

## SI GENOME TO PHENOME

# Crop genome-wide association study: a harvest of biological relevance

**Hai-Jun Liu\*** and **Jianbing Yan\****National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China*

Received 23 July 2018; revised 13 October 2018; accepted 22 October 2018; published online 28 October 2018.

\*For correspondence (e-mails [yjianbing@mail.hzau.edu.cn](mailto:yjianbing@mail.hzau.edu.cn); [heroalone@webmail.hzau.edu.cn](mailto:heroalone@webmail.hzau.edu.cn)).**SUMMARY**

With the advent of rapid genotyping and next-generation sequencing technologies, genome-wide association study (GWAS) has become a routine strategy for decoding genotype–phenotype associations in many species. More than 1000 such studies over the last decade have revealed substantial genotype–phenotype associations in crops and provided unparalleled opportunities to probe functional genomics. Beyond the many ‘hits’ obtained, this review summarizes recent efforts to increase our understanding of the genetic architecture of complex traits by focusing on non-main effects including epistasis, pleiotropy, and phenotypic plasticity. We also discuss how these achievements and the remaining gaps in our knowledge will guide future studies. Synthetic association is highlighted as leading to false causality, which is prevalent but largely underestimated. Furthermore, validation evidence is appealing for future GWAS, especially in the context of emerging genome-editing technologies.

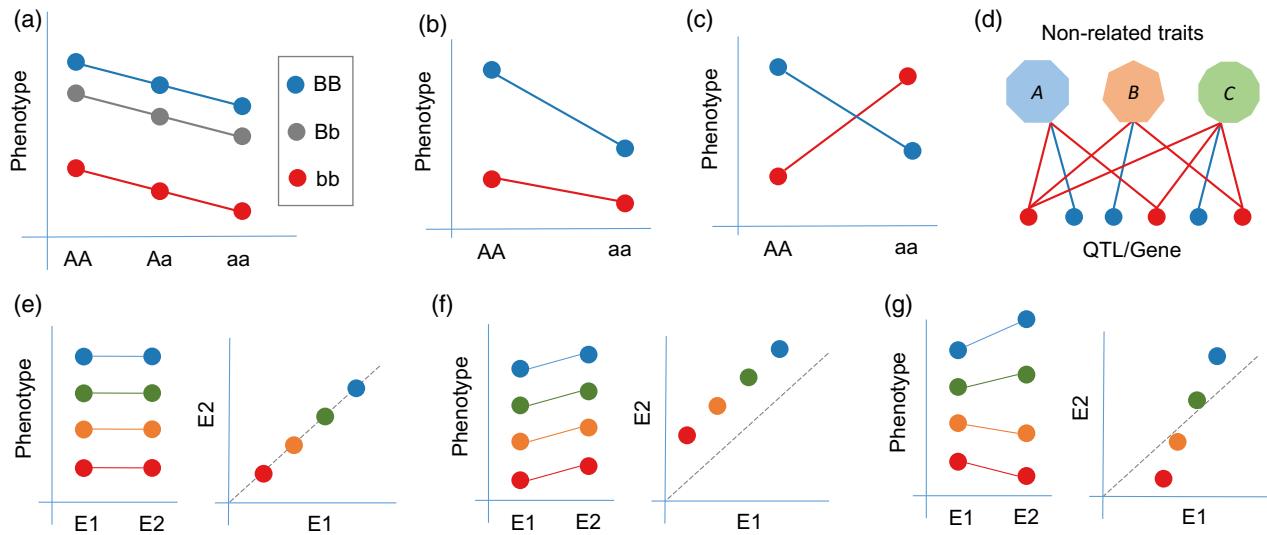
**Keywords:** GWAS, genetic architecture, epistasis, pleiotropy, phenotypic plasticity, synthetic association, genome editing.

**INTRODUCTION**

Since the concept was first applied in maize in 2001 (Thornberry *et al.*, 2001), association mapping studies in crop species have revealed links between tens of thousands of genomic regions and various traits. Association mapping is a quantitative approach for determining if a genomic variant is associated with a trait of interest using a natural population or a collection of diverse individuals. The main hypothesis states that a particular phenotype shared by a subset of individuals will be highly linked to neighboring genetic variations (linkage disequilibrium, LD; Glossary Box) in their recent ancestor, where the causal mutation and corresponding phenotype arose. Recent advances in high-throughput genotyping technologies and increases in computational power have made it possible to carry out association studies on genome-wide sets of genetic variants, an approach known as genome-wide association study (GWAS), thus greatly changing the mapping of quantitative traits. While relatively low-resolution and time-consuming methods such as linkage mapping are typically used for mapping in biparental populations, the emergence of GWAS provides an opportunity to discover genes or regions associated with

given traits in a relatively high resolution and unbiased manner in broad-based and diverse populations. GWAS can also reveal the global landscape of a trait, known as its genetic architecture (Figure 1), a term used to describe the genetic basis of a trait based on information regarding the number of causative genes or alleles, their interactions, and the distribution and patterns of their effects (Hansen, 2006).

Sequencing-based GWAS has become a routine tool in crop genetics over the last decade, making outstanding achievements in two major ways. The first is in redefining the concept of a ‘trait’, from conventional developmental traits (e.g. Buckler *et al.*, 2009; for maize flowering; Huang *et al.*, 2010; for rice agronomic traits), to individual responses to environmental factors (biotic or abiotic stress tolerance) (e.g. Li *et al.*, 2017 for rice blast resistance; Wang *et al.*, 2016 and Guo *et al.*, 2018 for drought tolerance in maize and rice, respectively; Kuroha *et al.*, 2018 for periodic flooding adaptation), to large-scale molecular-level quantification (e.g. Fu *et al.*, 2013 and Chen *et al.*, 2018 for determination of the structural transcriptome of the maize kernel, respectively; Chen *et al.*, 2014 and Wen



**Figure 1.** Complex principles of genetic architecture. (a) Demonstration of additive and dominant effects for a two-locus model. Locus A only presents an additive effect, and dominance of locus B occurs as the phenotype of the heterozygous allele deviates from the average of the two homozygous alleles. These two loci show no epistatic effects with each other, as displayed in (b) and (c). (b) The different alleles of locus A show distinct effects on trait variance among different states of locus B, with the same direction. (c) The alternative alleles of locus A express similar effects on trait variance with opposite direction under different backgrounds of locus B. Because inbred lines are typically studied in genome-wide association studies, the heterozygous allele has been removed to simplify the interaction models in (b) and (c). (d) Presence of pleiotropy in red quantitative trait loci (QTLs) or genes as these show effects on at least two non-correlated traits; blue QTLs or genes represent non-pleiotropic loci as they only contribute to one trait. (e) Absence of plasticity. No phenotypic difference exists under different environments (E1 and E2); each colored point represents a different genotype. (f) Presence of phenotypic plasticity without existence of a genotype–environment interaction ( $G \times E$ ), as all genotypes alter their phenotypes in parallel under different environments. (g) Co-existence of phenotypic plasticity and  $G \times E$ , as all genotypes alter their phenotypes but to distinct extents or/and in distinct directions under different environments.

et al., 2014a with respect to metabolism in rice and maize), to more complex phenotypic variations as long as they are heritable and measurable, such as rice heterosis (Huang et al., 2015) and maize haploid male fertility (Ma et al., 2018a). The level of precision is expected to extend further in the next decade, from single-trait variables to a variable vector depicting a dynamic developmental process with the recent development of plant phenomics techniques (Yang et al., 2013b; Tardieu et al., 2017; Singh et al., 2018). The second impressive achievement is that as increasing numbers of crop species enter the ‘omics era, GWAS is being performed not only in cereals (in particular rice and maize) but also in a broad range of crops, including soybean (i.e., Hwang et al., 2014; Wen et al., 2014b; Zhang et al., 2015a; Fang et al., 2017a; Leamy et al., 2017), cotton (Fang et al., 2017b; Wang et al., 2017b; Du et al., 2018; Ma et al., 2018b), tomato (Lin et al., 2014; Tieman et al., 2017; Zhu et al., 2018), cucumber (Shang et al., 2014; Zhang et al., 2015c), sesame (Wei et al., 2015), peanut (Pandey et al., 2014; Zhang et al., 2017b), peach (Cao et al., 2016), and lettuce (Zhang et al., 2017a). These studies, together with purpose-developed populations, catalogs of allelic variants, and corresponding genotype–phenotype associations, provide unprecedented resources for understanding crop functional genomics. These studies have not only validated known trait associations, but also identified new favorable haplotypes or, in some cases, revealed previously unknown pathways.

