

Homeologous Epistasis in Wheat: The Search for an Immortal Hybrid

Nicholas Santantonio,^{*1} Jean-Luc Jannink,^{*†} and Mark Sorrells^{*}

^{*}Cornell University, Plant Breeding and Genetics Section, School of Integrated Plant Sciences, College of Agriculture and Life Sciences, Ithaca, New York 14853 and [†]United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Robert W. Holley Center for Agriculture and Health, Ithaca, New York 14853

ORCID IDs: 0000-0002-4351-4023 (N.S.); 0000-0003-4849-628X (J.-L.J.); 0000-0002-7367-2663 (M.S.)

ABSTRACT Hybridization between related species results in the formation of an allopolyploid with multiple subgenomes. These subgenomes will each contain complete, yet evolutionarily divergent, sets of genes. Like a diploid hybrid, allopolyploids will have two versions, or homeoalleles, for every gene. Partial functional redundancy between homeologous genes should result in a deviation from additivity. These epistatic interactions between homeoalleles are analogous to dominance effects, but are fixed across subgenomes through self pollination. An allopolyploid can be viewed as an immortalized hybrid, with the opportunity to select and fix favorable homeoallelic interactions within inbred varieties. We present a subfunctionalization epistasis model to estimate the degree of functional redundancy between homeoallelic loci and a statistical framework to determine their importance within a population. We provide an example using the homeologous dwarfing genes of allohexaploid wheat, *Rht-1*, and search for genome-wide patterns indicative of homeoallelic subfunctionalization in a breeding population. Using the IWGSC RefSeq v1.0 sequence, 23,796 homeoallelic gene sets were identified and anchored to the nearest DNA marker to form 10,172 homeologous marker sets. Interaction predictors constructed from products of marker scores were used to fit the homeologous main and interaction effects, as well as estimate whole genome genetic values. Some traits displayed a pattern indicative of homeoallelic subfunctionalization, while other traits showed a less clear pattern or were not affected. Using genomic prediction accuracy to evaluate importance of marker interactions, we show that homeologous interactions explain a portion of the nonadditive genetic signal, but are less important than other epistatic interactions.

KEYWORDS allopolyploidy; homeologous; epistasis; subfunctionalization; heterosis; genomic prediction; wheat

WHOLE genome duplication events are ubiquitous in the plant kingdom. The impact of these duplications on angiosperm evolution was not truly appreciated until the ability to sequence entire genomes elucidated their omnipresence (Soltis *et al.* 2009). Haldane (1933), postulated that single gene duplication allowed one copy to diverge through mutation while metabolic function was maintained by the other copy. Ohno (1970) reintroduced this hypothesis, and it has since been validated both theoretically (Ohta 1987; Walsh 1995; Lynch and Conery 2000) and empirically (Blanc and Wolfe 2004; Duarte *et al.* 2005; Liu *et al.* 2011;

Assis and Bachtrog 2013). The duplicated gene hypothesis does not, however, generally explain the apparent advantage of duplicating an entire suite of genes. The necessity of genetic diversity for plant populations to survive and adapt to divergent or changing environments may help to explain this pervasive phenomenon.

The need for gene diversity can become more immediate in plants than in animals, where the latter can simply migrate to “greener pastures” when conditions become unfavorable. Plants lack substantial within-generation mobility and must therefore change gene expression to cope with changing environmental conditions. Many species maintain gene diversity through alternate splicing, but this has been shown to be less common in plants than in other eukaryotes (Nagasaki *et al.* 2005). Whole genome duplication can generate the raw materials for the maintenance of genetic diversity (Wendel 2000; Adams and Wendel 2005). Gault *et al.* (2018) demonstrated that similar sets of duplicated genes were preserved in two related genera, *Zea* and *Tripsacum*,

Copyright © 2019 by the Genetics Society of America
doi: <https://doi.org/10.1534/genetics.118.301851>

Manuscript received December 3, 2018; accepted for publication January 16, 2019; published Early Online January 23, 2019.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6913253>.

¹Corresponding author: Cornell University, Plant Breeding and Genetics Section, School of Integrated Plant Sciences, College of Agriculture and Life Sciences, 240 Emerson Hall, Ithaca, NY 14853. E-mail: ns722@cornell.edu

millions of years after a shared paleopolyploidization event. This conserved pattern in purifying selection suggests that, at least for some genes, there is a clear advantage to maintaining two copies.

The union of two complete, yet divergent, genomes during the formation of an allopolyploid introduces manifold novel gene pathways that can specialize to specific tissues or environments (Blanc and Wolfe 2004). Similar to diploid hybrids, the formation of an allopolyploid results in a homogeneous population, but heterozygosity is maintained across homeologous sites rather than homologous sites. Unlike diploid hybrids that lose heterozygosity in subsequent generations, the homeoallelic heterozygosity is fixed through selfing in the allopolyploid. Mac Key (1970) postulated a trade off between new-creating (allogamous) and self-preserving (autogamous) mating systems, where allopolyploids favor self-pollination to preserve diverse sets of alleles across their subgenomes. As such, an allopolyploid may be thought of as an immortalized hybrid, with heterosis fixed across subgenomes (Ellstrand and Schierenbeck 2000; Feldman *et al.* 2012). While still hotly debated, evidence is mounting that allopolyploids exhibit a true heterotic response as traditional hybrids have demonstrated (Wendel 2000; Adams and Wendel 2005; Chen 2010, 2013).

