed of the exact method by optimization of the mixed-model equation, and these are termed the factored spectrally transformed linear mixed model (FaST-LMM) (Lippert et al., 2011) and the genome-wide efficient mixed-model association (GEMMA) algorithm (Zhou and Stephens, 2012). Briefly, the improved exact methods typically focus on rewriting the mixed-model equations that refactor the conventional likelihood function of the mixed model to a form analogous to the likelihood of a linear regression model. This optimization simplifies the multi-dimensional parameters estimation into a one-dimensional numerical optimization problem, dramatically reducing computational burden for each iteration. These methods work at even greater speeds than the approximate methods in large-scale GWAS.

To solve a mixed-model equation, computing cost is increased most due to the matrix operations that estimate random effects, which are proportional to the cubed sample size. Hence, it should be possible to further improve GWAS speed by using a low-rank matrix in the mixed model. In animal breeding, animals' breeding value can be predicted by their sire origins; thus, animal relationships can be estimated by sire covariance matrices (this is termed the pedigree-based sire model). Similarly, Zhang et al. (2010) firstly proposed a low-rank matrix-based mixed-model approach, which they called the compressed MLM (CMLM). This model uses the GRM between pairs of groups to replace the GRM between pairs of individuals as the random effect. The CMLM method was further optimized by including a new parameter to define algorithms for calculating GRM between groups, termed enriched CMLM (ECMLM) (Li et al., 2014). Alternatively, Lippert et al. (2011) illustrate a new low-rank matrix method within

FaST-LMM, which still deals with an individual-pairs GRM matrix, but to estimate it using a much smaller subset of markers rather than by using all of them (Lippert et al., 2011).

The best markers to estimate the low-rank GSM matrix has not yet been determined. Under the assumption of the mixed model analogous to a Bayesian linear regression model, the extended version of FaST-LMM algorithm, FaST-LMM-Select, attempts to choose the best markers depending on the associations with a typical trait. Briefly, linear regression is firstly conducted to order all markers ascending by *p* values; the optimal subset of influential markers is then determined by looking for minimum genomic control index value, a parameter to estimate the genomic inflations due to background noise, which is commonly used in human disease GWAS studies (Devlin and Roeder, 1999). For each tested marker, the GSM matrix is iteratively established by using identified influential markers before omitting the tested markers (and all markers within 2 cM) (Listgarten et al., 2012). Beyond FaST-LMM-Select, a similar but more sophisticated algorithm, called Settlement of MLM under Progressively Exclusive Relationship (SUPER), provides another solution by treating the number and content of influential markers as genetic parameters of mixed-model functions for specific traits. Optimization of the likelihood would then be expected to increase statistical power and reduce false positives (Wang et al., 2014).

## New GWAS Methods for Multi-variant Test of Rare Variants

Traditional GWAS methods, including those mentioned above, assume the common disease caused by common variants model (CDCV) (Reich and Lander, 2001). Because of this bias, and because rare variants have such low statistical power, it is usual to filter out rare variants before any GWAS. However, rare alleles may be the cause of the phenotypic variant of interest. Thus, this is one source of the missing heritability, because a lack of sufficient power will not allow identification of rare causal variants unless their effect on the phenotype is extremely large (Eichler et al., 2010). Additionally, the case of multiple rare functional variants in nearby positions can potentially trigger indirect associations (synthetic association); these may be in positions that are megabases away from functional variants (Dickson et al., 2010). Synthetic associations caused by long-range LD blocks cause low GWAS resolution and increase difficulties in pinpointing causal genes based on GWAS signals (Cirulli and Goldstein, 2010). In species grown in very large numbers, recently emerged rare variants were found to be very prevalent, and thus intuitively believed to play vital functions on trait variations (Myles et al., 2009; Gibson, 2012); this was shown empirically to be true (Xiao et al., 2016). More robust evidence will follow as genotyping technologies rapidly advance, which enables a more complete assessment of rare variants with exceptionally large sample sizes and their roles in complex traits (Ozsolak and Milos, 2011; Grada and Weinbrecht, 2013).

When two or more independent causal variants, each with negligible effect, existing together in an inherited region (e.g., recombinant bin or LD block), it is statistically unlikely that any single variant will be detected, but would be possible to be detected by jointly testing the variants as a group in a multi-variant test (Gusev et al., 2013). Multi-variant methods, which benefit from rare variant detection demands (and potentially avoid synthetic associations), are now available to address multiple unlinked causal variants, and have been reviewed by Lee et al. (2014). These methods assume either fixed or random effects. On the basis of fixed-effect assumption for tested variants, burden tests propose the information from multi-variants are collapsed into a new statistical score that tests the association between the score and a trait (Madsen and Browning, 2009; Price et al., 2010; Asimit et al., 2012). For the basis of random assumption, variance-component tests have been proposed, and mixed models are used to test the likelihood ratio for a group of multi-variants treated as random-effect variables with independent normal distributions (Riggio et al., 2013; Sun et al., 2013; Casale et al., 2015).

Another option to analyze linked independent causal variants is the haplotype-based association analysis. This partially addresses the problem of synthetic association via the identification of the significance of allelic series or genetic heterogeneity attributed to multiple functional polymorphisms in a genomic (or gene-based) region (Zhang et al., 2012). Technically different to the multi-variants method, the haplotype-based method focuses on the identification of haplotypes, which unavoidably suffer bias of statistical ascertainment (Neale and Sham, 2004) but could be more biologically meaningful to geneticists. In maize, examples dealing with synthetic associations with haplotype-based method were presented by Lin et al. (2012); other examples can be found in tomato (Lin et al., 2014) and rice (Yano et al., 2016). However, the practical efficacy of the haplotype-based method is challenged by a trade-off between the benefits of modeling abundant variation and the cost of the extra degrees of freedom. Several statistical strategies are available to reduce the degrees of freedom in the haplotype-based method (Tzeng and Zhang, 2007), and the innovative biological, rather than statistical, significance of the results may improve the mapping power by unifying functional haplotypes for simplifying model complexity.