Despite great success, current GWAS analysis has clear limitations, especially issues of population structure correlation and low-frequency causal alleles leading to false-negative results (Korte and Farlow, 2013). For example, only one gene (*ZmCCT*) was revealed for flowering time using a diverse association mapping panel consisting of 500 inbred lines (Yang et al., 2013a), as flowering time is a typical adaptive trait and is always confounded (i.e., highly correlated) with population structure. It has been widely accepted that many false negatives occur for such confounded traits when correcting for population structure in GWAS (Huang and Han, 2014). Another example showed only five inbred lines (<1%) possessing functionally alternative alleles at the *Brachytic2* locus for plant height among 527 lines (Xing et al., 2015); it is thus not possible to identify this locus using routine association mapping analysis. A similar phenomenon is also seen in rice, where (putative) causal alleles within most of the cloned yield-related quantitative trait loci (QTLs) are at low frequency in diverse germplasms (2% for *Ghd7*, Xue et al., 2008; Lu et al., 2012; 1% for *GS3*, Fan et al., 2006; Mao et al., 2010; 2% for *qGL3*, Zhang et al., 2012; 6% for *TGW6*, Ishimaru et al., 2013). Solving these issues by developing novel statistical models to explore rare functional alleles (Zhu et al., 2011; Listgarten et al., 2013; Kaakinen et al., 2017) or employing artificially designed populations to balance allelic frequencies and control population structure (Buckler et al., 2009; Dell’Acqua et al., 2015; Romero Navarro et al.,

2017; Wen *et al.*, 2018) are of great importance for GWAS. As these matters have been extensively discussed previously (Gibson, 2012; Korte and Farlow, 2013; Xiao *et al.*, 2017; Cockram and Mackay, 2018), they will not be the further focus of the present review.

For the foreseeable future, GWAS will continue to be a common research tool for providing a bird's-eye view of the genetic structure of any trait of interest. The generation of various 'omics data will increase and enable OWAS ('omics-wide association study; Glossary Box) (Xiao *et al.*, 2017). As the number of GWAS-based candidate gene regions has greatly increased in recent years, a review of new insights into crop genetic architecture is required. We summarize here (Box 1) how recent GWAS results have led to an updated view of several important issues of genetic architecture, including epistasis, pleiotropy, and phenotypic plasticity. We also discuss the likely causes of synthetic association, an increasingly common GWAS artifact. We examine the impact of these new biological insights and problems on the practice of GWAS and emphasize the need to validate GWAS results, especially by using the increasingly popular genome-editing technology.

#### **EPISTASIS: NEGIGIBLE OR NEGLECTED?**

Epistasis represents a non-linear interaction between two or more segregating loci with different alleles across genetic backgrounds. This type of interaction between segregating loci is expected to contribute to phenotypes by biologically plausible mechanisms (Mackay, 2014). However, most studies have focused on additive genetic variance (Figure 1a), with relatively little attention paid to epistasis (Figure 1b, c), the importance of which is still being established in different biological systems. While its prominent role in heterosis has been widely recognized in various crops (Yu *et al.*, 1997; Li *et al.*, 2001; Melchinger *et al.*, 2007; Garcia *et al.*, 2008; Shen *et al.*, 2014; Jiang *et al.*, 2017), the prevalence of epistasis in maize trait architecture is thought to be small (Buckler *et al.*, 2009; Tian *et al.*, 2011; Xiao *et al.*, 2016), or results in a large effect only at specific loci (Studer and Doebley, 2011; Durand *et al.*, 2012). However, many studies indicate that epistasis is pervasive in contributing to various quantitative trait phenotypes (Manicacci *et al.*, 2009; Würschum *et al.*, 2011; Zhang *et al.*, 2015a; Wen *et al.*, 2016; He *et al.*, 2017; Luo *et al.*, 2017; Mathew *et al.*, 2018) and can be further used to improve the accuracy of trait prediction for both inbreds and hybrids (Maurer *et al.*, 2015; Santos *et al.*, 2015; Luo *et al.*, 2017). These results were observed in a variety of populations, such as a recombinant inbred line (RIL) population, a multi-parent advanced-generation inter-cross (MAGIC) population, and diversity panels, and for different traits, such as morphological characteristics, resistance to disease, and cellular metabolite levels.

Even though the importance of epistasis is increasingly recognized, the detection of epistatic effects is difficult, especially in GWAS using populations of unrelated individuals. Large numbers of variants, each present at a low frequency, create a major challenge, leading to low statistical confidence in the level of epistasis. Thanks to recently proposed efficient computing algorithms (Hemani *et al.*, 2011; Gynenesi *et al.*, 2012; Lishout *et al.*, 2015; Zhang *et al.*, 2016; Cowman and Koyuturk, 2017) and alternative, non-exhaustive modeling approaches (Guo *et al.*, 2014; Leem *et al.*, 2014; Karkkainen *et al.*, 2015; Wang *et al.*, 2015a; Zhang *et al.*, 2016; Mathew *et al.*, 2018), both the computational cost and multiple-testing burden can be effectively addressed and even high-order interactions can be uncovered, provided there is a sufficient number of individuals in the population under study. Still, because of the need to test alleles pairwise, those with relatively high frequency will provide the greatest probability of discovering epistasis, and low-frequency alleles at either locus will reduce the statistical power. This could be why epistatic effects appear to be population dependent. For a panel with diverse lines, a large population size is needed to reach the high statistical power necessary to uncover QTLs with moderate or subtle epistatic effects. However, artificially designed populations with balanced allele frequency, such as MAGIC (Mathew *et al.*, 2018), promise to greatly facilitate epistasis discovery. Interestingly, Wei *et al.* (2018) found that loci identified using genotypic-variability-based GWAS can be used to evaluate potential epistatic interactions.

In general, integrating improved algorithms with experimental crop populations will improve the accuracy of future interaction studies. These insights will enable future crop engineering, as demonstrated by a pioneering study in tomato that optimized inflorescence architecture and high yield by eliminating undesirable epistasis (Soyk *et al.*, 2017).

#### **PLEIOTROPY: WHAT IS THE PROMISE OF GWAS?**

Understanding pleiotropy, a well known phenomenon identified a century ago (Stearns, 2010) in which one allele or gene affects multiple phenotypes (Figure 1d), is crucial for understanding genetic mechanisms and for simultaneous breeding of multiple complex traits. While the presence of pleiotropy is often mistakenly assumed when a locus is found to be associated with two or more traits, there are at least two additional assumptions that need to be emphasized. The first is that the association with two or more traits comes from the same causal gene within the locus, which is particularly important as the mapping resolution in crop GWAS limits the ability to discriminate between multiple candidates in high LD with each other. The second assumption is that the associated traits should be uncorrelated or, more precisely,

they should be independently affected by the same causal gene rather than affected by confounding between the phenotypes (Solovieff *et al.*, 2013). This makes the identification of pleiotropy more difficult in both plant species and human beings.

