GENETICS, 2025, 230(1), iyaf045
https://doi.org/10.1093/genetics/iyaf045
Advance Access Publication Date: 18 March 2025
Plant Genetics and Genomics

A theory of heterosis

Zhao-Bang Zeng [1,*], Gabriel De Siqueira Gesteira [1], Lujia Mo [1], Yingjie Xiao [2], Jianbing Yan [2]

[1]Bioinformatics Research Center, Department of Horticulture Science, North Carolina State University, Raleigh, 27695 NC, USA
[2]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

*Corresponding author: Bioinformatics Research Center, North Carolina State University, Raleigh, 27695-7566 NC, USA. Email: szeng@ncsu.edu

Heterosis refers to the superior performance of a hybrid over its parents. It is the basis for hybrid breeding particularly for maize and rice. Genetically, it is due to interactions between alleles of quantitative trait loci (dominance and epistasis). Despite enormous interest and efforts to study the genetic basis of heterosis, the relative contribution of dominance vs epistasis to heterosis is still not clear. This is because most published studies estimate quantitative trait loci effects in pieces, not able to put them together to assess the overall pattern adequately. We propose a theoretical framework that focuses on the inference of the relationship between genome and traits that includes the identification of multiple quantitative trait loci and estimation of the whole set of quantitative trait loci (additive, dominant, and epistatic) effects. Used for heterosis, it gives a clear genetic definition and interpretation of heterosis. We applied the theory and methods to a large maize dataset with a factorial design of many male and female inbred lines and their hybrid crosses. Heterosis of ear weight in maize is primarily due to quantitative trait loci dominant effects, many are overdominant. The contribution to heterosis due to epistasis is small and diffused. For comparison, we also analyzed a rice dataset that is an F2-type population derived from a cross between 2 inbred lines. The result indicates that dominance is still the main contributor to heterosis, and epistasis contribution is small.

Keywords: heterosis; quantitative genetic model; dominance; epistasis; model selection; hybrid breeding; Plant Genetics and Genomics

Introduction

Heterosis refers to the superior performance of a hybrid over its parents. The utilization of heterosis is the basis of hybrid breeding, particularly in maize (Duvick 2005; Kusmec *et al.* 2021) and in rice (Gu and Han 2024). The genetic basis of heterosis is due to the interaction of alleles of quantitative trait loci (QTL)—dominance and epistasis (e.g. Fujimoto *et al.* 2018). QTL dominance has long been regarded as a major contributor to heterosis because it was thought that dominance can mask the effects of deleterious recessive alleles. Overdominance is an excessive form of dominance. There have been many reports of the detection of QTL overdominance (Stuber *et al.* 1992; Krieger *et al.* 2010; Li *et al.* 2017), although cautions have been voiced that some may be due to pseudo-overdominance of multiple QTL in close repulsion linkage. Epistasis could also play a significant role in heterosis and there have been reports of detection of statistically significant epistasis between QTL alleles (e.g. Jiang *et al.* 2017). Despite tremendous interests and efforts, it is still largely unclear how to systematically study the genetics of heterosis and assess the relative importance of dominance (including overdominance) vs epistasis on heterosis. One problem could be due to the limited sample sizes of many study populations, which would limit our ability to assess QTL epistasis adequately. Another problem is that QTL effects including epistasis were typically estimated and explained in pieces in most studies (e.g. Huang *et al.* 2016; Xiao *et al.* 2021), which would impose the difficulty to assess the overall effects consistently and in totality.

In this study, we propose a general multiple QTL genetic model to model the relationship from genome to phenotypes in a population for general quantitative genetics data analysis and interpretation and use it to study heterosis. This approach could produce a direct estimate of the genetic composition of heterosis, thus providing evidence to assess the relative importance of dominance vs epistasis on heterosis and many genetic questions in the study population.

Theory and methods

Although this theory is currently proposed for the study of heterosis, the theoretical framework can be used for many applications and studies of quantitative trait variation in complex populations and environments, particularly for the evolutionary study of complex traits in general. This approach would work better for a large population with dense genetic markers.

Traditionally, the study of heterosis uses a cross (F1) of 2 inbred lines (P1 and P2) and from F1 to produce an F2 population or further crosses. This design studies the heterosis between an F1 and the mean of P1 and P2 and uses F2 or other segregating populations to detect and estimate QTL effects and relate those

Received on 05 September 2024; accepted on 09 March 2025

estimates to the observed heterosis. An example is the study of Hua et al. (2002, 2003) on rice heterosis.

Recently, Xiao et al. (2021) reported a study of maize heterosis, which has 6210 crosses (F1's) between 30 male inbred lines (P1's) from 1 heterotic group and 207 female inbred lines [P2's, drawn from 1,404 recombinant inbred lines (RILs) resulted from multiple-way crosses of 24 original lines] from another heterotic group, with extensive genomic genotypes and trait phenotypes. This is a North Carolina Design II (factorial design) population. In this population, each F1 hybrid has different parents and different heterosis. We will use these 2 studies to lay out our theory and discuss the genetic basis of heterosis in maize and rice. Of course, the theory can be used or adapted for other experimental designs and data structures. We first focus on the study of Xiao et al. (2021).

Population

Let $P1_i$ for $i = 1, 2, \ldots, n_1$ and $P2_j$ for $j = 1, 2, \ldots, n_2$ be inbred lines of the 2 heterotic groups and F_{ij} be their hybrids. We treat inbred lines and their hybrids as one population for model specifications, genetic estimation, and interpretation.

The G2A genetic model

Suppose we observe genotypes of many genomic single-nucleotide polymorphism (SNP) markers for $P1_i$ and $P2_j$ and hence F_{ij} (deduced from the marker genotypes of $P1_i$ and $P2_j$). If the marker coverage is dense, we can practically treat some markers as potential QTL and perform marker selection analysis. As SNP makers are biallelic, we will treat potential QTL as biallelic. If some markers of $P1_i$ and $P2_j$ are heterozygous, those markers can be accommodated in the analysis or treated as missing data.

We will use a general 2-allele (G2A) model Zeng et al. (2005) and Wang and Zeng (2006) to model and analyze QTL in both inbred lines and hybrids. A G2A model, rather than an F2 model, is more appropriate here to describe the genetic variation of the population, as allelic frequencies of markers and potential QTL are very uneven among the male and female inbred lines (the source of variation).

What is the G2A model? We first give the G2A model a general derivation and explanation. Let $s \in (P1_i, P2_j, F_{ij};\ i = 1, 2, \ldots, n_1,\ j = 1, 2, \ldots, n_2)$. We use z_{slg} to index the allelic state of individual s, locus l ($l = 1,2, \ldots, m$), and gamete g ($g = 1,2$). We consider 2 alleles A_{l1} and A_{l2} that are segregating in the population with allele frequency p_l for A_{l1}. Let

$$z_{slg} = \begin{cases} 1 & \text{if allele is } A_{l1} \\ 0 & \text{if allele is } A_{l2} \end{cases}$$

We use a centralized variable $x_{slg} = z_{slg} - Prob(z_{slg} = 1) = z_{slg} - p_l$ for model setting.

$$x_{slg} = z_{slg} - Prob(z_{slg} = 1) = \begin{cases} 1 - p_l & \text{if allele is } A_{l1} \\ -p_l & \text{if allele is } A_{l2} \end{cases}$$

Why centralizing variables? This is a well-known statistical practice. Centralizing or standardizing variables is especially important when a regression model contains interaction terms. If variables are not centralized or standardized when a model contains these types of terms, there is a risk of missing statistically significant results or producing potentially conflicting results, i.e. the model is internally inconsistent. The consistency means that a lower-dimension model is consistent in a higher-dimension space *under certain conditions* (Zeng et al. 2005).