All the methods described above are parametric, as they assume the phenotype or marker effect will follow a specific (typically normal) distribution. To properly analyze the extremely unbalanced allelic frequencies common in association tests, nonparametric methods, which evaluate differences between two distributions with no prior assumptions, may be more appropriate. Yang et al. (2014b) proposed a nonparametric GWAS method, the Anderson-Darling (A-D) test, to conceptually test median differences between allelic groups rather than the mean value, which is usually tested in parametric methods. Compared with traditional mixed-model approaches, the A-D test showed a good balance between false positives and statistical power, especially for those traits with abnormal distribution and rare variants (Yang et al., 2014b).

The popularity of the mixed-model approach has made it an almost routine GWAS tool in plant studies in the era of big data; however, the complementary use of other analytical methods offers opportunities to more deeply explore biological questions such as where to find the ‘‘missing heritability’’. Several of the methods presented in this section would be appropriate for many GWAS analyses.

| Character | Population[a] | | |
|---|---|---|---|
| | NAM | MAGIC | ROAM |
| Cross pattern | Interconnect | Interconnect | Disconnect |
| Genetic diversity | High | High | High |
| Founder contribution | Extremely imbalanced | Balanced | Approximate balanced |
| Population size | Large | Intermediate | Large/very large |
| Algorithm complexity | Low | High | Intermediate |
| Recombinant events | Intermediate | High | Intermediate |
| Developmental cost | High | Very high | Low |
| Collaborative research | Possible | No | Suitable |
| Typical publication | Tian et al., 2011 | Dell'Acqua et al., 2015 | Xiao et al., 2016 |

**Table 3. Summary of the Characteristics of Three Multi-parent Population Designs in Maize.**
NAM, nested association mapping; MAGIC, multi-parent advanced generations intercross; ROAM, random-open-parent association mapping.
[a]See Figure 3 for details.

## NEW DESIGNS OF PLANT GENETIC RESOURCES FOR EFFICIENT IDENTIFICATION OF VARIANTS ENCODING COMPLEX TRAITS

In the last decade, association mapping in plants is almost exclusively studied on kinds of natural populations. While it is true that GWAS benefits from abundant diversity allowing the location of QTL to be inferred with a high resolution, the inherent population structure and presence of rare variants in natural populations will reduce GWAS statistical power. It is very difficult to correctly detect variants underlying traits of interest if it is significantly correlated to population structure (Flint-Garcia et al., 2005); while it is also hard to uncover rare variants, as indicated above. Although statistical innovations are proposed, it is still impossible to properly address most confounded or rare causal variants; new populations specifically designed to remove confounding variables or increase the frequency of the rare allele are thus required. This is usually done using bi-parental populations in plants; however, these insufficiently capture all (or most of) the existing variants underlying quantitative traits due to the limited diversity in any two parents (Yan et al., 2011). In maize, a set of pilot studies showcased the robust potential to identify QTL of minor effect and low frequency via GWAS in multi-parent populations (Table 1). In this section, we summarize the novelty and shortcomings of the currently available multi-parent populations for large-scale quantitative studies (Table 3).

### NAM

This large-scale, multi-parent design for plant is first initiated in maize by the creation of the nested association mapping (NAM) population, consisting of a set of 25 recombinant inbred line (RIL) populations derived from crosses between the maize inbred line B73 and another 25 genetically diverse inbred lines (Figure 3A) (Yu et al., 2008; McMullen et al., 2009). Because of the joint use of historical and recent recombination events and purified population structure, the maize NAM exhibits a powerful potential to thoroughly dissect the genetic architecture of complex quantitative traits, and many recent studies attest to the fulfillment of the potential (Buckler et al., 2009; Brown et al., 2011; Kump et al., 2011; Poland et al., 2011; Tian et al., 2011; Hung et al., 2012; Peiffer et al., 2013, 2014; Zhang et al., 2015; Wu et al., 2016; Li et al., 2016c). The structural design shares a common parent (B73) across all populations, enabling the additive effects of all identified QTL to be scaled against the common backdrop of the B73 allele, facilitating an intuitional application of QTL results to breeding programs. However, several statistical limitations to the use of NAM remain: (1) the absence of intercrossing between non-B73 parents potentially masks the effects of causal QTL by confounding population structure, in cases in which QTL segregate among RIL populations, but not within them; (2) imbalanced parental composition probably dilutes the GWAS efficiency at any given locus that assumes multi-allelic effects.

### MAGIC

In contrast to NAM, another multi-parent design, the multi-parent advanced generation intercross (MAGIC) population (Figure 3B), provides the opportunity to thoroughly intercross and effectively balance the contributions of all founder parents (Cavanagh et al., 2008). Initially, the MAGIC design is proposed in the mouse as the "Collaborative Cross" population, created as a community resource for genetic analyses of complex traits in mice (Churchill et al., 2004; Yalcin et al., 2005; Valdar et al., 2006). The idea has more recently been adopted as a prevalent design to identify QTL of agriculturally important traits in plants (Kover et al., 2009; Huang et al., 2011, 2012; Bandillo et al., 2013). Recently, a maize MAGIC population was established with eight diverse founder lines, which provides the maize community with a useful new resource (Dell'Acqua et al., 2015). Conceptually, The MAGIC design calls for intercrossing all parents in multiple rounds, leading to a balanced parental composition and numerous recombination events in the progeny, which is vital to increase statistical power and mapping resolution. However, inference of the genome-wide identity-by-descent (IBD) origins of MAGIC progenies is mathematically complicated, especially with increasing numbers of initial founder parents (Mott et al., 2000). The development of NAM and MAGIC populations requires extensive field and laboratory effort, which limits the application of these