Pleiotropic effects have largely not been systematically explored in crops, even though a number of pleiotropic genes have been implicated in specific case studies. For example, *Ghd7* controls heading date, grain number, plant height, and flag leaf area in rice (Xue *et al.*, 2008; Tan *et al.*, 2012; Weng *et al.*, 2014), and its ortholog in maize, *ZmCCT*, affects both flowering time and disease resistance (Hung *et al.*, 2012; Yang *et al.*, 2013a; Wang *et al.*, 2017a). Several recent studies have attempted to evaluate the shared genetic basis for crop genotype–phenotype links on a genome-wide basis. Schulthess *et al.* (2017) identified several potential pleiotropic loci for wheat yield-related traits using multiple-trait-based association mapping. This study also presented a simulation analysis of factors, including minor allele frequency and QTL effect sizes and distances for different traits, affecting the power to identify pleiotropy, which can be used as reference for future study designs. However, the population diversity and low genotypic density in this study limited the differentiation of true pleiotropy from the effects of closely linked genes or spurious pleiotropy (Glossary Box). By integrating a large and diverse soybean collection with whole-genome sequencing, Fang *et al.* (2017a) found that genetic sharing for different traits is widespread and, interestingly, the *E2* locus exhibits pleiotropy for both yield and seed quality. Pleiotropy has also been identified for maize carbon and nitrogen metabolism (Zhang *et al.*, 2015b).

Beyond the accumulative GWAS of various traits for crop species, improved methods provide an unprecedented opportunity to effectively dissect the contribution of pleiotropy to crop trait variation. Commonly applied multi-trait-based methods have been comprehensively reviewed (Solovieff *et al.*, 2013; Hackinger and Zeggini, 2017) and can be applied to crop studies. We emphasize that whole-genome gene expression QTL (eQTL) results can also be used to analyze pleiotropy, and Zhu *et al.* (2016) have proposed a strategy to integrate eQTL data with knowledge of QTL resulting in observable phenotypes to identify the variants displaying pleiotropic effects for both trait and gene expression. This method was further applied by Hannon *et al.* (2017) to study the pleiotropic variants associated with quantitative traits and DNA methylation.

Pleiotropy seems common and important for increasing trait prediction accuracy for complex human traits (Maier *et al.*, 2018), but our understanding of this phenomenon is still greatly limited in crop genetics. As the number of genotype–phenotype links grows, pleiotropy will draw greater attention from crop researchers.

## PHENOTYPIC PLASTICITY: A POWER TO NURTURE THE NATURE

Most current efforts focus on mapping the genetics of complex trait variance in populations; however, the phenotypic performance of an individual can change with fluctuating environment. The ability to respond to environmental change by expressing variable phenotypes without genotypic change is called phenotypic plasticity (Figure 1e–g). The phenomenon of different alleles displaying varied plastic responses is described as genotype–environment interaction ( $G \times E$ ). The significance of phenotypic plasticity, which was considered to be genetically heritable, has been realized for several decades (Bradshaw, 1965; Weaver and Ingram, 1969; Schlichting, 1986; Gavrilov and Scheiner, 1993; Dewitt *et al.*, 1998). These preliminary studies described morphological changes mainly by performing theoretical or simulation investigations. Various quantitative genetic models controlling phenotypic plasticity have been proposed, including the allelic sensitivity model and the over-dominance model (Scheiner, 1993). Current genetic mapping studies provide an unparalleled chance to explore the quantitative architecture of changing phenotypic responses and possibly even uncover the underlying genes (Wang *et al.*, 2013; Zhai *et al.*, 2014). Phenotypic plasticity adds a significantly complex layer to the genetic architecture of complex traits.

Phenotypic measures of plasticity include two aspects: the degree of change in the phenotypic mean across environments and the pattern of such change (Schlichting and Levin, 1984). The coefficient estimated from regression analysis, and summary statistics including range, standard deviation, and coefficient of variation are simple measures. However, these only describe the amount of change without indexing the pattern of change. The coefficient obtained from regression along an environmental gradient can be specified as the reaction norm to describe both the degree and pattern of plastic change (Eberhart and Russell, 1966; Freeman, 1973); however, the environmental gradient is trait and genotype dependent and therefore difficult to characterize when numbers of individuals are measured. After obtaining the measurements, phenotypic plasticity can be mapped genetically and compared in the same way as phenotypic mean, or modeled along specific frameworks (Wang *et al.*, 2013; Zhai *et al.*, 2014).

In a recent pioneering study, Kusmec *et al.* (2017) analyzed 23 agronomic traits in 4–11 environments using a nested association mapping (NAM) population consisting of about 5000 RILs. By partitioning the trait into phenotypic mean, linear, and non-linear plasticities with the Bayesian Finlay–Wilkinson Regression (FWR) (Su *et al.*, 2006), structurally and functionally distinct candidate genes were found in association with mean and plastic phenotypes. This distinct genetic architecture provides a further

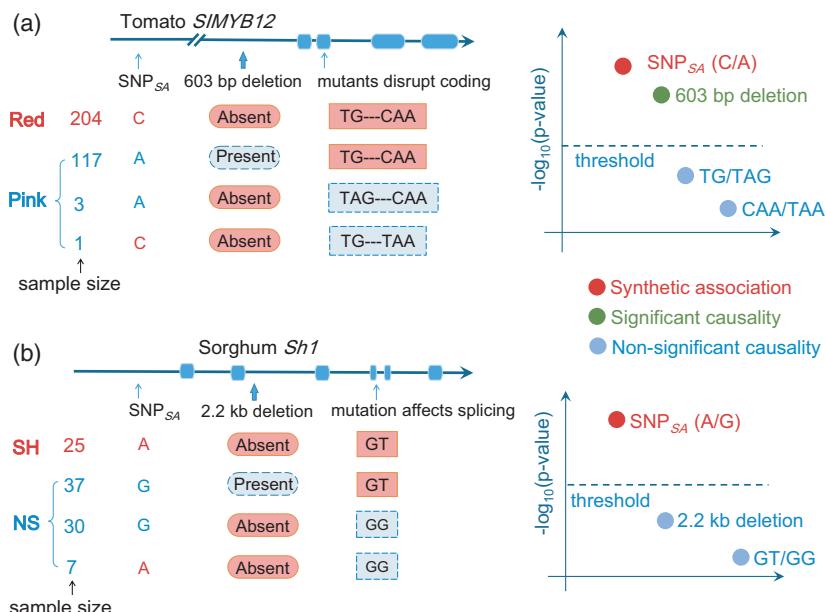
opportunity to simultaneously manage trait mean and plasticity for a given environment or for a changing climate (Nicotra *et al.*, 2010). Flowering time is a classic example used in model and crop plants to study the genetic control of both mean and plastic phenotypes (Ungerer *et al.*, 2003; Anderson *et al.*, 2012; Brachi *et al.*, 2013; Mendez-Vigo *et al.*, 2016; Mao *et al.*, 2017). Another recent study on sorghum flowering time (Li *et al.*, 2018) indicated that modeling plasticity can not only explain the genetic response to different environments, but also enable highly accurate prediction of trait performance in new environments. Beyond such studies on individual traits, a new model has been proposed (Zhou *et al.*, 2015) to explore the genetic architecture of phenotypic plasticity in multiple correlative traits.

Even though the statistical framework (Wang *et al.*, 2013; Zhai *et al.*, 2014; Zhou *et al.*, 2015) has been established for years and representative studies (Kusmec *et al.*, 2017; Li *et al.*, 2018) have been performed, our understanding of the mechanisms of plasticity and its effect on shaping crop diversity along environmental gradients is still limited. The lack of deep insights into this common and important issue may be because of the massive data sets needed, including environmental measurements in

addition to standard genotypic and phenotypic variations. Particularly, multiple environmental conditions along a gradient of variation are necessary to unravel the quantitative attributes of phenotypic responses. When describing the interaction between genome and environment, understanding phenotypic plasticity helps interpret the evolutionary and environmental forces modifying genetic architectures (Josephs, 2018). Therefore, investigating phenotypic plasticity will also contribute to crop improvement, when integrated with breeding programs that are facing climate change and instability.