With this specification, the relationship between a trait phenotype y_s and genotypes of multiple QTL (x'_{slg}s) can be modeled as

$$y_s = \mu + \sum_{l=1}^{m} a_l w_{sl} + \sum_{l=1}^{m} d_l v_{sl} + \sum_{k<l}^{m} aa_{kl} w_{sk} w_{sl} + \sum_{k\neq l}^{m} ad_{kl} w_{sk} v_{sl} + \sum_{k<l}^{m} dd_{kl} v_{sk} v_{sl} + \ldots + e_s \quad (1)$$

$$w_{sl} = x_{sl1} + x_{sl2} = \begin{cases} 2(1-p_l) & \text{for } A_{l1}A_{l1} \\ 1 - 2p_l & \text{for } A_{l1}A_{l2} \\ -2p_l & \text{for } A_{l2}A_{l2} \end{cases}$$

$$v_{sl} = -2x_{sl1}x_{sl2} = \begin{cases} -2(1-p_l)^2 & \text{for } A_{l1}A_{l1} \\ 2p_l(1-p_l) & \text{for } A_{l1}A_{l2} \\ -2p_l^2 & \text{for } A_{l2}A_{l2} \end{cases}$$

with $e_s \sim N(0, \sigma_e^2)$. In this model, w_{sl} is the additive allelic variable and a_l is the additive effect of QTL l, v_{sl} is the dominant (additive-by-additive interaction) variable, and d_l is the dominant effect of QTL l, aa_{kl}, ad_{kl}, and dd_{kl} are additive-by-additive, additive-by-dominant, and dominant-by-dominant epistatic effects between QTL k and l , respectively.

We used the G2A model form of Zeng et al. (2005) with $v_{sl} = -2x_{sl1}x_{sl2}$, differing on the specification of $v_{sl} = x_{sl1}x_{sl2}$ of Wang and Zeng (2006) by a factor −2. This is for the purpose to be in line with the specification of the F2 model (symmetric model) with $p_l = 1/2$. The F2 model was first introduced by Anderson and Kempthorne (1954) and has been used extensively in literature for many applications.

Why G2A model for general applications? This G2A model is a digital model or binary model that has simplicity in setting and can represent any complexity in multitude. Biologically, since many genomic studies contain dense SNP markers that are biallelic, it is reasonable to assume that some SNP markers are targeted QTL or very closely linked to casual variants, and a model selection from those SNP markers can have a good representation of the genetic structure of quantitative trait variation in a population. Theoretically, this model is akin to the infinite site mutation model (Crow and Kimura 1970). How about multiple alleles? The case of multiple alleles can be conveniently represented by multiple 2 alleles (SNP).

Incidentally, this G2A model can be readily extended to polyploids. Let ρ be ploidy number (2 for diploid, 4 for tetraploid, 6 for hexaploid, etc.). In the model above, we can extend the gametic index to ρ ($g = 1,2, \ldots, \rho$). Then, the allelic dosage additive variable can be extended to $w_{sl} = \sum_{g=1}^{\rho} x_{slg}$. The summation is the concept of allelic dosage for diploid and polyploids. Correspondingly, the pair dosage (2 allelic interaction) dominant variable $v_{sl} = \sum_{g<g'}^{\rho} x_{slg}x_{slg'}$, the triplet dosage (3 allelic interaction) dominant variable $u_{sl} = \sum_{g<g'<g''}^{\rho} x_{slg}x_{slg'}x_{slg''}$, etc. With these specifications, a polyploid G2A model can be expressed in a similar form of equation (1) for all levels of allelic effects and interactions within and between loci. A more general discussion on the implications and applications of this polyploid G2A model will be presented elsewhere.

Heterosis

Heterosis of a trait for F_{ij} from $P1_i$ and $P2_j$ is defined as

$$H_{ij} = y_{F_{ij}} - (y_{P1_i} + y_{P2_j})/2 \quad (2)$$

To study the genetic basis of heterosis, we need to show what constitutes heterosis in genetic terms. Denote $z_{P1_{i,l}} = A_{l1}A_{l1}$ or $A_{l2}A_{l2}$ for $l = 1, 2, \ldots, m$ and similarly for $z_{P2_{j,l}}$ as well for marker genotypes of inbred lines. Then

$$Z_{F_{ij,l}} = \begin{cases} A_{l1}A_{l1} \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ A_{l1}A_{l2} \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l2}A_{l2} \\ A_{l2}A_{l1} \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ A_{l2}A_{l2} \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} and\ z_{P2_{j,l}} = A_{l2}A_{l2} \end{cases}$$

By using the above genetic model, the heterosis is expected to be as follows if we restrict the analysis to additive, dominant, and additive-by-additive epistatic effects.

$$\begin{aligned} E(H_{ij}) = E(y_{F_{ij}}) - E(y_{P1_i} + y_{P2_j})/2 = \sum_{l=1}^{m} (a_l \tilde{w}_{ij,l} + d_l \tilde{v}_{ij,l}) \\ + \sum_{k<l}^{m} (aa_{kl} w_{ij,\widetilde{k}} \widetilde{w}_{ij,l}) \end{aligned} \quad (3)$$

with $\tilde{w}_{ij,l} = w_{F_{ij,l}} - (w_{P1_{i,l}} + w_{P2_{j,l}})/2$

$$\tilde{v}_{ij,k} = v_{F_{ij,l}} - (v_{P1_{i,l}} + v_{P2_{j,l}})/2$$

$$w_{ij,\widetilde{k}} \widetilde{w}_{ij,l} = w_{F_{ij,k}} w_{F_{ij,l}} - (w_{P1_{i,k}} w_{P1_{i,l}} + w_{P2_{j,k}} w_{P2_{j,l}})/2$$

For a QTL locus if both parents have the same homozygote (either both $A_{l1}A_{l1}$ or $A_{l2}A_{l2}$), the hybrid genotype is still $A_{l1}A_{l1}$ or $A_{l2}A_{l2}$; $\tilde{w}_{ij,l} = \tilde{v}_{ij,l} = 0$. When parental genotypes have different homozygotes (i.e. one is $A_{l1}A_{l1}$ and the other is $A_{l2}A_{l2}$), the hybrid is heterozygote ($A_{l1}A_{l2}$) and

$$\tilde{w}_{ij,l} = (1 - 2p_l) - \frac{2(1-p_l) - 2p_l}{2} = 0$$

$$\tilde{v}_{ij,l} = 2p_l(1-p_l) - \frac{-2(1-p_l)^2 - 2p_l^2}{2} = 1$$

Thus,

$$\tilde{w}_{ij,l} = \begin{cases} 0 \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ 0 \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l2}A_{l2} \\ 0 \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ 0 \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} \text{ and } z_{P2_{j,l}} = A_{l2}A_{l2} \end{cases}$$

$$\tilde{v}_{ij,l} = \begin{cases} 0 \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ 1 \text{ if } z_{P1_{i,l}} = A_{l1}A_{l1} \text{ and } z_{P2_{j,l}} = A_{l2}A_{l2} \\ 1 \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} \text{ and } z_{P2_{j,l}} = A_{l1}A_{l1} \\ 0 \text{ if } z_{P1_{i,l}} = A_{l2}A_{l2} \text{ and } z_{P2_{j,l}} = A_{l2}A_{l2} \end{cases}$$

This shows that a does not contribute to heterosis and d contributes to heterosis for loci in heterozygote in hybrid.