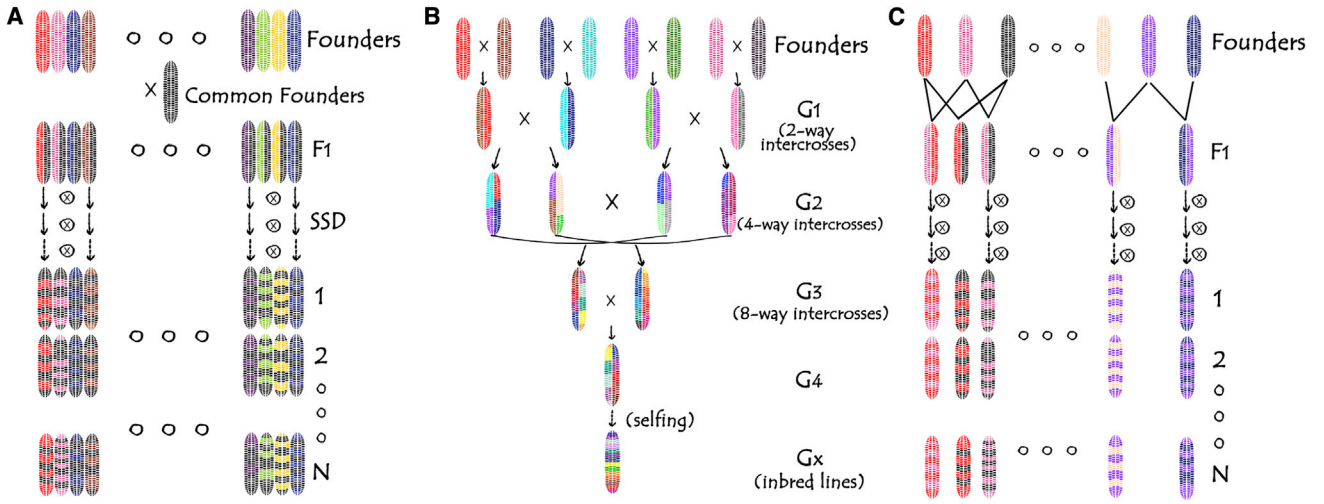


**Figure 3. Designs of Current Multi-Parent Populations in Maize.**

**(A)** The NAM population, a set of RIL families derived from a common parent (reference line) and diverse founder inbreds (modified from Figure 1, Yu et al., 2008). SSD, single seed descent.

**(B)** The MAGIC population, a set of RILs in which the genome of multiple founders (usually number of $2^n$) are reshuffled by several rounds of intercrossing, following several inbreeding generations (modified from Figure 1, Dell'Acqua et al., 2015).

**(C)** The ROAM population, which provides a manageable way to integrate existing population resources for large-scale genetic analysis in plants (modified from Supplemental Figure 1B, Xiao et al., 2016).

multi-parent designs. In addition, expansion of progeny numbers in existing NAM or MAGIC populations for traits displaying low variation is not possible.

### ROAM

A new design recently reported in maize combining different existing RIL populations is the random-open-parent association mapping (ROAM) population. This design improves genetic resolution and statistical power for identifying variants of minor effect and low frequency (Pan et al., 2016; Xiao et al., 2016). This design is created with 14 randomly intercrossed founder parents, from which 10 RIL populations are obtained (Figure 3C). The ROAM design allows construction of mapping populations de novo as it does not rely strictly on intercrossing between specific parents, hence permitting direct integration of new populations into currently existing population resources for large-scale genetic analysis whenever necessary. ROAM allows direct inference of IBD status within each bi-parental population, similar to NAM. The continuous inclusion of new populations into ROAM allows a better balance of founder parent contribution to be achieved as necessary to enhance QTL detection power. However, it is impossible to ensure that each new parent contributes one allele for a given QTL, and the robustness of the ROAM GWAS models may decrease somewhat due to the assumption that each parent will independently contribute one allele (Xiao et al., 2016). A haplotype-based GWAS would be used to better capture QTL by transforming allelic effect from each parent to inferred ancestral haplotypes in multi-parent populations (Giraud et al., 2014; Leroux et al., 2014).

For all these multi-parent designs, GWAS mapping resolution is commonly limited by the number of founders. Inclusion of more diverse founders is one way to further increase mapping resolution, with time- and labor-consuming effort. In traditional natural population, abundant historical recombinant events still drive the possibilities to increase mapping resolution, as long as statistical methods are powerful enough to identify missed heritability due to rare variants and population structure.

## PERSPECTIVES OF GWAS IN MAIZE: OPPORTUNITIES FROM CHALLENGES

The ultimate goal of breeding is to select better progeny, those with inherited favorable combination of alleles from their parents; thus, the particular importance of linking specific alleles to corresponding trait variation is evident. GWAS has come into its own in genetic mapping through large-scale and reliable practices, which shapes our understanding of maize functional genomics and genetics. Soon, for important crop traits, it will be possible to identify all underlying genes and their functions, which will directly accelerate molecular breeding designs. GWAS is still a powerful tool to fully exploit NGS technology advancements, but innovations in statistical method and genetic design can further promote this goal.

A deeper understanding of genetic basis of complex traits can be achieved through the application of diverse statistical methods, and different assumptions inherent to each method remind us to carefully choose the analysis methodology for each experiment, or to integrate complementary methods. In addition to new sequencing and statistical techniques, new population designs could aid in long-standing attempts to retrieve missed heritability in the identification and manipulation of genetic factors causing quantitative variation. These new designs will allow an increased detection power, which will be especially beneficial to overcome inherent statistical barriers, because these designs focus congruently on reforming allelic spectra and diluting confounding effects.

The genomics era has focused on studying the relationship between phenotype and genotype, and all GWAS methods

reviewed here attempt to find new QTL. In the new omics era, all layers of data that can be collected downstream of the level of the DNA sequence, the over-genomics data, have been shown empirically valuable for disclosing hidden heritability that seems to be incompletely penetrable when concerned only in the genomic layer. This hypothesis has been recently confirmed by a study, in which they found that the co-expression networks from transcriptome and proteome rarely overlap (Walley et al., 2016). Hence, integrating multi-layer data to GWAS will surely enhance the understanding of complex traits, not only to aid in narrowing down a candidate gene list but also to disclose functional pathways underlying interesting traits beyond the genomic hints, on which research focused in past decades. The joint use of omics data, specific genetic designs, and relevant analytical methods will open opportunities to move the past GWAS era to the OWAS (omic-wide association study) era, which will systematically integrate all layers of data to identify many more biological causes underlying phenotypic variation.