## SYNTHETIC ASSOCIATION: MISLEADING FOR CAUSALITY

Confusingly, sometimes the non-causative loci show more significant associations in GWAS than the causative ones. In other words, the causative genes are sometimes located away from the GWAS peaks. This has been observed in a number of association studies in plants including *Arabidopsis* (Atwell *et al.*, 2010; Kerdaffrec *et al.*, 2016), rice (Huang *et al.*, 2010, 2011; Yano *et al.*, 2016), sorghum (Lin *et al.*, 2012), and tomato (Lin *et al.*, 2014). This misleading association is called synthetic association (or ‘ghost association’), which is presumed to be caused by LD between common tagged markers and rare causative variants



**Figure 2.** Synthetic association is likely to be caused by the presence of multiple-causal alleles within a gene. (a) Modified example of tomato *S/MYB12* for fruit color (Lin *et al.*, 2014). The single-nucleotide polymorphism (SNP) showing synthetic association (SNP<sub>SA</sub>) is the most significant, while three causal alleles are identified with less or no significance. Each mutation in the promoter and coding regions can individually alter the phenotype from red to pink, which causes no perfect match of any variant to the trait. Interestingly, the SNP<sub>SA</sub> alleles (205 versus 120) are correlated with the combination of mutated alleles versus wild-type genotypes (204 versus 121) of the other variants. Even though it is less significant, the deletion locus can be identified because of the rare ( $n = 4$ ) frequency of the other two causal variants. (b) Another example is simplified from sorghum *Sh1* for shattering (Lin *et al.*, 2012). Two causal alleles are identified by experimental exploration. For the deletion locus, while the deletion is present in 37 individuals, plants carrying the absent allele show shattering (SH) and non-shattering (NS) in comparable numbers (25 versus 37). This makes the deletion locus undetectable in standard single-variant-based association mapping. A similar situation occurs for the splicing variant. Any variant (SNP<sub>SA</sub>) correlated with the wild-type allele (with the SH trait) and the combination of both functional mutants, if present, will unexpectedly be uncovered but have nothing to do with causality. This case is different from (a) as both causal loci (the deletion locus and the splicing variant) are common and undetectable (62 versus 37).

(Dickson *et al.*, 2010; Chang and Keinan, 2012). The rare-allele hypothesis can be understood such that the common variant would show significant association signal with a given trait when it is linked to a low frequency but large-effect causative variant. This can also explain the ‘missing heritability’ issue, because the common markers (identified with greater significance) can explain only a limited fraction of trait variance as a result of their imperfect association with the causative low-frequency variant. However, some cases do not follow the simple rare-allele assumption and can be explained by another phenomenon, that the trait variation is caused by multiple alleles within one gene (Lin *et al.*, 2012; Yano *et al.*, 2016).

We tend to interpret the synthetic association issue as the ‘presence of multiple causative alleles’, in which the standard single-variant-based GWAS has insufficient power to detect any of these because of the genetic interference between different alleles (Figure 2). Given that mutation constantly generates new variants, multiple independent alleles within one gene leading to the same phenotype could be common. Haplotype- or gene-based methodologies have good potential for identifying such situations, although current haplotype-based association mapping is still imperfect (Hayes, 2013) and both inferring accurate haplotypes and incorporating haplotypes into association mapping remain particularly challenging in plants. Additionally, the multiple functional alleles in such cases of synthetic association with high LD are unlikely to

be captured by haplotype phasing. A better understanding of the underlying causes of synthetic associations would help in the design of future studies to detect causative genetic variants while avoiding artifacts.

### VALIDATION: A NECESSARY ADOPTION FOR FUTURE GWAS

Even with a strong theoretical foundation and efforts to remove undesired noise (i.e., population structure) and employ strict probability cut-offs, false-positive associations will still occur due to the enormous number of statistical inferences and other unaccounted factors, such as low-accuracy genotype calling at some loci (Browning and Yu, 2009), small population size (Finno *et al.*, 2014), and synthetic associations introduced above. This calls for an independent validation process, which has seldom been incorporated into GWAS design. Validation includes at least two methodologies: one validating candidates of interest in different populations, the associations would be assumed to be more likely when being detected in independent studies; and the other using laboratory experiments, such as candidate gene knock-out, over-expression, or genetic complementation. Cross-population validation is currently achieved by integrating association mapping in diverse panels or linkage mapping in RIL population(s) or  $F_2$  populations. Taking as an example the recent study on cloning *ZmCCT9*, which affects maize flowering time (Huang *et al.*, 2018), this locus was simultaneously

#### **Box 1 Highlights**

- Crop GWAS has ushered a transition to ‘omics-wide association mapping (OWAS), promising a better understanding of genetic architecture of complex traits.
- The large number of studies provides an unprecedented opportunity to increase in-depth understanding of the classical concepts of epistasis and pleiotropy.
- Phenotypic plasticity is largely ignored and requires intensive data collection and general statistical modeling.
- Synthetic association exists frequently in GWAS and is considered to result from the presence of multiple independent alleles within a locus.
- Emerging novel technologies such as genome editing can be used for further GWAS validation.

#### **Box 2 Open questions**

- What is the pervasiveness of the non-additive effect, is this effect dominant for certain traits, and what is the underlying mechanism?
- Is it possible to map causal genes (or variants) with ultra-high resolution to understand the generality of pleiotropy and how to apply it in future genetic improvement?
- Which genes and mechanisms contribute to phenotypic plasticity, and will identify these help to predict crop responses to climate change?
- Is synthetic association widespread, and how can it be addressed effectively?
- How far are we away from the precise design of new cultivars by integrating natural and created variations?

## Glossary Box

- **Linkage disequilibrium (LD):** is a phenomenon in which co-occurrence of alleles at different loci is non-random in a given population, being either higher or lower than expected if they were independent. The presence of genetic linkage makes it unnecessary for association studies to examine every polymorphism since linked variants are strongly correlated.
- **Genome-wide association study (GWAS):** is a statistical approach of mapping quantitative trait loci to link phenotypes of interest to whole-genomic genotypes by taking advantage of historic linkage disequilibrium.
- **'Omics-wide association study (OWAS):** extends genome-wide association studies to multiple 'omic variations, including genomics, transcriptomics, proteomics, and metabolomics, with the aim of characterizing a full molecular functional and dynamic picture of variations in phenotype.
- **Quantitative trait locus (QTL):** is a genomic confidence interval associated with a trait of interest, which varies in degree of effect size and physical length, and includes at least one causal gene or other functional element. QTLs exert main, epistatic, and interaction with the environment effects, while the main effects can be additive or dominant.
- **Epistasis:** describes an interaction relationship between genes (or loci) in which the effect of one gene can vary (in size or even direction) among different alleles of the other gene, leading to non-linear consequences for phenotypic variation. Epistasis can modify effects in an additive and/or dominant manner at the interacting loci, with epistatic effect models described as additive-by-additive, additive-by-dominance, and dominance-by-dominance. Additive-by-additive interactions are the most studied as only a small population is required.
- **Pleiotropy:** is a phenomenon in which one gene directly contributes to more than one seemingly unrelated phenotypic trait. The likely underlying mechanism is that the product of a given gene can be either used by various cells or function in cascade-like signaling to various targets. As genes usually function as networks and developmental phenotypes are interactive, it is very hard to distinguish true biological pleiotropy from mediated pleiotropy and spurious pleiotropy. Mediated pleiotropy occurs when phenotype 1 lies on a causal path to phenotype 2; an association between gene(s) with phenotype 1 will thus also occur with phenotype 2. Spurious pleiotropy is reflected by both genotypic and phenotypic aspects: the identified variant is in high linkage disequilibrium with two causal variants in distinct genes that contribute to different phenotypes; or different phenotypes are misclassified into one phenotype such that any causal variant/gene affects one of them and a spurious association occurs for the other.
- **Phenotypic plasticity:** is used to describe all kinds of phenotypic responses to environmental change without any change of genome sequence. In theory, not all phenotypic plasticity is exactly the same, but it covers the concept of genotype-environment interaction ( $G \times E$ ), which presents a narrower situation that different alleles respond to environmental change to different degrees. However, in practice,  $G \times E$  is almost equivalent to phenotypic plasticity with only a slight difference in emphasis. Phenotypic plasticity is considered more important for plants than animals due to their immobility and has been revealed to be highly relevant to plant traits including flowering timing, leaf shape variations, allocation of soil nutrients, and size of seeds.
- **CRISPR-based genome editing:** is a genetic engineering technology targeting specific genome locations with operations including insertion, deletion, modification, or replacement. CRISPR is an abbreviation of Clustered Regularly Interspaced Short Palindromic Repeats, a family of DNA sequences in Bacteria and Archaea that play key roles in the prokaryotic defense system. CRISPR associated proteins (Cas) process these sequences and cut matching viral DNA sequences. CRISPR/Cas form the basis of the emerging highly efficient and specific CRISPR-based genome-editing technology.