For a pair of QTL (k and l), if $z_{P1_{i,kl}} = A_{k1}A_{k1}, A_{l1}A_{l1}\}$ and $z_{P2_{j,kl}} = A_{k2}A_{k2}, A_{l2}A_{l2}\}$ or $z_{P1_{i,kl}} = A_{k2}A_{k2}, A_{l2}A_{l2}\}$ and $z_{P2_{j,kl}} = A_{k1}A_{k1}, A_{l1}A_{l1}\}$

$$w_{ij,\widetilde{k}} \widetilde{w}_{ij,l} = (1-2p_k)(1-2p_l) - \frac{[2(1-p_k)2(1-2p_l) + (-2p_k)(-2p_l)]}{2} = -1$$

If $z_{P1_{i,kl}} = A_{k1}A_{k1}, A_{l2}A_{l2}\}$ and $z_{P2_{j,kl}} = A_{k2}A_{k2}, A_{l1}A_{l1}\}$ or $z_{P1_{i,kl}} = A_{k2}A_{k2}, A_{l1}A_{l1}\}$ and $z_{P2_{j,kl}} = A_{k1}A_{k1}, A_{l2}A_{l2}\}$

$$w_{ij,\widetilde{k}} \widetilde{w}_{ij,l} = (1-2p_k)(1-2p_l) - \frac{[2(1-p_k)(-2p_l) + (-2p_k)2(1-p_l)]}{2} = 1$$

For all other cases, $w_{ij,\widetilde{k}} \widetilde{w}_{ij,l} = 0$. Thus,

$$w_{ij,\widetilde{k}} \widetilde{w}_{ij,l} = \begin{cases} -1 \text{ if } z_{P1_{i,kl}} = \{A_{k1}A_{k1}, A_{l1}A_{l1}\} and\ z_{P2_{j,kl}} = \{A_{k2}A_{k2}, A_{l2}A_{l2}\} \\ 1 \text{ if } z_{P1_{i,kl}} = \{A_{k1}A_{k1}, A_{l2}A_{l2}\} and\ z_{P2_{j,kl}} = \{A_{k2}A_{k2}, A_{l1}A_{l1}\} \\ 1 \text{ if } z_{P1_{i,kl}} = \{A_{k2}A_{k2}, A_{l1}A_{l1}\} and\ z_{P2_{j,kl}} = \{A_{k1}A_{k1}, A_{l2}A_{l2}\} \\ -1 \text{ if } z_{P1_{i,kl}} = \{A_{k2}A_{k2}, A_{l2}A_{l2}\} and\ z_{P2_{j,kl}} = \{A_{k1}A_{k1}, A_{l1}A_{l1}\} \\ 0 \quad others \end{cases}$$

This shows that aa contributes to heterosis if both parents have different homozygotes on the 2 loci.

Thus, given estimates of genetic model parameters ($\hat{d}_k$ and $\widehat{aa}_{kl}$), we can estimate genetic partitions and components of heterosis for each hybrid.

$$\hat{H}_{ij} = \hat{D}_{ij} + \widehat{AA}_{ij} = \sum_{l=1}^{m} (\hat{d}_l \tilde{v}_{ij,l}) + \sum_{k<l}^{m} (\widehat{aa}_{kl} w_{ij,\widetilde{k}} \widetilde{w}_{ij,l}) \quad (4)$$

This can be compared with the observed heterosis H_{ij}.

Here, we ignored additive-by-dominant $\sum_{k\neq l}^{m} (\widehat{ad}_{kl} w_{ij,\widetilde{k}} \widetilde{v}_{ij,l})$, dominant-by-dominant $\sum_{k<l}^{m} (\widehat{dd}_{kl} v_{ij,\widetilde{k}} \widetilde{v}_{ij,l})$, and higher-order epistatic effects because they are higher-order statistics and less important. The terms $w_{ij,\widetilde{k}} \widetilde{v}_{ij,l}$ and $v_{ij,\widetilde{k}} \widetilde{v}_{ij,l}$ are complex and nonzero in general but are zero when allele frequency $p_k = p_l = 1/2$. That is, for the F2 model with allele frequency half, additive-by-dominant and dominant-by-dominant epistasis do not contribute to heterosis (Melchinger *et al.* 2007; Garcia *et al.* 2008). If deemed necessary, we could include additive-by-dominant and dominant-by-dominant epistasis in the analysis.

Note that, as pointed out in Garcia *et al.* (2008), the interpretation of heterosis genetics depends on the genetic model used. A commonly used genetic model in quantitative genetics literature is the F_∞ genetic model largely because of its simplicity in expression. Based on the F_∞ model, heterosis also depends on dominant-by-dominant interaction. In this study, we used a G2A model and its special form F_2 model for the reason explained in Zeng *et al.* (2005). It is based on the principle of partition of genetic variances, the legacy of Fisher (1918).

Now, we summarize and highlight a few points of our model-based analysis approach. First, we put inbred lines and hybrids together as one population and model the genetic variation for the whole population. As such, a model selection and estimation can explain the genetic variation within and between inbred lines and hybrids. The heterosis is defined specifically for each pair of inbred lines and their hybrid and is shown to be due to dominant effects of QTL that are heterozygote in the hybrid and aa epistatic effects of QTL that have opposite homozygote genotypes of the 2 loci in the inbred lines. In this model system, we try to identify individual QTL and analyze additive effects of QTL alleles and significant pair-wise interaction effects of QTL alleles (dominant effects within loci and aa epistatic effects between loci). Heterosis is all about the interactions of QTL alleles, primarily the pair-wise allelic interaction effects of QTL (dominance within loci and aa epistasis between loci).

Data and analysis

We applied this model to the dataset of Xiao *et al.* (2021). The data consist of the high-quality whole genome SNP marker (~4.5 million) genotypes of 1,428 inbred lines from the CUBIC (complete-diallel plus unbalanced breeding-derived inter-cross) synthetic population as a maternal pool and 30 paternal tester lines from

diverse genetic backgrounds. We will focus our analysis and discussion on a cross population that includes 207 maternal lines (randomly selected from the 1,428 CUBIC lines) and 30 paternal lines and their 6,210 hybrids. Twenty quantitative traits were measured in 5 locations for all inbred lines and hybrids. Our analysis is primarily focused on the inference of genetic structure of the population on each quantitative trait, and through the inference to study the genetic basis of heterosis.

Model selection

Model selection is at the core of a model-based analysis. The results will depend on the selection procedures and criteria. We used the LASSO method and combined it with stability selection for model selection. LASSO (Tibshirani 1996) is a statistical method that shrinks regression coefficient estimates through the L1 regularization, leading many small estimates to zero to achieve a subset selection. Stability selection (Meinshausen and Bühlmann 2010) uses LASSO model selection in subsampling data to explore the model structure. It provides an algorithm for selecting a model while controlling the number of false discoveries.

Specifically, this is the procedure we used for data analysis. First, there are too many markers in the data for marker subset selection analysis. To achieve a feasible and efficient computation, we initially generated an evenly spaced marker subset by sampling one marker every 800 markers across the whole genomes of 10 chromosomes. After removing markers with minor allele frequency (MAF) less than $1/(30+24) = 0.0185$ and with genotype errors and contamination, a total of 4,701 markers were retained in the marker pool for subsequent model selection. We used stability selection (Meinshausen and Bühlmann 2010) to enhance the consistency and robustness of QTL selection while controlling false discoveries. The procedure involved 100 times of random subsampling from the total 6,447 samples, with each round drawing 50% of the samples without replacement. Marker selection for each of the 100 sample sets was conducted using group LASSO (Simon *et al.* 2013) via the R package "grplasso" (Meier *et al.* 2008) along with 10-fold cross-validation, in which the additive and dominance effects of a QTL were selected together as a pair. Given the tunning parameter λ with the minimum cross-validation error, QTL consistently selected in at least 50 of the 100 subsamples were retained for further analysis. Conditional on the selected additive and dominant QTL effects, the same selection procedure was used again to identify additive-by-additive epistatic interactions among all combinations of those selected QTL. The QTL (a, d, and aa) effects in the final selected model were reestimated in the full sample.

Thus, we used a 2-step selection procedure, first selecting QTL main effects (putting a and d together in selection) from candidate markers, then selecting QTL aa effects only from the QTL pairs selected in the first step. This procedure was in line with a similar procedure used in Laurie *et al* (2014), which explained the justification and rational for the multiple step selection procedure.

During the investigation, we selected and compared many different models. A detailed discussion on model comparison is complex. Although some model details may vary for different model selection procedures, the genetic result pattern and conclusions reported below are robust. To simplify the result report and discussion, we report the results based on a representative selected model.

Results and discussion

Given an inferred genetic model for a study population, one can explore the genetic structure of the population for a quantitative trait, estimate or predict any quantity including its genetic components, trace the causes (QTL changes) and process (a sequence of changes) of evolutionary events or selection responses to breeding efforts, and utilize the inferred information for a more proactive or creative intervention (e.g. breeding design). These are just a few examples of advantages and opportunities that a model-based quantitative genetic inference can and should play for an agricultural breeding program or any biological inquiry on complex traits.