## FUNDING

This research was supported by the National Key Research and Development Program of China (2016YFD0101003), the National Natural Science Foundation of China (31525017, 31401389), and the National Youth Top-notch Talent Support Program.

## ACKNOWLEDGMENTS

No conflict of interest declared.

Received: September 10, 2016
Revised: December 2, 2016
Accepted: December 20, 2016
Published: December 27, 2016

## REFERENCES

**Asimit, J., Day-Williams, A., Morris, A., and Zeggini, E.** (2012). ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. Hum. Hered. **73**:84–94.

**Aulchenko, Y., de Koning, D., and Haley, C.** (2007). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics **177**:577–585.

**Bandillo, N., Raghavan, C., Muyco, P., Sevilla, M., Lobina, I., Dilla-Ermita, C., Tung, C., McCouch, S., Thomson, M., and Mauleon, R.** (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice **6**:11.

**Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D., Li, B., Tingey, S., and Rafalski, A.** (2008). Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Mol. Genet. Genomics **279**:1–10.

**Benke, A., Urbany, C., and Stich, B.** (2015). Genome-wide association mapping of iron homeostasis in the maize association population. BMC Genet. **16**:1.

**Brown, P., Upadyayula, N., Mahone, G., Tian, F., Bradbury, P., Myles, S., Holland, J., Flint-Garcia, S., McMullen, M., Buckler, E., et al.** (2011). Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genet. **7**:e1002383.

**Buckler, E., Holland, J., Bradbury, P., Acharya, C., Brown, P., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J., et al.** (2009). The genetic architecture of maize flowering time. Science **325**:714–718.

**Casale, F., Rakitsch, B., Lippert, C., and Stegle, O.** (2015). Efficient set tests for the genetic analysis of correlated traits. Nat. Methods **12**:755–793.

**Cavanagh, C., Morell, M., Mackay, I., and Powell, W.** (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr. Opin. Plant Biol. **11**:215–221.

**Chen, G., Wang, X., Hao, J., Yan, J., and Ding, J.** (2015). Genome-wide association implicates candidate genes conferring resistance to maize rough dwarf disease in maize. PLoS One **10**:e0142001.

**Chen, J., Shrestha, R., Ding, J., Zheng, H., Mu, C., Wu, J., and Mahuku, G.** (2016). Genome-wide association study and QTL mapping reveal genomic loci associated with fusarium ear rot resistance in tropical maize germplasm. G3 (Bethesda) **6**:3803–3815.

**Churchill, G., Airey, D., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W., et al.** (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. **36**:1133–1137.

**Cirulli, E., and Goldstein, D.** (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. **11**:415–425.

**Cook, J., McMullen, M., Holland, J., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E., and Flint-Garcia, S.** (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. **158**:824–834.

**Cui, Z., Luo, J., Qi, C., Ruan, Y., Li, J., Zhang, A., Yang, X., and He, Y.** (2016). Genome-wide association study (GWAS) reveals the genetic architecture of four husk traits in maize. BMC Genomics **17**:946.

**Dell'Acqua, M., Gatti, D., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A., Aung, H., Nelissen, H., Baute, J., et al.** (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol. **16**:167.

**Devlin, B., and Roeder, K.** (1999). Genomic control for association studies. Biometrics **55**:997–1004.

**Dickson, S., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.** (2010). Rare variants create synthetic genome-wide associations. PLoS Biol. **8**:e1000294.

**Ding, J., Ali, F., Chen, G., Li, H., Mahuku, G., Yang, N., Narro, L., Magorokosho, C., Makumbi, D., and Yan, J.** (2015). Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. BMC Plant Biol. **15**:206.

**Dinka, S., Campbell, M., Demers, T., and Raizada, M.** (2007). Predicting the size of the progeny mapping population required to positionally clone a gene. Genetics **176**:2035–2054.

**Eichler, E., Flint, J., Gibson, G., Kong, A., Leal, S., Moore, J., and Nadeau, J.** (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. **11**:446–450.

**Eichten, S., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P., Waters, A., Starr, E., West, P., Tiffin, P., et al.** (2013). Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell **25**:2783–2797.

**Farfan, I., De La Fuente, G., Murray, S., Isakeit, T., Huang, P., Warburton, M., Williams, P., Windham, G., and Kolomiets, M.** (2015). Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the sub-tropics. PLoS One **10**:e0117737.

**Fiedler, J., Afidra, R., Mugambi, G., Tehinse, J., Kabaghe, G., Zulu, R., Lividini, K., Smitz, M., Jallier, V., Guyondet, C., et al.** (2014). Maize flour fortification in Africa: markets, feasibility, coverage, and costs. Ann. N. Y. Acad. Sci. **1312**:26–39.

**Flint-Garcia, S., Thornsberry, J., and Buckler, E.** (2003). Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**:357–374.

**Flint-Garcia, S., Thuillet, A., Yu, J., Pressoir, G., Romero, S., Mitchell, S., Doebley, J., Kresovich, S., Goodman, M., and Buckler, E.** (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. **44**:1054–1064.

**Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., Zhang, J., He, C., Du, X., Peng, Z., et al.** (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. Nat. Commun. **4**:2832.

**Gibson, G.** (2012). Rare and common variants: twenty arguments. Nat. Rev. Genet. **13**:135–145.

**Giraud, H., Lehermeier, C., Bauer, E., Falque, M., Segura, V., Bauland, C., Camisan, C., Campo, L., Meyer, N., and Ranc, N.** (2014). Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the Flint and Dent heterotic groups of maize. Genetics **198**:1717–1734.

**Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M., Bright, J., Beyene, Y., and Prasanna, B.** (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor. Appl. Genet. **128**:1957–1968.

**Grada, A., and Weinbrecht, K.** (2013). Next-generation sequencing: methodology and application. J. Invest. Dermatol. **133**:e11.

**Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B., Diogo, D., Stahl, E., Gregersen, P., Worthington, J., Klareskog, L., Raychaudhuri, S., et al.** (2013). Quantifying missing heritability at known GWAS loci. PLoS Genet. **9**:e1003993.