identified by NAM (Buckler *et al.*, 2009) and maize-teosinte RIL populations under association and linkage mapping. Furthermore, the causal allele, an InDel of a harbinger-like transposon, has been identified in an association panel containing 513 diverse maize inbred lines (Li *et al.*, 2013) and validated in the above two populations used to map the locus. Another model case is the association mapping of rice chlorophyll content in a diverse panel of 529 individuals followed by three customized  $F_2$  populations to

validate GWAS signals (Wang *et al.*, 2015b). With increasing numbers of populations and corresponding genotypic and phenotypic variations, association mapping in multiple populations is becoming possible and should make the observed hits much more effective and reliable. At this point, cross-population analysis could be extended to cross-species analysis for similar or homologous traits. A better understanding of the conserved, differentiated, and/or dynamic genetic architecture for any trait of interest will

be valuable, and such cross-species analysis has already been implemented in rice (Huang *et al.*, 2011) and cereals (Chen *et al.*, 2016; Liu *et al.*, 2017).

Beyond statistical inference, molecular and genetic experimentation is a reliable way to validate GWAS hits. This is still difficult because of the low throughput and case dependence of most wet experiments. Fortunately, the emerging genome-editing technologies promise an effective and high-throughput approach. Recently, two Chinese teams have simultaneously created genome-wide targeted mutant libraries using the CRISPR/Cas9 technique in rice (Lu *et al.*, 2017; Meng *et al.*, 2017). Obtaining the functional gene within each GWAS peak should be quicker than ever expected by combining high-throughput forward- and reverse-genetic techniques.

## CONCLUSIONS: PROSPECTS AND CHALLENGES

The power of crop GWAS to explore the genetic architecture of complex traits has been demonstrated in multiple species, and this number will continue to increase rapidly. However, most studies are of limited scope to the main (additive) effect of genetic architecture. This is why the present review attempts to restate the complexity of the concept of genetic architecture, exploring the architecture beyond additive effects and underscoring the importance of understanding trait variability (Box 2). This complexity is not only the result of differences in gene action, but also determined by ontogenetic gene networks or even epigenetic effects, and the interaction with surroundings, including living neighbors and greatly changing environmental conditions. Compared with unprecedented achievements in the study of main effects, the application of GWAS to non-linear effects has been limited, providing only a rudimentary view of the comprehensive picture of genetic architecture.

Sun and Wu (2015) draw a complete picture of genetic architecture and propose a conceptual framework. They consider the current genetic mapping, i.e., linking genomic variants to individual phenotypes, as the first stage, followed by functional mapping, systems mapping, network and ecosystem mapping, and consideration of biological mechanisms, treating each trait as a dynamic vector, integrating full 'omics variations, and incorporating the role of ecological interactions in the formation of complex traits. Such complexity makes future mapping a great challenge for data collection, integration, and statistical modeling.

## ACKNOWLEDGEMENTS

This research was supported by the National Key Research and Development Program of China (2016YFD0100803, 2016YFD0101003), National Natural Science Foundation of China (31525017), the Hubei Provincial Natural Science Foundation of China (2015CFA008), and the Postdoctoral Talent Innovation Program of China (BX201700092).

## CONFLICT OF INTEREST

The authors declare no competing interests.

## REFERENCES

- Anderson, J.T., Inouye, D.W., McKinney, A.M., Colautti, R.I. and Mitchell-Olds, T. (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proc. Biol. Sci.* **279**, 3843–3852. <https://doi.org/10.1098/rspb.2012.1051>.
- Atwell, S., Huang, Y.S., Vilhjálmsson, B.J. *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627–631. <https://doi.org/10.1038/nature08800>.
- Brachi, B., Faure, N., Bergelson, J., Cuguen, J. and Roux, F. (2013) Genome-wide association mapping of flowering time in *Arabidopsis thaliana* in nature: genetics for underlying components and reaction norms across two successive years. *Acta Bot. Gallica*, **160**, 205–219. <https://doi.org/10.1080/12538078.2013.807302>.
- Bradshaw, A.D. (1965) Evolutionary significance of phenotypic plasticity in plants. *Adv. Genet.* **13**, 115–155. [https://doi.org/10.1016/S0065-2660\(08\)60048-6](https://doi.org/10.1016/S0065-2660(08)60048-6).
- Browning, B.L. and Yu, Z. (2009) Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* **85**, 847–861. <https://doi.org/10.1016/j.ajhg.2009.11.004>.
- Buckler, E.S., Holland, J.B., Bradbury, P.J. *et al.* (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714–718. <https://doi.org/10.1126/science.1174276>.
- Cao, K., Zhou, Z., Wang, Q. *et al.* (2016) Genome-wide association study of 12 agronomic traits in peach. *Nat. Commun.* **7**, 13246. <https://doi.org/10.1038/ncomms13246>.
- Chang, D. and Keinan, A. (2012) Predicting signatures of “synthetic associations” and “natural associations” from empirical patterns of human genetic variation. *PLoS Comput. Biol.* **8**(7), e1002600. <https://doi.org/10.1371/journal.pcbi.1002600>.
- Chen, W., Gao, Y., Xie, W. *et al.* (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721. <https://doi.org/10.1038/ng.3007>.
- Chen, W., Wang, W., Peng, M. *et al.* (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* **7**, 12767. <https://doi.org/10.1038/ncomm12767>.
- Chen, Q., Han, Y., Liu, H. *et al.* (2018) Genome-wide association analyses reveal the importance of alternative splicing in diversifying gene function and regulating phenotypic variation in maize. *Plant Cell*, **30**, 1404–1423. <https://doi.org/10.1105/tpc.18.00109>.
- Cockram, J. and Mackay, I. (2018) Genetic mapping populations for conducting high-resolution trait mapping in plants. *Adv. Biochem. Eng. Biotechnol.* **164**, 109–138. [https://doi.org/10.1007/10\\_2017\\_48](https://doi.org/10.1007/10_2017_48).
- Cowman, T. and Koyuturk, M. (2017) Prioritizing tests of epistasis through hierarchical representation of genomic redundancies. *Nucleic Acids Res.* **45**(14), e131. <https://doi.org/10.1093/nar/gkx505>.
- Dell'Acqua, M., Gatti, D.M., Pea, G. *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. <https://doi.org/10.1186/s13059-015-0716-z>.
- Dewitt, T.J., Sih, A. and Wilson, D.S. (1998) Costs and limits of phenotypic plasticity. *Trends Ecol. Evol.* **13**, 77–81. [https://doi.org/10.1016/S0169-5347\(97\)01274-3](https://doi.org/10.1016/S0169-5347(97)01274-3).
- Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. and Goldstein, D.B. (2010) Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**(1), e1000294. <https://doi.org/10.1371/journal.pbio.1000294>.
- Du, X., Huang, G., He, S. *et al.* (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796–802. <https://doi.org/10.1038/s41588-018-0116-x>.
- Durand, E., Bouchet, S., Bertin, P., Ressayre, A., Jamin, P., Charcosset, A., Dillmann, C. and Tenaillon, M.I. (2012) Flowering time in maize: linkage and epistasis at a major effect locus. *Genetics*, **190**, 1547–1562. <https://doi.org/10.1534/genetics.111.136903>.