Selected QTL model for ear weight and heterosis composition

Since ear weight is the most important trait economically and has the most significant heterosis, we will discuss the model selection results mostly using ear weight as an example and then discuss other traits for comparisons. Based on the selection procedure, we selected a model of 139 QTL distributed all over the genome with additive (a) and dominant (d) effects and selected 413 additive-by-additive (aa) epistatic effects (Supplementary Table 1).

Figure 1 plots some details of this inferred genetic model of ear weight: genotypes (A_1A_1 or A_2A_2), estimated additive (a) and dominant (d) effects (red and blue dots) of 139 QTL for 30 male lines, 24 female founder lines, and 183 derived female lines. This figure provides a visual picture of the genetic differences between the 2 heterotic groups (male and female lines) on ear weight. It shows what contributes to the heterosis and how one might design a better breeding plan to further improve the breeding lines.

More genetic details are plotted in Fig. 2. Figure 2a plots the distribution of 139 QTL with additive (a) and dominant (d) effects and 413 additive-by-additive (aa) epistatic effects on ear weight. A distinct feature emerges that a and aa effects are centered around zero, about equally positive or negative, and d effects are predominantly positive. This goes right in the heart of the genetic interpretation of the causes of heterosis. As seen clearly in Fig. 2e which shows the estimates of heterosis components [dominance vs aa epistasis, equation (4)], the conclusion is clear that the heterosis in ear weight is primarily due to QTL dominant effects. The contribution due to epistasis is small in magnitude and non-directional. Further, by Fig. 2b which plots dominant (d) effects of QTL against the degree of dominance ($d/|a|$), QTL effects on heterosis on ear weight are overwhelmingly overdominant. It is possible that some of those overdominant QTL could be pseudo-overdominance due to multiple underlying genes in close repulsion linkage. But the evidence on overdominance is overwhelming.

Figure 2c and d shows the model fit, the comparison of estimated vs observed ear weights for inbred lines and hybrids, and their differences (heterosis). More details are provided in Table 1 on the partition of variances and covariances of genetic components (A, D, and AA) and residuals in male inbred lines, female inbred lines, hybrids, and heterosis (the difference between hybrid and mean of inbred parental lines) and the broad-sense heritability. For male inbred lines, female inbred lines, and hybrids, the intercept (μ), A, D, AA, and residual are specified by equation (1), and for heterosis, D, AA, and residual are specified by equations (2) and (4). Figure 2c and d and Table 1 show the fit of the genetic model to the population on ear weight in different ways.

The model fit for female parents and hybrids is better than that for male parents. This is partly because there are more female parents (207) and hybrids (6210) than male parents (30) that differentially contributed to the overall model fit, but probably more because the female parents have their genomes randomized during

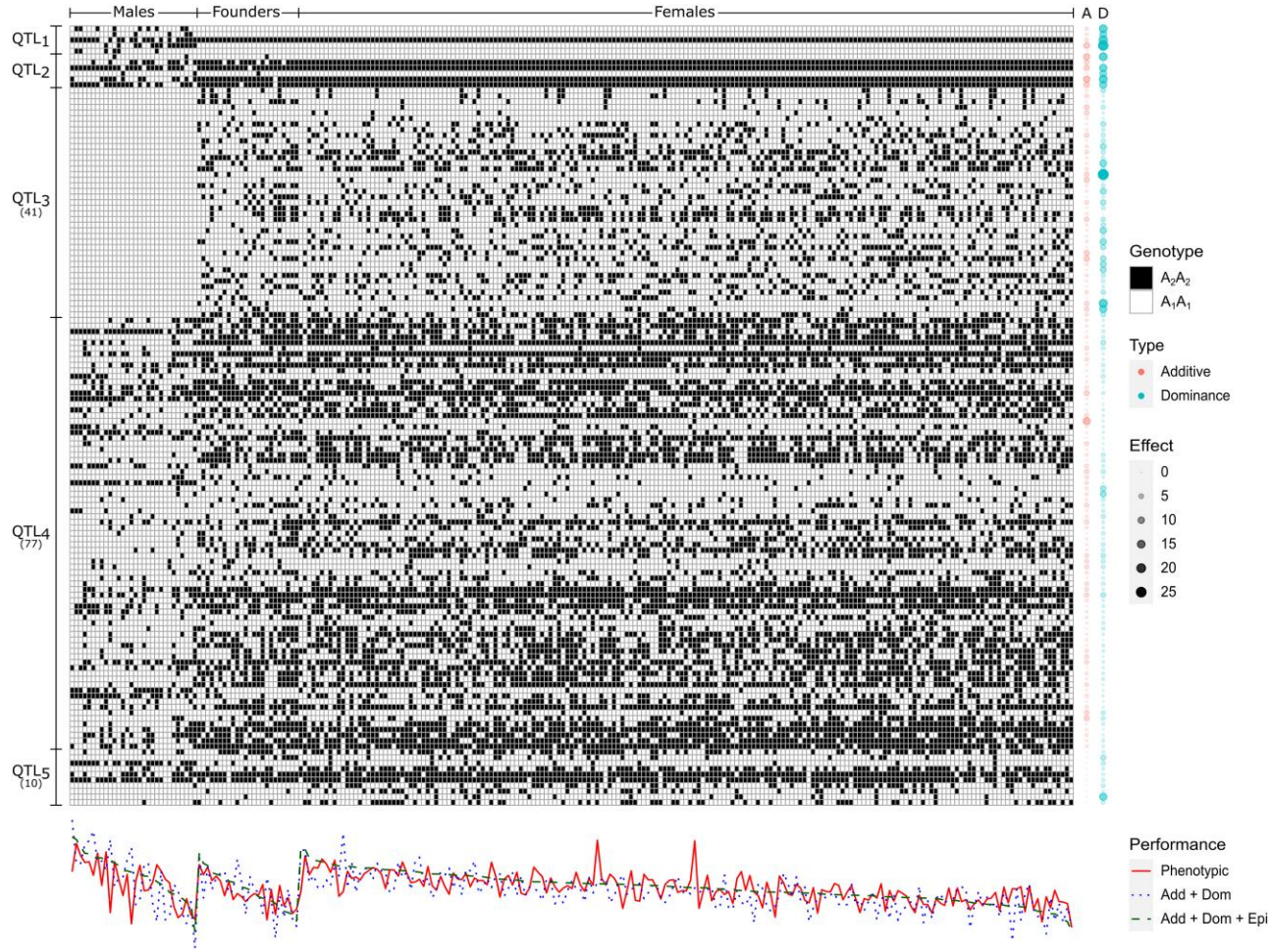

Fig. 1. The inferred genetic model of ear weight: genotypes (A_1A_1 or A_2A_2), estimated additive (A) and dominant effects (D) (dots grouped in absolute size) of 139 QTL for 30 male lines, 24 female founder lines, and 183 derived female lines arranged in columns. QTLs are arranged in rows and grouped to show the pattern of the genetic structure in the parental populations. Group 1: 5 QTL that segregate only in males; Group 2: 6 QTL that segregate only in males and founders; Group 3: 41 QTL that segregate in founders and derived females, but not in males; Group 4: 77 QTL that segregate in males, founders and derived females; and Group 5: 10 QTL with additive effects close to zero, but significant dominant effects. The lower panel shows observed phenotypic values of ear weight, estimated ear weights with QTL additive and dominant effects, and with QTL additive, dominant, and additive-by-additive effects.

the construction of CUBIC lines and the male parents are genetically more diverse. It is notable that hybrids have smaller residual variances (Table 1) than their parents, which is generally observed in hybrid experiments.