**Hao, Z., Li, X., Xie, C., Weng, J., Li, M., Zhang, D., Liang, X., Liu, L., Liu, S., and Zhang, S.** (2011). Identification of functional genetic variations underlying drought tolerance in maize using SNP markers. J. Integr. Plant Biol. **53**:641–652.

**Harjes, C., Rocheford, T., Bai, L., Brutnell, T., Kandianis, C., Sowinski, S., Stapleton, A., Vallabhaneni, R., Williams, M., Wurtzel, E., et al.** (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science **319**:330–333.

**He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., Yang, M., Lu, L., Qi, Y., Wang, X., et al.** (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. Genome Biol. **14**:R57.

**Henderson, C.** (1975). Best linear unbiased estimation and prediction under a selection model. Biometrics **31**:423–447.

**Henderson, C.** (1984). Application of Linear Models in Animal Breeding (Ontario: University of Guelph).

**Horn, F., Habekuss, A., and Stich, B.** (2014). Genes involved in barley yellow dwarf virus resistance of maize. Theor. Appl. Genet. **127**: 2575–2584.

**Huang, X., Paulo, M., Boer, M., Effgen, S., Keizer, P., Koornneef, M., and van Eeuwijk, F.** (2011). Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. Proc. Natl. Acad. Sci. USA **108**:4488–4493.

**Huang, B., George, A., Forrest, K., Kilian, A., Hayden, M., Morell, M., and Cavanagh, C.** (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. **10**:826–839.

**Huang, J., Zhang, J., Li, W., Hu, W., Duan, L., Feng, Y., Qiu, F., and Yue, B.** (2013). Genome-wide association analysis of ten chilling tolerance indices at the germination and seedling stages in maize. J. Integr. Plant Biol. **55**:735–744.

**Hung, H., Shannon, L., Tian, F., Bradbury, P., Chen, C., Flint-Garcia, S., McMullen, M., Ware, D., Buckler, E., Doebley, J., et al.** (2012). *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA **109**:1913–1921.

**Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., Wang, Y., Xie, W., Wang, G., and Yan, J.** (2016). Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. Sci. Rep. **6**:18936.

**Johnson, D., and Thompson, R.** (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J. Dairy Sci. **78**:449–456.

**Kang, H., Zaitlen, N., Wade, C., Kirby, A., Heckerman, D., Daly, M., and Eskin, E.** (2008). Efficient control of population structure in model organism association mapping. Genetics **178**:1709–1723.

**Kang, H., Sul, J., Service, S., Zaitlen, N., Kong, S., Freimer, N., Sabatti, C., and Eskin, E.** (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354.

**Kawakatsu, T., Huang, S., Jupe, F., Sasaki, E., Schmitz, R.J., Urich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., et al.** (2016). Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. Cell **166**:492–505.

**Kover, P., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I., Purugganan, M., Durrant, C., and Mott, R.** (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet. **5**:e1000551.

**Kump, K., Bradbury, P., Wisser, R., Buckler, E., Belcher, A., Oropeza-Rosas, M., Zwonitzer, J., Kresovich, S., McMullen, M., Ware, D., et al.** (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. **43**:163–168.

**Lander, E., and Botstein, D.** (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**:185–199.

**Lee, S., Abecasis, G., Boehnke, M., and Lin, X.** (2014). Rare-variant association analysis: study designs and statistical tests. Am. J. Hum. Genet. **95**:5–23.

**Leiboff, S., Li, X., Hu, H., Todt, N., Yang, J., Yu, X., Muehlbauer, G., Timmermans, M., Yu, J., Schnable, P., et al.** (2015). Genetic control of morphometric diversity in the maize shoot apical meristem. Nat. Commun. **6**:8974.

**Leroux, D., Rahmani, A., Jasson, S., Ventelon, M., Louis, F., Moreau, L., and Mangin, B.** (2014). Clusthaplo: a plug-in for MCQTL to enhance QTL detection using ancestral alleles in multi-cross design. Theor. Appl. Genet. **127**:921–933.

**Li, Q., Yang, X., Xu, S., Cai, Y., Zhang, D., Han, Y., Li, L., Zhang, Z., Gao, S., Li, J., et al.** (2012a). Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. PLoS One **7**:e36807.

**Li, X., Zhu, C., Yeh, C., Wu, W., Takacs, E., Petsch, K., Tian, F., Bai, G., Buckler, E., Muehlbauer, G., et al.** (2012b). Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res. **22**:2436–2444.

**Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., et al.** (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45**:43–50.

**Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y., Todhunter, R., Buckler, E., and Zhang, Z.** (2014). Enrichment of statistical power for genome-wide association studies. BMC Biol. **12**:73.

**Li, Q., Suzuki, M., Wendt, J., Patterson, N., Eichten, S., Hermanson, P., Green, D., Jeddeloh, J., Richmond, T., Rosenbaum, H., et al.** (2015). Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res. **43**:e81.

**Li, K., Wang, H., Hu, X., Liu, Z., Wu, Y., and Huang, C.** (2016a). Genome-wide association study reveals the genetic basis of stalk cell wall components in maize. PLoS One **11**:e0158906.

**Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., Ma, J., Wang, S., Zhang, X., Xia, Z., et al.** (2016b). Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. Front Plant Sci. **7**:833.

**Li, Y., Li, C., Bradbury, P., Liu, X., Lu, F., Romay, C., Glaubitz, J., Wu, X., Peng, B., Shi, Y., et al.** (2016c). Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J. **86**:391–402.

**Lin, Z., Li, X., Shannon, L., Yeh, C., Wang, M., Bai, G., Peng, Z., Li, J., Trick, H., Clemente, T., et al.** (2012). Parallel domestication of the *Shattering1* genes in cereals. Nat. Genet. **44**:720–724.

**Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., et al.** (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**:1220–1226.

**Lipka, A., Gore, M., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., Chen, C., Buell, C., Buckler, E., Rocheford, T., et al.** (2013). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 (Bethesda) **3**:1287–1299.