- Eberhart, S.A. and Russell, W.A.** (1966) Stability parameters for comparing varieties. *Crop Sci.* **6**, 36–40. <https://doi.org/10.2135/cropsci1966.0011183X000600010011x>.
- Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X. and Zhang, Q.** (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171. <https://doi.org/10.1007/s00122-006-0218-1>.
- Fang, C., Ma, Y., Wu, S. et al.** (2017a) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**, 161. <https://doi.org/10.1186/s13059-017-1289-9>.
- Fang, L., Wang, Q., Hu, Y. et al.** (2017b) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **49**, 1089–1098. <https://doi.org/10.1038/ng.3887>.
- Finno, C.J., Aleman, M., Higgins, R.J., Madigan, J.E. and Bannasch, D.L.** (2014) Risk of false positive genetic associations in complex traits with underlying population structure: a case study. *Vet. J.* **202**, 543–549. <https://doi.org/10.1016/j.tvjl.2014.09.013>.
- Freeman, G.H.** (1973) Statistical methods for the analysis of genotype–environment interactions. *Heredity (Edinb.)* **31**, 339–354. <https://doi.org/10.1038/hdy.1973.90>.
- Fu, J., Cheng, Y., Linghu, J. et al.** (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat. Commun.* **4**, 2832. <https://doi.org/10.1038/ncomms3832>.
- Garcia, A.A., Wang, S., Melchinger, A.E. and Zeng, Z.B.** (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics*, **180**(3), 1707–1724. <https://doi.org/10.1534/genetics.107.082867>.
- Gavrilovs, S. and Scheiner, S.** (1993) The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *J. Evol. Biol.* **6**, 31–48. <https://doi.org/10.1046/j.1420-9101.1993.6010031.x>.
- Gibson, G.** (2012) Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145. <https://doi.org/10.1038/nrg3118>.
- Guo, X., Meng, Y., Yu, N. and Pan, Y.** (2014) Cloud computing for detecting high-order genome-wide epistatic interaction via dynamic clustering. *BMC Bioinformatics*, **15**, 102. <https://doi.org/10.1186/1471-2105-15-102>.
- Guo, Z., Yang, W., Chang, Y. et al.** (2018) Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. *Mol. Plant*, **11**, 789–805. <https://doi.org/10.1016/j.molp.2018.03.018>.
- Gyenesi, A., Moody, J., Semple, C.A., Haley, C.S. and Wei, W.H.** (2012) High-throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics*, **28**, 1957–1964. <https://doi.org/10.1093/bioinformatics/bts304>.
- Hackinger, S. and Zeggini, E.** (2017) Statistical methods to detect pleiotropy in human complex traits. *Open Biol.* **7**, pii: 170125. <https://doi.org/10.1098/rsob.170125>.
- Hannon, E., Weedon, M., Bray, N., O'Donovan, M. and Mill, J.** (2017) Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. *Am. J. Hum. Genet.* **100**(6), 954–959. <https://doi.org/10.1016/j.ajhg.2017.04.013>.
- Hansen, T.F.** (2006) The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* **37**(1), 123–157. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110224>.
- Hayes, B.** (2013) Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* **1019**, 149–169. [https://doi.org/10.1007/978-1-62703-447-0\\_6](https://doi.org/10.1007/978-1-62703-447-0_6).
- He, S., Reif, J.C., Korzun, V., Bothe, R., Ebmeyer, E. and Jiang, Y.** (2017) Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. *Theor. Appl. Genet.* **130**, 635–647. <https://doi.org/10.1007/s00122-016-2840-x>.
- Hemani, G., Theoscharidis, A., Wei, W. and Haley, C.** (2011) EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics*, **27**, 1462–1465. <https://doi.org/10.1093/bioinformatics/btr172>.
- Huang, X. and Han, B.** (2014) Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **65**, 531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>.
- Huang, X., Sang, T., Zhao, Q. et al.** (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967. <https://doi.org/10.1038/ng.695>.
- Huang, X., Zhao, Y., Li, C. et al.** (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39.
- Huang, X., Yang, S., Gong, J. et al.** (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* **6**, 6258. <https://doi.org/10.1038/ncomms7258>.
- Huang, C., Sun, H., Xu, D. et al.** (2018) ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci.* **115**, E334–E341. <https://doi.org/10.1073/pnas.1718058115>.
- Hung, H.Y., Shannon, L.M., Tian, F. et al.** (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc. Natl. Acad. Sci.* **109**, E1913–E1921. <https://doi.org/10.1073/pnas.1203189109>.
- Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J. and Creagan, P.B.** (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **15**, 1. <https://doi.org/10.1186/1471-2164-15-1>.
- Ishimaru, K., Hirotsu, N., Madoka, Y. et al.** (2013) Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* **45**, 707–711. <https://doi.org/10.1038/ng.2612>.
- Jiang, Y., Schmidt, R.H., Zhao, Y. and Reif, J.C.** (2017) A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* **49**(12), 1741–1746. <https://doi.org/10.1038/ng.3974>.
- Josephs, E.B.** (2018) Determining the evolutionary forces shaping G × E. *New Phytol.* **219**, 31–36. <https://doi.org/10.1111/nph.15103>.
- Kaakinen, M., Magi, R., Fischer, K., Heikkilä, J., Jarvelin, M.R., Morris, A.P. and Prokopenko, I.** (2017) MARV: a tool for genome-wide multi-phenotype analysis of rare variants. *BMC Bioinformatics*, **18**, 110. <https://doi.org/10.1186/s12859-017-1530-2>.
- Karkkainen, H.P., Li, Z. and Sillanpää, M.J.** (2015) An efficient genome-wide multilocus epistasis search. *Genetics*, **201**, 865–870. <https://doi.org/10.1534/genetics.115.182444>.
- Kerdaffrec, E., Filiault, D.L., Korte, A., Sasaki, E., Nizhynska, V., Seren, U. and Nordborg, M.** (2016) Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis. *Elife*, **5**, e22502. <https://doi.org/10.7554/elife.22502>.
- Korte, A. and Farlow, A.** (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29. <https://doi.org/10.1186/1746-4811-9-29>.
- Kuroha, T., Nagai, K., Gamuyao, R. et al.** (2018) Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Science*, **361**, 181–186. <https://doi.org/10.1126/science.aat1577>.
- Kusmec, A., Srinivasan, S., Nettleton, D. and Schnable, P.S.** (2017) Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nat. Plants*, **3**, 715–723. <https://doi.org/10.1038/s41477-017-0007-7>.
- Leamy, L.J., Zhang, H., Li, C., Chen, C.Y. and Song, B.H.** (2017) A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genom.* **18**, 18. <https://doi.org/10.1186/s12864-016-3397-4>.
- Leem, S., Jeong, H.H., Lee, J., Wee, K. and Sohn, K.A.** (2014) Fast detection of high-order epistatic interactions in genome-wide association studies using information theoretic measure. *Comput. Biol. Chem.* **50**, 19–28. <https://doi.org/10.1016/j.combiolchem.2014.01.005>.
- Li, Z.K., Luo, L.J., Mei, H.W. et al.** (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics*, **158**, 1737–1753.
- Li, H., Peng, Z., Yang, X. et al.** (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **45**, 43–50. <https://doi.org/10.1038/ng.2484>.
- Li, W., Zhu, Z., Chern, M. et al.** (2017) A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. *Cell*, **170**, 114–126 e15. <https://doi.org/10.1016/j.cell.2017.06.008>.
- Li, X., Guo, T., Mu, Q., Li, X. and Yu, J.** (2018) Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *Proc. Natl. Acad. Sci. USA*, **115**, 6679–6684. <https://doi.org/10.1073/pnas.1718326115>.
- Lin, Z., Li, X., Shannon, L.M. et al.** (2012) Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **44**, 720–724. <https://doi.org/10.1038/ng.2281>.