Degree of heterosis

In this study, each hybrid has different parents and thus different heterosis. The degree of heterosis varies from hybrid to hybrid and from trait to trait. We define the degree of heterosis (DH) by

$$\mathrm{DH}_{ij} = \left[y_{F_{ij}} - \frac{y_{P1_i} + y_{P2_j}}{2}\right] \Big/ \left[\frac{y_{P1_i} + y_{P2_j}}{2}\right]$$

which is plotted in Fig. 3a. There is a big difference in DH between yield-related traits, such as EW (ear weight) and KWPE (kernel weight per ear), and other traits. This reflects the fact that the selection for heterosis has been primarily concentrated on yield in the past hybrid breeding in maize and resulted in a profound degree of heterosis. Genetically, the difference in DH between traits is related to the mean degree of dominance MDD = $\overline{d_i}/\overline{|a_i|}$ of QTL. Figure 3b plots the mean degree of heterosis (MDH) against the mean degree of dominance (MDD). This means that selection on heterosis can increase the degree of dominance on QTL.

Selection target

The distribution of estimated QTL effects in Fig. 2a immediately suggests a measure that measures the direction of QTL effects, which may be called selection target as it tends to reflect the impact of selection: $T_a^{\mathrm{EW}} = \sum a_i / \sum |a_i| = 0.06$, $T_d^{\mathrm{EW}} = \sum d_i / \sum |d_i| = 0.94$, and $T_{aa}^{\mathrm{EW}} = \sum aa_{ij} / \sum |aa_{ij}| = -0.06$. Figure 3c shows the measures for all traits in the study.

This is of course due to hybrid breeding that has been practiced for over a century (Duvick 2005; Kusmec *et al.* 2021). Hybrid breeding explored heterosis between different heterotic groups by using reciprocal recurrent selection (RRS) that is targeted to preserve and enhance heterosis. As a result, hybrid performance and heterosis have been consistently enhanced over time, although surely inbred lines have also been continuously improved particularly with significant efforts to eliminate deleterious recessive alleles that hinder inbreeding efforts. Thus, the primary selection targets are QTL dominant effects. Male and female inbred line performance is similar for ear weight

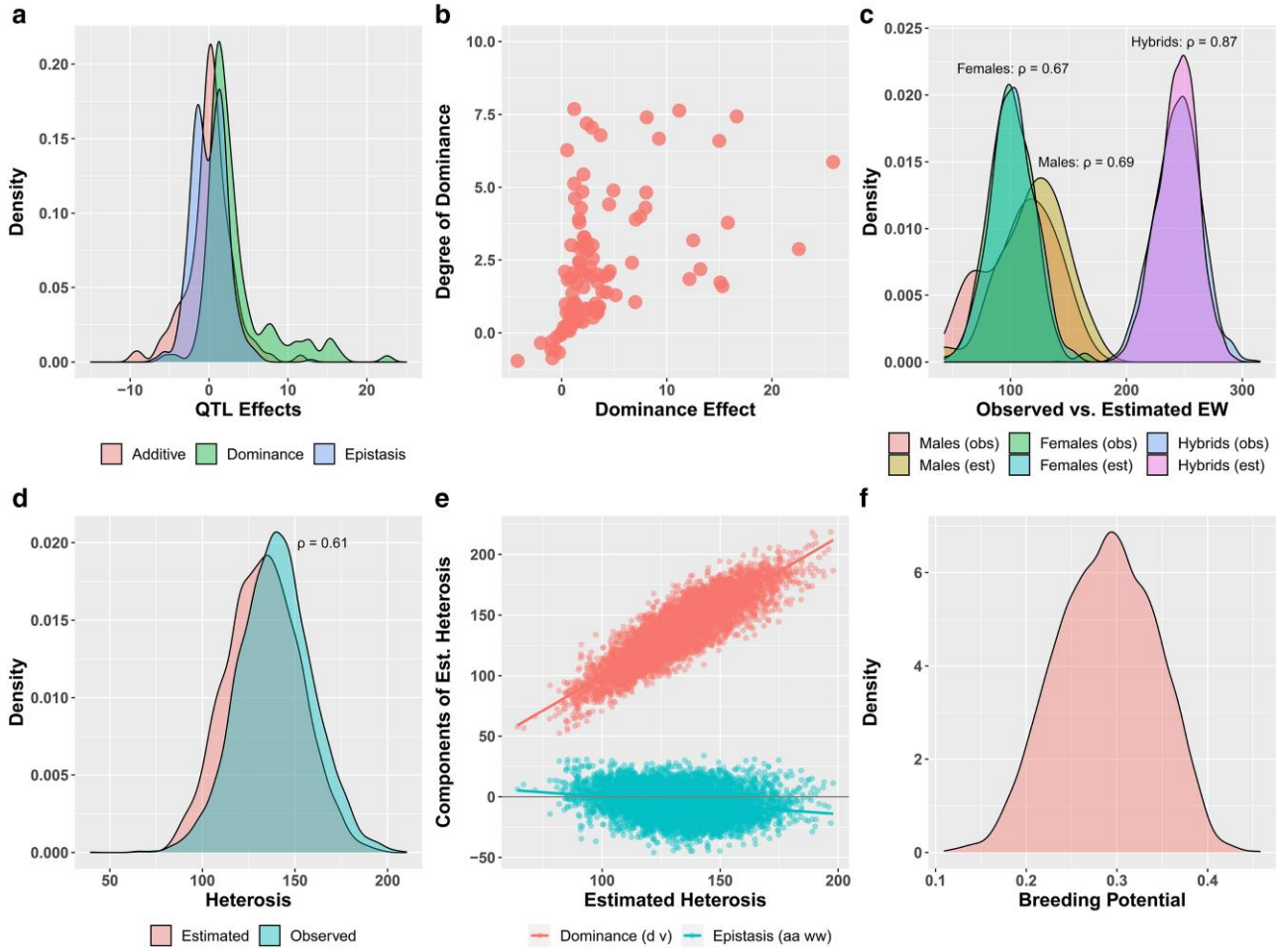

Fig. 2. Estimation of genetic model and associated properties on ear weight in the population. a) The distribution of 139 QTL with additive (a) and dominant (d) effects and 413 additive-by-additive (aa) epistatic effects on ear weight. b) Dominant (d) effects of QTL are plotted against the degree of dominance ($d/|a|$). c) The comparison of estimated vs observed ear weights for (male and female) inbred lines and hybrids. d) The comparison of estimated vs observed ear weights for heterosis. e) The distribution of the components [dominance vs aa epistasis, equation (4)] of estimated heterosis for 6,210 hybrids is plotted against the estimated heterosis. f) This plots what may be called the breeding potential of 6,210 hybrid crosses: the dominant contribution $\sum_l (\hat{d}_l v_{ij,l})$ of each hybrid to heterosis as a ratio of the maximal contribution of QTL dominant effects to heterosis $\sum_l (\hat{d}_l)$, if all segregating QTL alleles are in heterozygote in a hybrid, i.e $v_{ij,l} = 1$ for all l.

Table 1. Partition of variances and covariances of genetic components (A, D, and AA) and residuals in male inbred lines, female inbred lines, hybrids, and heterosis (difference between hybrid and mean of male and female parental lines) and the broad-sense heritability.

			Partition of variance			
	Intercept					Heritability
Intercept	218.86	A	D	AA	Residuals	H^2
Male	A	355.60				
	D	273.78	804.59			
	AA	−261.72	−522.73	635.98		
	Residuals				519.33	0.71
Female	A	448.20				
	D	−215.04	486.45			
	AA	−103.86	−101.73	268.73		
	Residuals				237.06	0.77
Hybrid	A	195.45				
	D	−56.10	181.60			
	AA	−20.50	11.67	59.70		
	Residuals				97.95	0.79
Heterosis	D		674.70			
	AA		−222.35	165.45		
	Residuals				308.44	0.67

and other traits; thus, QTL additive effects are not directional. Additive-by-additive epistatic effects are also not directional for the population. This, of course, agrees with the conclusion that the QTL additive-by-additive epistatic effects are not the main cause and contributor to the maize heterosis as shown in Fig. 2e.