**Lipka, A., Kandianis, C., Hudson, M., Yu, J., Drnevich, J., Bradbury, P., and Gore, M.** (2015). From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Curr. Opin. Plant Biol. **24**:110–118.

**Lippert, C., Listgarten, J., Liu, Y., Kadie, C., Davidson, R., and Heckerman, D.** (2011). FaST linear mixed models for genome-wide association studies. Nat. Methods **8**:833–835.

**Listgarten, J., Lippert, C., Kadie, C., Davidson, R., Eskin, E., and Heckerman, D.** (2012). Improved linear mixed models for genome-wide association studies. Nat. Methods **9**:525–526.

**Liu, N., and Yan, J.** (2015). The progress and perspective of vitamin A biofortification for maize. Chin. Bull. Life Sci. **27**:1028–1036.

**Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., Li, J., Tran, L., Shinozaki, K., Yamaguchi-Shinozaki, K., et al.** (2013). Genome-wide analysis of *ZmDREB* genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. PLoS Genet. **9**:e1003790.

**Liu, H., Wang, X., Warburton, M., Wen, W., Jin, M., Deng, M., Liu, J., Tong, H., Pan, Q., Yang, X., et al.** (2015a). Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. Mol. Plant **8**:871–884.

**Liu, L., Du, Y., Huo, D., Wang, M., Shen, X., Yue, B., Qiu, F., Zheng, Y., Yan, J., and Zhang, Z.** (2015b). Genetic architecture of maize kernel row number and whole genome prediction. Theor. Appl. Genet. **128**:2243–2254.

**Liu, L., Du, Y., Shen, X., Li, M., Sun, W., Huang, J., Liu, Z., Tao, Y., Zheng, Y., Yan, J., et al.** (2015c). *KRN4* controls quantitative variation in maize kernel row number. PLoS Genet. **11**:e1005670.

**Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X., Jin, M., Li, W., Zhang, Q., et al.** (2016a). Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. Mol. Plant http://dx.doi.org/10.1016/j.molp.2016.06.016.

**Liu, N., Xue, Y., Guo, Z., Li, W., and Tang, J.** (2016b). Genome-wide association study identifies candidate genes for starch content regulation in maize kernels. Front Plant Sci. **7**:1046.

**Lu, F., Romay, M., Glaubitz, J., Bradbury, P., Elshire, R., Wang, T., Li, Y., Li, Y., Semagn, K., Zhang, X., et al.** (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. Nat. Commun. **6**:6914.

**Madsen, B., and Browning, S.** (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. **5**:e1000384.

**Maher, B.** (2008). Personal genomes: the case of the missing heritability. Nature **456**:18–21.

**Mahuku, G., Chen, J., Shrestha, R., Narro, L., Guerrero, K., Arcos, A., and Xu, Y.** (2016). Combined linkage and association mapping identifies a major QTL (qRtsc8-1), conferring tar spot complex resistance in maize. Theor. Appl. Genet. **129**:1217–1229.

**Mammadov, J., Sun, X., Gao, Y., Ochsenfeld, C., Bakker, E., Ren, R., Flora, J., Wang, X., Kumpatla, S., Meyer, D., et al.** (2015). Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). BMC Genomics **16**:916.

**Manolio, T., Collins, F., Cox, N., Goldstein, D., Hindorff, L., Hunter, D., McCarthy, M., Ramos, E., Cardon, L., Chakravarti, A., et al.** (2009). Finding the missing heritability of complex diseases. Nature **461**:747–753.

**Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L., and Qin, F.** (2015). A transposable element in a *NAC* gene is associated with drought tolerance in maize seedlings. Nat. Commun. **6**:8326.

**Matsuoka, Y., Vigouroux, Y., Goodman, M., Sanchez, G., Buckler, E., and Doebley, J.** (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA **99**:6080–6084.

**McGeachie, M., Chang, H., and Weiss, S.** (2014). CGBayesNets: conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data. PLoS Comput. Biol. **10**:e1003676.

**McMullen, M., Kresovich, S., Villeda, H., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., et al.** (2009). Genetic properties of the maize nested association mapping population. Science **325**:737–740.

**Mott, R., Talbot, C., Turri, M., Collins, A., and Flint, J.** (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. USA **97**:12649–12654.

**Multani, D., Briggs, S., Chamberlin, M., Blakeslee, J., Murphy, A., and Johal, G.** (2003). Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science **302**:81–84.

**Myles, S., Peiffer, J., Brown, P., Ersoz, E., Zhang, Z., Costich, D., and Buckler, E.** (2009). Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell **21**:2194–2202.

**Neale, B., and Sham, P.** (2004). The future of association studies: gene-based analysis and replication. Am. J. Hum. Genet. **75**:353–362.

**Olukolu, B., Negeri, A., Dhawan, R., Venkata, B., Sharma, P., Garg, A., Gachomo, E., Marla, S., Chu, K., Hasan, A., et al.** (2013). A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics **193**:609–620.

**Olukolu, B., Wang, G., Vontimitta, V., Venkata, B., Marla, S., Ji, J., Gachomo, E., Chu, K., Negeri, A., Benson, J., et al.** (2014). A genome-wide association study of the maize hypersensitive defense response identifies genes that cluster in related pathways. PLoS Genet. **10**:e1004562.

**Owens, B., Lipka, A., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C., Kandianis, C., Kim, E., Cepela, J., Mateos-Hernandez, M., Buell, C., et al.** (2014). A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics **198**:1699–1716.

**Ozsolak, F., and Milos, P.** (2011). RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. **12**:87–98.

**Pace, J., Gardner, C., Romay, C., Ganapathysubramanian, B., and Lubberstedt, T.** (2015). Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). BMC Genomics **16**:47.

**Pan, Q., Li, L., Yang, X., Tong, H., Xu, S., Li, Z., Li, W., Muehlbauer, G., Li, J., and Yan, J.** (2016). Genome-wide recombination dynamics are associated with phenotypic variation in maize. New Phytol. **210**:1083–1094.

**Peiffer, J., Flint-Garcia, S., De Leon, N., McMullen, M., Kaeppler, S., and Buckler, E.** (2013). The genetic architecture of maize stalk strength. PLoS One **8**:e67066.