- Lin, T., Zhu, G., Zhang, J. et al. (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226. <https://doi.org/10.1038/ng.3117>.
- Lishout, F.V., Gadaleta, F., Moore, J.H., Wehenkel, L. and Steen, K.V. (2015) gammaMAXT: a fast multiple-testing correction algorithm. *BioData Min.* **8**, 36. <https://doi.org/10.1186/s13040-015-0069-x>.
- Listgarten, J., Lipper, C. and Heckerman, D. (2013) FaST-LMM-Select for addressing confounding from spatial structure and rare variants. *Nat. Genet.* **45**, 470–471. <https://doi.org/10.1038/ng.2620>.
- Liu, J., Huang, J., Guo, H. et al. (2017) The conserved and unique genetic architecture of kernel size and weight in maize and rice. *Plant Physiol.* **175**, 774–785. <https://doi.org/10.1104/pp.17.00708>.
- Lu, L., Yan, W., Xue, W., Shao, D. and Xing, Y. (2012) Evolution and association analysis of Ghhd7 in rice. *PLoS One*, **7**, e34021. <https://doi.org/10.1371/journal.pone.0034021>.
- Lu, Y., Ye, X., Guo, R., Huang, J., Wang, W., Tang, J., Tan, L., Zhu, J.K., Chu, C. and Qian, Y. (2017) Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. *Mol. Plant*, **10**, 1242–1245. <https://doi.org/10.1016/j.molp.2017.06.007>.
- Luo, X., Ding, Y., Zhang, L., Yue, Y., Snyder, J.H., Ma, C. and Zhu, J. (2017) Genomic prediction of genotypic effects with epistasis and environment interactions for yield-related traits of rapeseed (*Brassica napus* L.). *Front. Genet.* **8**, 15. <https://doi.org/10.3389/fgene.2017.00015>.
- Ma, H., Li, G., Würschum, T., Zhang, Y., Zheng, D., Yang, X., Li, J., Liu, W., Yan, J. and Chen, S. (2018a) Genome-wide association study of haploid male fertility in maize (*Zea mays* L.). *Front. Plant Sci.* **9**, 974. <https://doi.org/10.3389/fpls.2018.00974>.
- Ma, Z., He, S., Wang, X. et al. (2018b) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **50**, 803–813. <https://doi.org/10.1038/s41588-018-0119-7>.
- Mackay, T.F. (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* **15**(1), 22–33. <https://doi.org/10.1038/nrg3627>.
- Maier, R.M., Zhu, Z., Lee, S.H. et al. (2018) Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat. Commun.* **9**, 989. <https://doi.org/10.1038/s41467-017-02769-6>.
- Manicacci, D., Camus-Kulandaivelu, L., Fourmann, M. et al. (2009) Epistatic interactions between Opaque2 transcriptional activator and its target gene CvPPDK1 control kernel trait variation in maize. *Plant Physiol.* **150**, 506–520. <https://doi.org/10.1104/pp.108.131888>.
- Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., Li, X. and Zhang, Q. (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. USA*, **107**, 19579–19584. <https://doi.org/10.1073/pnas.1014419107>.
- Mao, T., Li, J., Wen, Z. et al. (2017) Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. *BMC Genom.* **18**, 415. <https://doi.org/10.1186/s12864-017-3778-3>.
- Mathew, B., Leon, J., Sannemann, W. and Sillanpaa, M.J. (2018) Detection of epistasis for flowering time using Bayesian multilocus estimation in a barley MAGIC population. *Genetics*, **208**, 525–536. <https://doi.org/10.1534/genetics.117.300546>.
- Maurer, A., Draba, V., Jiang, Y., Schnaitmann, F., Sharma, R., Schumann, E., Kilian, B., Reif, J.C. and Pillen, K. (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genom.* **16**, 290. <https://doi.org/10.1186/s12864-015-1459-7>.
- Melchinger, A.E., Piepho, H.P., Utz, H.F., Muminovic, J., Wegenast, T., Törjek, O., Altmann, T. and Kusterer, B. (2007) Genetic basis of heterosis for growth-related traits in *Arabidopsis* investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics*, **177**, 1827–1837. <https://doi.org/10.1534/genetics.107.080564>.
- Mendez-Vigo, B., Savic, M., Ausin, I., Ramiro, M., Martin, B., Pico, F.X. and Alonso-Blanco, C. (2016) Environmental and genetic interactions reveal FLOWERING LOCUS C as a modulator of the natural variation for the plasticity of flowering in *Arabidopsis*. *Plant, Cell Environ.* **39**, 282–294. <https://doi.org/10.1111/pce.12608>.
- Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., Gao, C. and Li, J. (2017) Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant*, **10**(9), 1238–1241. <https://doi.org/10.1016/j.molp.2017.06.006>.
- Nicotra, A.B., Atkin, O.K., Bonser, S.P. et al. (2010) Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* **15**, 684–692. <https://doi.org/10.1016/j.tplants.2010.09.008>.
- Pandey, M.K., Upadhyaya, H.D., Rathore, A. et al. (2014) Genomewide association studies for 50 agronomic traits in peanut using the ‘reference set’ comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. *PLoS One*, **9**, e105228. <https://doi.org/10.1371/journal.pone.0105228>.
- Romero Navarro, J.A., Willcox, M., Burgueño, J. et al. (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**, 476–480. <https://doi.org/10.1038/ng.3784>.
- Santos, J.P., Pereira, H.D., Von Pinho, R.G., Pires, L.P., Camargos, R.B. and Balestre, M. (2015) Genome-wide prediction of maize single-cross performance, considering non-additive genetic effects. *Genet. Mol. Res.* **14**, 18471–18484. <https://doi.org/10.4238/2015.December.23.35>.
- Scheiner, S.M. (1993) Genetics and evolution of phenotypic plasticity. *Annu. Rev. Ecol. Syst.* **24**, 35–68. <https://doi.org/10.1146/annurev.es.24.110193.000343>.
- Schlücht, C.D. (1986) The evolution of phenotypic plasticity in plants. *Annu. Rev. Ecol. Syst.* **17**, 667–693. <https://doi.org/10.1146/annurev.es.17.110186.003315>.
- Schlücht, C.D. and Levin, D.A. (1984) Phenotypic plasticity of annual phlox: tests of some hypotheses. *Am. J. Bot.* **71**, 252–260. <https://doi.org/10.1002/j.1537-2197.1984.tb12511.x>.
- Schulthess, A.W., Reif, J.C., Ling, J. et al. (2017) The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum* L.). *J. Exp. Bot.* **68**, 4089–4101. <https://doi.org/10.1093/jxb/erx214>.
- Shang, Y., Ma, Y., Zhou, Y. et al. (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*, **346**, 1084–1088. <https://doi.org/10.1126/science.1259215>.
- Shen, G., Zhan, W., Chen, H. and Xing, Y. (2014) Dominance and epistasis are the main contributors to heterosis for plant height in rice. *Plant Sci.* **215–216**, 11–18. <https://doi.org/10.1016/j.plantsci.2013.10.004>.
- Singh, A.K., GanapathySubramanian, B., Sarkar, S. and Singh, A. (2018) Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* **23**, 883–898. <https://doi.org/10.1016/j.tplants.2018.07.004>.
- Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. and Smoller, J.W. (2013) Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14**, 483–495. <https://doi.org/10.1038/nrg3461>.
- Soyk, S., Lemmon, Z.H., Oved, M. et al. (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell*, **169**, 1142–1155.e12. <https://doi.org/10.1016/j.cell.2017.04.032>.
- Stearns, F.W. (2010) One hundred years of pleiotropy: a retrospective. *Genetics*, **186**, 767–773. <https://doi.org/10.1534/genetics.110.122549>.
- Studer, A.J. and Doebley, J.F. (2011) Do large effect QTL fractionate? A case study at the maize domestication QTL teosinte branched1. *Genetics*, **188**(3), 673–681. <https://doi.org/10.1534/genetics.111.126508>.
- Su, G., Madsen, P., Lund, M.S., Sorensen, D., Korsgaard, I.R. and Jensen, J. (2006) Bayesian analysis of the linear reaction norm model with unknown covariates. *J. Anim. Sci.* **84**, 1651–1657. <https://doi.org/10.2527/jas.2005-517>.
- Sun, L. and Wu, R. (2015) Mapping complex traits as a dynamic system. *Phys. Life Rev.* **13**, 155–185. <https://doi.org/10.1016/j.plrev.2015.02.007>.
- Tan, C., Weng, X.Y., Yan, W.H., Bai, X.F. and Xing, Y.Z. (2012) Ghhd7, a pleiotropic gene controlling flag leaf area in rice. *Yi Chuan*, **34**, 901–906.
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T. and Bennett, M. (2017) Plant phenomics, from sensors to knowledge. *Curr. Biol.* **27**, R770–R783. <https://doi.org/10.1016/j.cub.2017.05.055>.
- Thornberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S.T. (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**, 286–289. <https://doi.org/10.1038/90135>.
- Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rochefford, T.R., McMullen, M.D., Holland, J.B. and Buckler, E.S. (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**(2), 159–162. <https://doi.org/10.1038/ng.746>.
- Tieman, D., Zhu, G., Resende, M.F. et al. (2017) A chemical genetic roadmap to improved tomato flavor. *Science*, **355**, 391–394. <https://doi.org/10.1126/science.aal1556>.