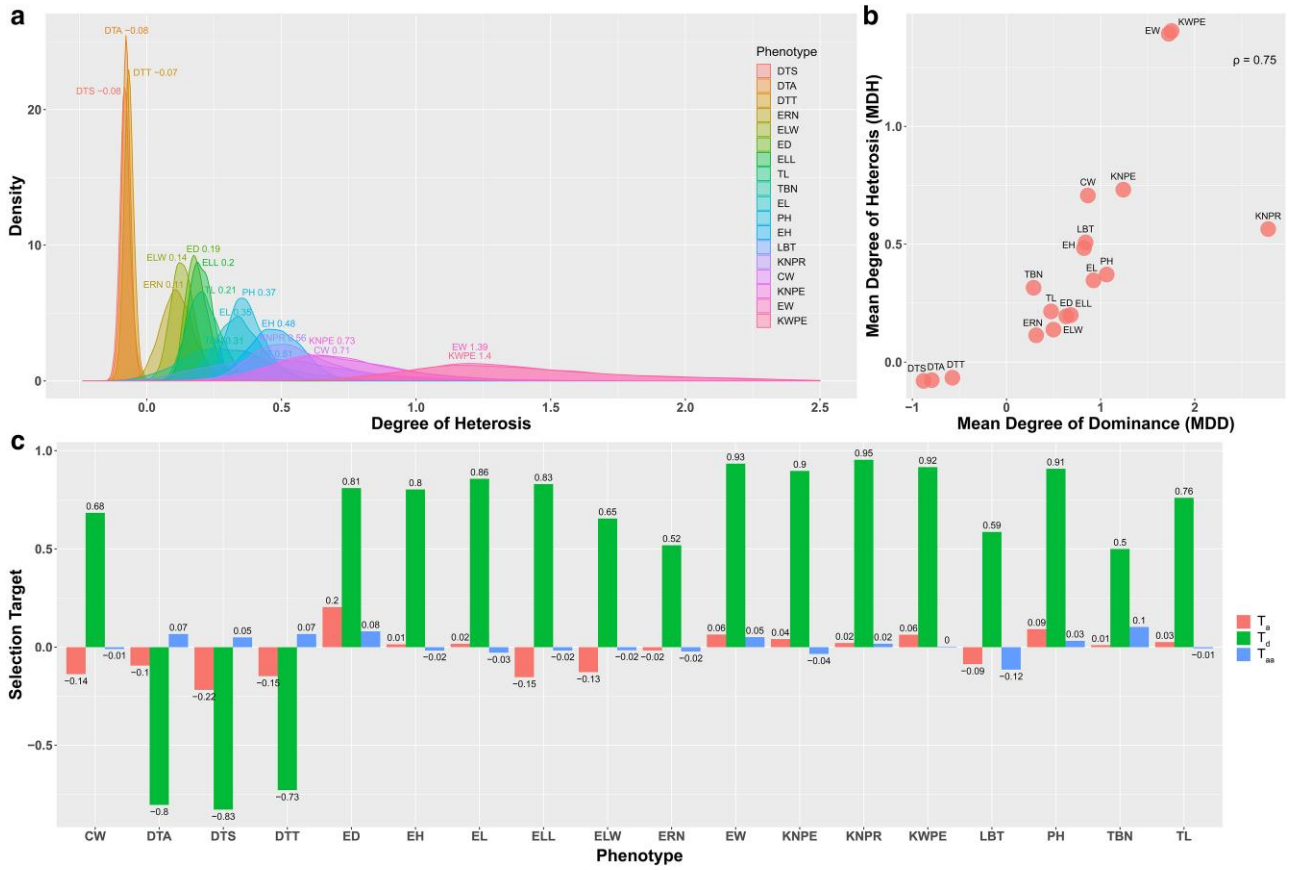

Fig. 3. a) The degree of heterosis of 6,210 hybrids is plotted for the 18 traits measured in the experiment. b) The mean degree of QTL dominant effects inferred for the 18 traits are plotted against the MDH for the 18 traits. c) Selection target of QTL effects (T_a, T_d, and T_{aa}) inferred for the 18 traits. Name of the traits: CW (cob weight), DTA (days to anthesis), DTS (days to silking), DTT (days to tasseling), ED (ear diameter), EH (ear height), EL (ear length), ELL (ear leaf length), ELW (ear leaf width), ERN (ear row number), EW, KNPE (kernel number per ear), KNPR (kernel number per row), KWPE, LBT (length of the barren tip), PH (plant height), TBN (tassel branch number), and TL (tassel length).

This QTL effect-based measure is very useful to detect and explain the cause and/or consequence of evolutionary process. In a large-scale QTL mapping study between *Drosophila simulans* and *Drosophila mauritiana* on male genital arch (MGA), 19 QTL were detected with estimates of QTL effects in 2 backcrosses (Zeng *et al.* 2000). From Table 2 of Zeng *et al.* (2000), we can take the averages of QTL effects in 2 backcrosses as the estimates of QTL additive effects. This gives $T_a^{MGA} = 0.98$ with 18 of the 19 QTL having additive effects in one direction. We can also take the differences of QTL effects in 2 backcrosses as the estimates of QTL dominant effects. This gives $T_d^{MGA} = -0.53$. This gives strong evidence that the MGA may have been strongly selected in opposite directions during and since speciation due to differential female preferences in the 2 species (True *et al.* 1997) and the impact of directional selection is primarily reflected on QTL additive effects.

This conclusion is mirrored in a large-scale directional artificial selection experiment in *Drosophila melanogaster* on wing shape (WS) (Weber 1990, 1992). After intensive bidirectional selection on WS for 16 generations, a high line and a low line were created and then crossed to create segregating RILs for QTL analysis. For chromosome 3, 11 QTL were detected with 10 QTL additive effects in one direction [Table 5 of Weber *et al.* (1999)]. This gives $T_{a.ch3}^{WS} = 0.98$. QTL analysis also detected 9 additive-by-additive epistatic effects with $T_{aa.ch3}^{WS} = -0.20$. For chromosome 2, this QTL pattern was also repeated with 10 QTL detected and all additive effects in one direction, thus $T_{a.ch2}^{WS} = 1$ and 14 additive-by-additive epistatic effects detected with $T_{aa.ch2}^{WS} = 0.44$ [from Table 3 of Weber *et al.* (2001)].

Table 2. Observed and estimated heterosis and parental line differences for the 4 traits in the rice heterosis study.

		Yield	Tillers	Grains	Grain weight
Heterosis	Observed	28.9	3.4	55.05	2.48
	Estimated	33.35	8.18	49.12	1.26
Parental	Observed	−23.03	−3	−43.5	−5.95
Differences	Estimated	−10.33	−1.77	−25.62	−3.68
$\sum d$		32.55	7.78	23.45	0.96
$\sum aa$		−0.79	−0.4	−25.67	−0.3
% of $\sum d$		0.98	0.95	0.48	0.76
% of $\sum aa$		0.02	0.05	0.52	0.24
Mean degree of dominance		2.86	2.54	1.69	0.56

The estimated heterosis includes the components due to QTL dominance and epistasis and their respective percentages. Estimated mean degree of dominance MDD = $\overline{d_i}/\overline{|a_i|}$ of QTL for the 4 traits.

When we put all these together, we can conclude that a QTL model-based analysis can reveal a selection target that reflects the action and consequence of the selection process. For directional selection, it is reflected in QTL additive effects, whether it is natural directional selection as in the case of *D. simulans* and *D. mauritiana* on MGA or artificial directional selection as in the case of *D. melanogaster* on WS. For hybrid breeding, it is reflected in QTL dominant effects.