**Peiffer, J., Romay, M., Gore, M., Flint-Garcia, S., Zhang, Z., Millard, M., Gardner, C., McMullen, M., Holland, J., Bradbury, P., et al.** (2014). The genetic architecture of maize height. Genetics **196**:1337–1356.

**Poland, J., Bradbury, P., Buckler, E., and Nelson, R.** (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Natl. Acad. Sci. USA **108**:6893–6898.

**Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., and Reich, D.** (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**: 904–909.

**Price, A., Kryukov, G., de Bakker, P., Purcell, S., Staples, J., Wei, L., and Sunyaev, S.** (2010). Pooled association tests for rare variants in exon-resequencing studies. Am. J. Hum. Genet. **86**:832–838.

**Pritchard, J., Stephens, M., and Donnelly, P.** (2000). Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

**Reich, D., and Lander, E.** (2001). On the allelic spectrum of human disease. Trends Genet. **17**:502–510.

**Revilla, P., Rodriguez, V., Ordas, A., Rincent, R., Charcosset, A., Giauffret, C., Melchinger, A., Schon, C., Bauer, E., Altmann, T., et al.** (2016). Association mapping for cold tolerance in two large maize inbred panels. BMC Plant Biol. **16**:127.

**Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A.** (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc. Natl. Acad. Sci. USA **109**:8872–8877.

**Riedelsheimer, C., Brotman, Y., Méret, M., Melchinger, A., and Willmitzer, L.** (2013). The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. Sci. Rep. **3**:2479.

**Riggio, V., Matika, O., Pong-Wong, R., Stear, M., and Bishop, S.** (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity **110**:420–429.

**Romay, M., Millard, M., Glaubitz, J., Peiffer, J., Swarts, K., Casstevens, T., Elshire, R., Acharya, C., Mitchell, S., Flint-Garcia, S., et al.** (2013). Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. **14**:R55.

**Samayoa, L.F., Malvar, R., Olukolu, B., Holland, J., and Butron, A.** (2015). Genome-wide association study reveals a set of genes associated with resistance to the Mediterranean corn borer (*Sesamia nonagrioides* L.) in a maize diversity panel. BMC Plant Biol. **15**:35.

**Schmitz, R., Schultz, M., Urich, M., Nery, J., Pelizzola, M., Libiger, O., Alix, A., McCosh, R., Chen, H., Schork, N., et al.** (2013). Patterns of population epigenomic diversity. Nature **495**:193–198.

**Schnable, P., Ware, D., Fulton, R., Stein, J., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T., et al.** (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326**:1112–1115.

**Setter, T., Yan, J., Warburton, M., Ribaut, J., Xu, Y., Sawkins, M., Buckler, E., Zhang, Z., and Gore, M.** (2011). Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J. Exp. Bot. **62**:701–716.

**Sherwin, J., Reacher, M., Dean, W., and Ngondi, J.** (2012). Epidemiology of vitamin A deficiency and xerophthalmia in at-risk populations. Trans. R. Soc. Trop. Med. Hyg. **106**:205–214.

**Smith, S.** (1990). Estimation of genetic parameters in non linear models. In Advances in Statistical Methods for Genetic Improvement of Livestock, D. Gianola and K. Hammond, eds. (New York: Springer-Verlag), pp. 190–206.

**Springer, N., Ying, K., Fu, Y., Ji, T., Yeh, C., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., et al.** (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. **5**:e1000734.

**Strigens, A., Freitag, N., Gilbert, X., Grieder, C., Riedelsheimer, C., Schrag, T., Messmer, R., and Melchinger, A.** (2013). Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. Plant Cell Environ. **36**:1871–1887.

**Sun, J., Zheng, Y., and Hsu, L.** (2013). A unified mixed-effects model for rare-variant association in sequencing studies. Genet. Epidemiol. **37**:334–344.

**Suwarno, W., Pixley, K., Palacios-Rojas, N., Kaeppler, S., and Babu, R.** (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor. Appl. Genet. **128**:851–864.

**Swanson-Wagner, R., Eichten, S., Kumari, S., Tiffin, P., Stein, J., Ware, D., and Springer, N.** (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. **20**:1689–1699.

**Tang, J., Perkins, A., Williams, W., and Warburton, M.** (2015). Using genome-wide associations to identify metabolic pathways involved in maize aflatoxin accumulation resistance. BMC Genomics **16**:673.

**Thirunavukkarasu, N., Hossain, F., Arora, K., Sharma, R., Shiriga, K., Mittal, S., Mohan, S., Namratha, P., Dogga, S., Rani, T., et al.** (2014). Functional mechanisms of drought tolerance in subtropical maize (*Zea mays* L.) identified using genome-wide association mapping. BMC Genomics **15**:1182.

**Thornsberry, J., Goodman, M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E.** (2001). *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28**:286–289.

**Tian, F., Bradbury, P., Brown, P., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T., McMullen, M., Holland, J., and Buckler, E.** (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. **43**:159–162.

**Tzeng, J., and Zhang, D.** (2007). Haplotype-based association analysis via variance-components score test. Am. J. Hum. Genet. **81**:927–938.

**Valdar, W., Solberg, L., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W., Taylor, M., Rawlins, J., Mott, R., and Flint, J.** (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. **38**:879–887.

**Van Inghelandt, D., Melchinger, A., Martinant, J., and Stich, B.** (2012). Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. BMC Plant Biol. **12**:56.

**Visscher, P.** (2008). Sizing up human height variation. Nat. Genet. **40**:489–490.

**Visscher, P., Brown, M., McCarthy, M., and Yang, J.** (2012). Five years of GWAS discovery. Am. J. Hum. Genet. **90**:7–24.

**Wallace, J., Bradbury, P., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E.** (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. **10**:e1004845.

**Walley, J., Sartor, R., Shen, Z., Schmitz, R., Wu, K., Urich, M., Nery, J., Smith, L., Schnable, J., Ecker, J., et al.** (2016). Integration of omic networks in a developmental atlas of maize. Science **353**:814–818.