- Ungerer, M.C., Halldorsdottir, S.S., Purugganan, M.D. and Mackay, T.F.** (2003) Genotype–environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics*, **165**(1), 353–365.
- Wang, Z., Pang, X., Lv, Y., Xu, F., Zhou, T., Li, X., Feng, S., Li, J., Li, Z. and Wu, R.** (2013) A dynamic framework for quantifying the genetic architecture of phenotypic plasticity. *Brief. Bioinform.* **14**, 82–95. <https://doi.org/10.1093/bib/bbs009>.
- Wang, J., Joshi, T., Valliyodan, B., Shi, H., Liang, Y., Nguyen, H.T., Zhang, J. and Xu, D.** (2015a) A Bayesian model for detection of high-order interactions among genetic variants in genome-wide association studies. *BMC Genom.* **16**, 1011. <https://doi.org/10.1186/s12864-015-2217-6>.
- Wang, Q., Xie, W., Xing, H. et al.** (2015b) Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. *Mol. Plant*, **8**, 946–957. <https://doi.org/10.1016/j.molp.2015.02.014>.
- Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X. and Qin, F.** (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* **48**, 1233–1241. <https://doi.org/10.1038/ng.3636>.
- Wang, C., Yang, Q., Wang, W., Li, Y., Guo, Y., Zhang, D., Ma, X., Song, W., Zhao, J. and Xu, M.** (2017a) A transposon-directed epigenetic change in ZmCCT underlies quantitative resistance to Gibberella stalk rot in maize. *New Phytol.* **215**, 1503–1515. <https://doi.org/10.1111/nph.14688>.
- Wang, M., Tu, L., Lin, M. et al.** (2017b) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587. <https://doi.org/10.1038/ng.3807>.
- Weaver, M.E. and Ingram, D.L.** (1969) Morphological changes in swine associated with environmental temperature. *Ecology*, **50**, 710–713. <https://doi.org/10.2307/1936264>.
- Wei, X., Liu, K., Zhang, Y. et al.** (2015) Genetic discovery for oil production and quality in sesame. *Nat. Commun.* **6**, 8609. <https://doi.org/10.1038/ncomms9609>.
- Wei, W.H., Massey, J., Worthington, J., Barton, A. and Warren, R.B.** (2018) Genotypic variability-based genome-wide association study identifies non-additive loci HLA-C and IL12B for psoriasis. *J. Hum. Genet.* **63**, 289–296. <https://doi.org/10.1038/s10038-017-0350-6>.
- Wen, W., Li, D., Li, X. et al.** (2014a) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**, 3438.
- Wen, Z., Tan, R., Yuan, J. et al.** (2014b) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genom.* **15**, 809. <https://doi.org/10.1186/1471-2164-15-809>.
- Wen, W., Liu, H., Zhou, Y. 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. <https://doi.org/10.1104/pp.15.01444>.
- Wen, Y.J., Zhang, Y.W., Zhang, J., Feng, J.Y., Dunwell, J.M. and Zhang, Y.M.** (2018) An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F2. *Brief. Bioinform.* [Epub ahead of print]. <https://doi.org/10.1093/bib/bby058>.
- Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J. and Zhang, Q.** (2014) Grain number, plant height, and heading date7 is a central regulator of growth, development, and stress response. *Plant Physiol.* **164**, 735–747. <https://doi.org/10.1104/pp.113.231308>.
- Würschum, T., Maurer, H.P., Kraft, T., Janssen, G., Nilsson, C. and Reif, J.C.** (2011) Genome-wide association mapping of agronomic traits in sugar beet. *Theor. Appl. Genet.* **123**, 1121–1131. <https://doi.org/10.1007/s00122-011-1653-1>.
- Xiao, Y., Tong, H., Yang, X. et al.** (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol.* **210**, 1095–1106. <https://doi.org/10.1111/nph.13814>.
- Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J.** (2017) Genome-wide association studies in maize: praise and stargaze. *Mol. Plant*, **10**(3), 359–374. <https://doi.org/10.1016/j.molp.2016.12.008>.
- Xing, A., Gao, Y., Ye, L. 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. <https://doi.org/10.1093/jxb/erv182>.
- Xue, W., Xing, Y., Weng, X. et al.** (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **40**, 761–767. <https://doi.org/10.1038/ng.143>.
- Yang, Q., Li, Z., Li, W. et al.** (2013a) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci.* **110**, 16969–16974. <https://doi.org/10.1073/pnas.1310949110>.
- Yang, W., Duan, L., Chen, G., Xiong, L. and Liu, Q.** (2013b) Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* **16**, 180–187. <https://doi.org/10.1016/j.pbi.2013.03.005>.
- Yano, K., Yamamoto, E., Aya, 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. <https://doi.org/10.1038/ng.3596>.
- Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q. and Saghai Maroof, M.A.** (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA*, **94**, 9226–9231. <https://doi.org/10.1073/pnas.94.17.9226>.
- Zhai, Y., Lv, Y., Li, X., Wu, W., Bo, W., Shen, D., Xu, F., Pang, X., Zheng, B. and Wu, R.** (2014) A synthetic framework for modeling the genetic basis of phenotypic plasticity and its costs. *New Phytol.* **201**, 357–365. <https://doi.org/10.1111/nph.12458>.
- Zhang, X., Wang, J., Huang, J. et al.** (2012) Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. USA*, **109**, 21534–21539. <https://doi.org/10.1073/pnas.1219776110>.
- Zhang, J., Singh, A., Mueller, D.S. and Singh, A.K.** (2015a) Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *Plant J.* **84**, 1124–1136. <https://doi.org/10.1111/tpj.13069>.
- Zhang, N., Gibon, Y., Jason, G.W. et al.** (2015b) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol.* **168**, 575–583. <https://doi.org/10.1104/pp.15.00025>.
- Zhang, Z., Mao, L., Chen, H. et al.** (2015c) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell*, **27**, 1595–1604. <https://doi.org/10.1105/tpc.114.135848>.
- Zhang, W., Dai, X., Wang, Q., Xu, S. and Zhao, P.X.** (2016) PEPIS: a pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies. *PLoS Comput. Biol.* **12**, e1004925. <https://doi.org/10.1371/journal.pcbi.1004925>.
- Zhang, L., Su, W., Tao, R. et al.** (2017a) RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat. Commun.* **8**, 2264. <https://doi.org/10.1038/s41467-017-02445-9>.
- Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X. and Yin, D.** (2017b) Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. Plant Sci.* **8**, 1611. <https://doi.org/10.3389/fpls.2017.01611>.
- Zhou, T., Lyu, Y., Xu, F., Bo, W., Zhai, Y., Zhang, J., Pang, X., Zheng, B. and Wu, R.** (2015) A QTL model to map the common genetic basis for correlative phenotypic plasticity. *Brief. Bioinform.* **16**, 24–31. <https://doi.org/10.1093/bib/bbt089>.
- Zhu, C., Li, X. and Yu, J.** (2011) Integrating rare-variant testing, function prediction, and gene network in composite resequencing-based genome-wide association studies (CR-GWAS). *G3 (Bethesda)*, **1**, 233–243. <https://doi.org/10.1534/g3.111.000364>.
- Zhu, Z., Zhang, F., Hu, H. et al.** (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487. <https://doi.org/10.1038/ng.3538>.
- Zhu, G., Wang, S., Huang, Z. et al.** (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e12. <https://doi.org/10.1016/j.cell.2017.12.019>.