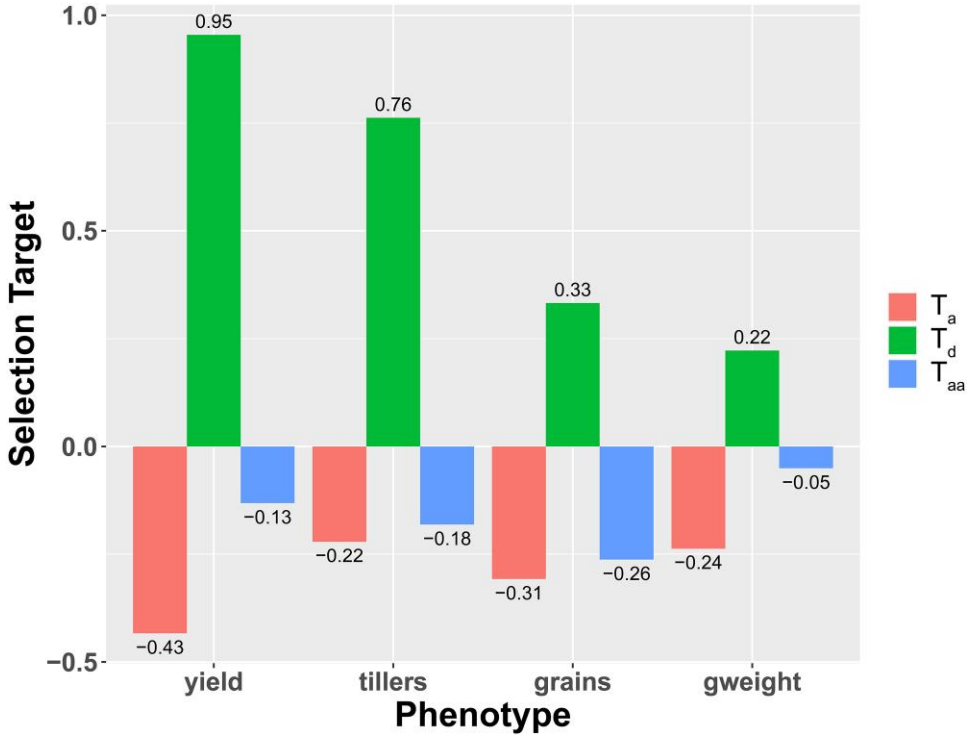

Fig. 4. Selection target of QTL effects (T_a, T_d, and T_{aa}) for the rice study.

Breeding potential

Since QTL dominant effects are the primary contributor to heterosis, it is of interest to examine the dominant contribution $\sum_{l=1}^{m}(\hat{d}_k v_{ij,l})$ of each hybrid to heterosis as a ratio of the maximal contribution of QTL dominant effects to heterosis $\sum_{l}(\hat{d}_l)$, if all segregating QTL alleles are in heterozygote in a hybrid, i.e. $\tilde{v}_{ij,l}=1$ for all l. This is plotted in Fig. 2e. This has important implications on experimental design for further breeding to explore heterosis.

If we would use this panel of inbred lines as a breeding population, based on the inferred genetic models for different traits we can explore computational means to optimize breeding experimental designs. In maize breeding, the selection scheme is RRS, i.e. repeated crossing within each heterotic group to create new RILs and testing with candidates of the opposite heterotic group for improving breeding gains. This is aimed to preserve and enhance the heterosis while improving inbred line performance as well. Given a comprehensive estimation of QTL positions (τ_k, hence r_{kl} estimated recombination frequencies between QTL), genotypes ($z_{P1_{i,k}}$, $z_{P2_{j,k}}$), and effects (a_k, d_k, aa_{kl}) for all inbred lines and hybrids for the targeted traits, we can compute the consequence (in probability) of a cross between any 2 lines within heterotic groups for line development (genome reshuffling to develop new RILs) or between heterotic groups for choosing parents (testing hybrid). Hence, this would be a very efficient computational searching strategy for the selection of mating pairs, a big advantage. This can be combined in a targeted simulation study to explore the short- and long-term breeding strategy in this population. This topic will be explored elsewhere.

Comparison of heterosis in maize and rice

Hua et al. (2002, 2003) reported a rice heterosis study. Starting with a cross between Zhenshan 97 and Minghui 63, the 2 parental inbred lines that produced F_1 Shanyou 63, the most widely cultivated hybrid at the time, 240 F9 RILs were produced by single seed descent. RILs were then randomly paired to produce 360 crosses, called immortalized F2 (IMF2). Originally, 231 molecular markers were genotyped in RILs and later with extensive genome sequences an ultrahigh-density marker coverage was obtained to produce 1,619 bins for RILs (Zhou et al. 2012). Both RILs and IMF2 were planted together and measured for 4 traits (yield, tiller, grains, and grain weight).

For this population, we orient marker genotypes of P_1 (Zhenshan 97) as $A_{l1}A_{l1}$ and those of P_2 (Minghui 63) as $A_{l2}A_{l2}$ for all l markers. After removing uninformative data, 209 RILs and 276 IMF2 as well as P_1 (Zhenshan 97), P_2 (Minghui 63), and F_1 (Shanyou 63) with a total of 488 samples were used for genetic model fitting and estimation in a joint analysis. Thus, let $s \in (P_1, P_2, F_1,$ RILs and IMF2). The genetic model for a quantitative trait y_s is defined as

$$y_s = u + \sum_{l=1}^{m}(a_l w_{sl} + d_l v_{sl}) + \sum_{k<l}^{m}(aa_{kl} w_{sk} w_{sl}) + e_s$$

$$w_{sl} = \begin{cases} 1 & \text{for } A_{l1}A_{l1} \\ 0 & \text{for } A_{l1}A_{l2} \\ -1 & \text{for } A_{l2}A_{l2} \end{cases}$$

$$v_{sl} = \begin{cases} -1/2 & \text{for } A_{l1}A_{l1} \\ 1/2 & \text{for } A_{l1}A_{l2} \\ -1/2 & \text{for } A_{l2}A_{l2} \end{cases}$$

This is the F2 genetic model (G2A model with $p_k = 1/2$) as the population is an F2-type segregating population. The heterosis is $H = y_{F_1} - (y_{P_1} + y_{P_2})/2$ and is expected to be $E(H) = \sum_{l=1}^{m} d_l - \sum_{k<l}^{m} aa_{kl}$ (Melchinger et al 2007; Garcia et al 2008). Also, $E(y_{P_1} - y_{P_2}) = \sum_{l=1}^{m} 2a_l$. Thus, by estimating the genetic model parameters

($\hat{a}_l$, $\hat{d}_l$, and $\widehat{aa}_{kl}$), we can compare the estimated and observed heterosis and the parental difference on a quantitative trait.

For this dataset, we used the same model selection procedure as for the maize dataset, except that 80% (rather than 50%) of the samples without replacement were drawn for stability selection. This is to ensure enough QTL with epistasis selected for the final estimation for a proper interpretation of the genetic basis of heterosis, as the sample size in the rice dataset is much smaller than that in the maize dataset.

The selected genetic model information is in Supplementary Table 2. The main results are in Table 2, which shows the comparison of the observed and estimated heterosis and parental mean differences for the 4 traits. Table 2 shows that they agree with each other very well, a general property of the model-based estimation. For yield, QTL dominance contributes 98% of heterosis and epistasis contributes 2% of heterosis. However, for grains, the contributions of QTL dominance and epistasis are similar (48% vs 52%). Table 2 also reports the estimates of the mean degree of dominance MDD = $\overline{d_i}/\overline{|a_i|}$ of QTL for the 4 traits.

Figure 4 shows the selection target of QTL effects (T_a, T_d, and T_{aa}) for the 4 traits. However, it needs to be pointed out that due to the relatively small sample size, the model fitting and estimation for the rice data are not stable. The sample size of this study may be insufficient to have a reliable model-based inference with epistasis. If we remove QTL epistasis, an estimation of QTL additive and dominant effects is stable.

Advantages of G2A model

A genetic model comes with a genetic explanation and thus can be used to explain many genetic phenomena, such as heterosis, as demonstrated in this study. This is the main advantage of a model-based genetic inference. This G2A model can fit to any population and thus can be conveniently used to study the genetic structure of a population. As the G2A model is a binary model, the genetic structure can be represented through a bifurcating tree structure with tree branches proportional to genetic effects and interaction networks connecting tree leaves. Combined with external variables in time (e.g. pedigree, breeding lines, response to selection) and space (e.g. environments), this genetic structure can help to illuminate the evolutionary process and to project the future.

Conclusion

We presented a general theory to analyze and interpret the genetics of heterosis. We applied it to the dataset of a maize study that is a factorial design between a group of male and female inbred lines and their hybrid crosses, and to a rice study that is an F2-type population derived from a cross between 2 inbred lines and the F1 hybrid. The conclusion on the relative contribution of dominance vs epistasis to yield heterosis is clear. For maize ear weight, the main contribution is QTL dominance (overdominance to be precise) and the contribution of epistasis is relatively minor. For rice yield, the main contribution is still QTL dominance.