**Wang, Q., Tian, F., Pan, Y., Buckler, E., and Zhang, Z.** (2014). A SUPER powerful method for genome wide association study. PLoS One **9**:e107684.

**Wang, H., Li, K., Hu, X., Liu, Z., Wu, Y., and Huang, C.** (2016a). Genome-wide association analysis of forage quality in maize mature stalk. BMC Plant Biol. **16**:227.

**Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X., and Qin, F.** (2016b). Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. Nat. Genet. **48**:1233–1241.

**Warburton, M., Tang, J., Windham, G., Hawkins, L., Murray, S., Xu, W., Boykin, D., Perkins, A., and Williams, W.** (2015). Genome-wide association mapping of *Aspergillus flavus* and aflatoxin accumulation resistance in maize. Crop Sci. **55**:1857–1867.

**Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., et al.** (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. **5**:3438.

**Wen, W., Li, K., Alseekh, S., Omranian, N., Zhao, L., Zhou, Y., Xiao, Y., Jin, M., Yang, N., Liu, H., et al.** (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. Plant Cell **27**:1839–1856.

**Wen, W., Liu, H., Zhou, Y., Jin, M., Yang, N., Li, D., Luo, J., Xiao, Y., Pan, Q., Tohge, T., et al.** (2016). Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. Plant Physiol. **170**:136–146.

**Weng, J., Xie, C., Hao, Z., Wang, J., Liu, C., Li, M., Zhang, D., Bai, L., Zhang, S., and Li, X.** (2011). Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. PLoS One **6**:e29229.

**Wu, X., Li, Y., Shi, Y., Song, Y., Zhang, D., Li, C., Buckler, E., Zhang, Z., and Wang, T.** (2016). Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. Plant Biotechnol. J. **14**:1551–1562.

**Xiao, Y., Tong, H., Yang, X., Xu, S., Pan, Q., Qiao, F., Raihan, M., Luo, Y., Liu, H., Zhang, X., et al.** (2016). Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytol. **210**:1095–1106.

**Xing, A., Gao, Y., Ye, L., Zhang, W., Cai, L., Ching, A., Llaca, V., Johnson, B., Liu, L., Yang, X., et al.** (2015). A rare SNP mutation in *Brachytic2* moderately reduces plant height and increases yield potential in maize. J. Exp. Bot. **66**:3791–3802.

**Xue, Y., Warburton, M., Sawkins, M., Zhang, X., Setter, T., Xu, Y., Grudloyma, P., Gethi, J., Ribaut, J., Li, W., et al.** (2013). Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor. Appl. Genet. **126**:2587–2596.

**Xue, S., Bradbury, P., Casstevens, T., and Holland, J.** (2016). Genetic architecture of domestication-related traits in maize. Genetics **204**:99–113.

**Yalcin, B., Flint, J., and Mott, R.** (2005). Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics **171**:673–681.

**Yan, J., Kandianis, C., Harjes, C., Bai, L., Kim, E., Yang, X., Skinner, D., Fu, Z., Mitchell, S., Li, Q., et al.** (2010). Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. Nat. Genet. **42**:322–327.

**Yan, J., Warburton, M., and Crouch, J.** (2011). Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci. **51**:433–449.

**Yang, J., Benyamin, B., McEvoy, B., Gordon, S., Henders, A., Nyholt, D., Madden, P., Heath, A., Martin, N., Montgomery, G., et al.** (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. **42**:565–569.

**Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K., Yang, N., Li, Y., Zhong, T., et al.** (2013). CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. USA **110**:16969–16974.

**Yang, J., Zaitlen, N., Goddard, M., Visscher, P., and Price, A.** (2014a). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. **46**:100–106.

**Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., Wen, W., Liu, J., Li, J., and Yan, J.** (2014b). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet. **10**:e1004573.

**Yang, J., Jiang, H., Yeh, C., Yu, J., Jeddeloh, J., Nettleton, D., and Schnable, P.** (2015). Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J. **84**:587–596.

**Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., et al.** (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. **48**:927–934.

**Yu, J., and Buckler, E.** (2006). Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. **17**:155–160.

**Yu, J., Pressoir, G., Briggs, W., Bi, I.V., Yamasaki, M., Doebley, J., McMullen, M., Gaut, B., Nielsen, D., and Holland, J.** (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**:203–208.

**Yu, J., Holland, J., McMullen, M., and Buckler, E.** (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178**:539–551.

**Zeng, Z.** (1994). Precision mapping of quantitative trait loci. Genetics **136**:1457–1468.

**Zhang, Y., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S.** (2005). Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics **169**:2267–2275.

**Zhang, Z., Ersoz, E., Lai, C., Todhunter, R., Tiwari, H., Gore, M., Bradbury, P., Yu, J., Arnett, D., Ordovas, J., et al.** (2010). Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **44**:355–360.

**Zhang, Z., Guillaume, F., Sartelet, A., Charlier, C., Georges, M., Farnir, F., and Druet, T.** (2012). Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification. Bioinformatics **28**:2467–2473.

**Zhang, N., Gibon, Y., Wallace, J., Lepak, N., Li, P., Dedow, L., Chen, C., So, Y., Kremling, K., Bradbury, P., et al.** (2015). Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiol. **168**:575–583.

**Zhang, X., Warburton, M., Setter, T., Liu, H., Xue, Y., Yang, N., Yan, J., and Xiao, Y.** (2016). Genome-wide association studies of drought-related metabolic changes in maize using an enlarged SNP panel. Theor. Appl. Genet. **129**:1449–1463.

**Zhou, X., and Stephens, M.** (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821–824.

**Zila, C., Samayoa, L., Santiago, R., Butron, A., and Holland, J.** (2013). A genome-wide association study reveals genes associated with fusarium ear rot resistance in a maize core diversity panel. G3 (Bethesda) **3**:2095–2104.

**Zila, C., Ogut, F., Romay, M., Gardner, C., Buckler, E., and Holland, J.** (2014). Genome-wide association study of *Fusarium* ear rot disease in the U.S.A. maize inbred line collection. BMC Plant Biol. **14**:372.