What we presented in this paper is a vision for general quantitative genetic data analysis and interpretation. First, we need to recognize that the fundamental genetic basis of quantitative trait variation is multiple genes. Thus, only when those QTL were identified and fitted in a model for a joint estimation and interpretation, can we have a fuller understanding of the genetic structure in a population, particularly pertaining to the history and evolutionary process that brought about the population. In this inference, a quantitative genetic model or QTL model is the key and the bridge that connects genome to phenome. This connection can implicate the past—evolutionary history. It is the genome that connects individuals, populations, and past, and it is phenotypes that are directly subject to (natural or artificial) selection or other evolutionary driving forces. The inference of the genetic structure in a population, represented by an inferred QTL model, can also be used to project the future that can be explored for a more efficient breeding design and selection scheme in the context of plant and animal breeding.

The underpinning of this vision is that QTL are the elements of quantitative trait variation. The mapping of QTL establishes the physical link to the underlying genes and the joint estimation of QTL effects can reveal the properties of gene actions in a population. This joint estimation and inference are broadly speaking statistically consistent and evolutionary continuous and thus is the right path for general quantitative genetic analysis and interpretation.

Data availability

The original experimental data were from Xiao *et al.* (2021) and Zhou *et al.* (2012). The information of the inferred genetic model is provided in Supplementary Table 1 for the maize study and Supplementary Table 2 for the rice study and is the basis for all the figures and tables. All relevant data information (data, R source codes, and analysis results) of this study is publicly available in Dryad (https://doi.org/10.5061/dryad.vq83bk44r).

Supplemental material available at GENETICS online.

Acknowledgments

Z.-B.Z. conceived the idea and wrote the paper, G.D.S.G. and L.M. performed the data analysis, Y.X. and J.Y. provided the data and stimulating discussion that led to this study. All contributed to the revision of the paper. We thank the reviewers for the constructive comments.

Funding

This project was funded by National Institute of Food and Agriculture, United States Department of Agriculture (2020-51181-32156 and 2022-67013-36269) (for Z.-B.Z., G.D.S.G., and L.M.) and the National Natural Science Foundation of China (32122066) (for Y.X. and J.Y.).

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

Anderson VL, Kempthorne O. 1954. A model for the study of quantitative inheritance. Genetics. 39(6):883–898. doi:10.1093/genetics/39.6.883.

Crow JF, Kimura, M. 1970. An Introduction to Population Genetics Theory. New York (NY): Harper and Row. doi:10.1002/tera.1420050318.

Duvick DN. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). Maydica. 50:193.

Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 52(2):399–433. doi:10.1017/S0080456800012163.

Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock W, Dennis E. 2018. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breed Sci. 68(2): 145–158. doi:10.1270/jsbbs.17155.

Garcia AAF, Wang S, Melchinger AE, Zeng Z-B. 2008. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics. 180(3):1707–1724. doi:10.1534/genetics.107.082867.

Gu ZL, Han B. 2024. Unlocking the mystery of heterosis opens the era of intelligent rice breeding. Plant Physiol. 196(2):735–744. doi:10.1093/plphys/kiae385.

Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang QF. 2002. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics. 162(4): 1885–1895. doi:10.1093/genetics/162.4.1885.

Hua J, Xing Y, Wu W, Xu C, Sun X, Yu S, Zhang Q. 2003. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci U S A. 100(5):2574–2579. doi:10.1073/pnas.0437907100.

Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, *et al.* 2016. Genomic architecture of heterosis for yield traits in rice. Nature. 537(7622):629–633. doi:10.1038/nature19760.

Jiang Y, Schmidt R, Zhao Y, Reif JC. 2017. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nat Genet. 49(12):1741–1746. doi:10.1038/ng.3974.

Krieger U, Lippman ZB, Zamir D. 2010. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat Genet. 42(5):459–463. doi:10.1038/ng.550.

Kusmec A, Zheng Z, Archontoulis S, Ganapathysubramanian B, HuG, Wang L, Yu J, Schnable PS. 2021. Interdisciplinary strategies to enable data-driven plant breeding in a changing climate. One Earth. 4(3):372–383. doi:10.1016/j.oneear.2021.02.005.

Laurie C, Wang S, Carlini-Garcia LA, Zeng ZB. 2014. Mapping epistatic quantitative trait loci. BMC Genet. 15(1):112. doi:10.1186/s12863-014-0112-9.

Li H, Yang Q, Fan N, Zhang M, Zhai H, Ni Z, Zhang Y. 2017. Quantitative trait locus analysis of heterosis for plant height and ear height in an elite maize hybrid zhengdan 958 by design III. BMC Genet. 18(1):36. doi:10.1186/s12863-017-0503-9.

Meier L, Van De Geer S, Bühlmann P. 2008. The group lasso for logistic regression. J R Stat Soc B Stat Methodol. 70(1):53–71. doi:10.1111/j.1467-9868.2007.00627.x.

Meinshausen N, Bühlmann P. 2010. Stability selection. J R Stat Soc B Stat Methodol. 72(4):417–473. doi:10.1111/j.1467-9868.2010.00740.x.

Melchinger AE, Utz HF, Piepho HP, Zeng Z-B, Schon CC. 2007. Quantitative genetic theory to elucidate the role of epistasis in the manifestation of heterosis. Genetics. 177(3):1815–1825. doi:10.1534/genetics.107.077537.

Simon N, Friedman J, Hastie T, Tibshirani R. 2013. A sparse-group lasso. J Comput Graph Stat. 22(2):231–245. doi:10.1080/10618600.2012.681250.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics. 132(3):823–839. doi:10.1093/genetics/132.3.823.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. J R Stat Soc B Methodol. 58(1):267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

True JR, Liu J, Stam LF, Zeng ZB, Laurie CC. 1997. Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*. Evolution. 51(3): 816–832. doi:10.1111/j.1558-5646.1997.tb03664.x.

Wang T, Zeng ZB. 2006. Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibrium. BMC Genet. 7(1):9. doi:10.1186/1471-2156-7-9.

Weber KE. 1990. Selection on wing allometry in *Drosophila melanogaster*. Genetics. 126(4):975–989. doi:10.1093/genetics/126.4.975.

Weber KE. 1992. How small are the smallest selectable domains of form? Genetics. 130(2):345–353. doi:10.1093/genetics/130.2.345.

Weber K, Eisman R, Morey L, Patty A, Sparks J, Tausek M, Zeng ZB. 1999. An analysis of polygenes affecting wing shape on chromosome three in *Drosophila melanogaster*. Genetics. 153(2):773–786. doi:10.1093/genetics/153.2.773.

Weber K, Eisman R, Higgins S, Morey L, Patty A, Tausek M, Zeng ZB. 2001. An analysis of polygenes affecting wing shape on chromosome 2 in *Drosophila melanogaster*. Genetics. 159(3):1045–1057. doi:10.1093/genetics/159.3.1045.

Xiao Y, Jiang S, Cheng Q, Wang X, Yan J, Zhang R, Qiao F, Ma C, Luo J, Li W, *et al.* 2021. The genetic mechanism of heterosis utilization in maize improvement. Genome Biol. 22(1):148. doi:10.1186/s13059-021-02370-7.

Zeng ZB, Liu J, Stam LF, Kao CH, Mercer JM, Laurie CC. 2000. Genetic architecture of a morphological shape difference between two Drosophila species. Genetics. 154(1):299–310. doi:10.1093/genetics/154.1.299.

Zeng ZB, Wang T, Zou W. 2005. Modeling quantitative trait loci and interpretation of models. Genetics. 169(3):1711–1725. doi:10.1534/genetics.104.035857.

Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q. 2012. Genetic composition of yield heterosis in an elite rice hybrid. Proc Natl Acad Sci U S A. 109(39):15847–15852. doi:10.1073/pnas.1214141109.

Editor: Y. Yuan