Birchler *et al.* (2010) note that newly synthesized allopolyploids often outperform their subgenome progenitors, and that the heterotic response appears to be exaggerated in wider interspecific crosses. This seems to hold true even within species, where autopolyploids tend to exhibit higher vigor from wider crosses (Bingham *et al.* 1994; Segovia-Lerma *et al.* 2004). The overwhelming prevalence of allopolyploidy to autopolyploidy in plant species (Soltis and Soltis 2009) may suggest that it is the increase in allelic diversity *per se* that is the primary driver for this observed tendency toward genome duplication. Instead of allowing genes to change function after a duplication event, alleles may develop novel function prior to their reunion during an allopolyploidization event. The branched gene networks of the allopolyploid may provide the organism with the versatility to thrive in a broader ecological landscape than those of its subgenome ancestors (Mac Key 1970; Ellstrand and Schierenbeck 2000; Osborn *et al.* 2003).

Subfunctionalization and neofunctionalization are often described as distinct evolutionary processes. Neofunctionalization implies the duplicated genes have completely novel, nonredundant function (Ohno 1970). Subfunctionalization is described as a partitioning of ancestral function through degenerative mutations in both copies, such that both genes must be expressed for physiological function (Force *et al.* 1999; Stoltzfus 1999; Lynch and Force 2000). However, barring total functional gene loss, many mutations will have some quantitative effect on protein kinetics or expression (Zeng and Cockerham 1993). Duplicated genes will demonstrate some quantitative degree of functional redundancy until the ultimate fate of neofunctionalization (*i.e.*, complete additivity) or gene loss (pseudogenization) of one copy. It

has been proposed that essentially all neofunctionalization processes undergo a subfunctionalization transition state (Rastogi and Liberles 2005).

If the mutations occur before the duplication event, as in allopolyploidy, the two variants are unlikely to have degenerative mutations. Instead, they may have differing optimal conditions in which they function or are expressed. The advantage of different variants at a single locus (alleles; Allard and Bradshaw 1964) or at duplicated loci (homeoalleles; Mac Key 1970) can result in greater plasticity to environmental changes. Allopolyploidization has been suggested as an evolutionary strategy to obtain the genic diversity necessary for invasive plant species to adapt to the new environments they invade (Ellstrand and Schierenbeck 2000; te Beest *et al.* 2012).

Adams *et al.* (2003) showed that some homeoallelic genes in cotton were expressed in an organ-specific manner, such that expression of one homeolog effectively suppressed the expression of the other in some tissues. These results have since been confirmed in other crops such as wheat (Pumphrey *et al.* 2009; Akhunova *et al.* 2010; Feldman *et al.* 2012; Pfeifer *et al.* 2014), and evidence for neofunctionalization of homeoallelic genes has been observed (Chaudhary *et al.* 2009). Differential expression of homeologous gene transcripts has also been shown to shift upon challenge with heat, drought (Liu *et al.* 2015), and salt stress (Zhang *et al.* 2016) in wheat, as well as water submersion and cold in cotton (Liu and Adams 2007).

Common wheat (*Triticum aestivum*) provides an example of an allopolyploid that has surpassed its diploid ancestors in its value to humans as a staple source of calories. Hexaploid wheat has undergone two allopolyploid events, the most recent of which occurred between 10,000 and 400,000 years ago, adding the D genome to the A and B genomes (Marcussen *et al.* 2014). The gene diversity provided by these three genome ancestors may explain why allohexaploid wheat has adapted from its source in southwest Asia to wide spread cultivation around the globe (Dubcovsky and Dvořák 2007; Feldman and Levy 2012).

In the absence of outcrossing in inbred populations, selection can act only on individuals, changing their frequency within the population. If the selection pressure changes (*e.g.*, for modern agriculture), combinations of homeoalleles within existing individuals may not be ideal for the new set of environments and traits. This presents an opportunity for plant breeders to capitalize on this feature of allopolyploids by making crosses to form new individuals with complementary sets of homeoalleles. Many of these advantageous combinations have likely been indirectly selected throughout the history of wheat domestication and modern breeding.

Dominance of homeologous genes is known to exist in wheat. For example, a single dominant red allele at any of the three homeologous kernel color genes on 3A, 3B, and 3D will confer a red kernel color (Allan and Vogel 1965; Metzger and Silbaugh 1970). Another crucial example involves the two

homeologous dwarfing genes (Allan *et al.* 1959; Gale *et al.* 1975; Gale and Marshall 1976; McVittie *et al.* 1978) important in the Green Revolution, which implemented semidwarf varieties to combat crop loss due to nitrogen application and subsequent lodging. These genes have been shown to exhibit a quantitative semidominant response (Börner *et al.* 1996). We discuss this example in detail, and use it as a starting point to justify the search for quantitative homeologous interactions genome-wide. While the effect of allopolyploidy has been demonstrated at both the transcript level and whole plant level, we are unaware of attempts to use genome-wide homeologous interaction predictors to model whole plant level phenotypes such as growth, phenology, and grain yield traits.

Using a soft winter wheat breeding population, we demonstrate that epistatic interactions account for a significant portion of genetic variance and are abundant throughout the genome. Some of these interactions occur between homeo-allelic regions and we demonstrate their potential as targets for selection. If advantageous homeoallelic interactions can be identified, they could be directly selected to increase homeoallelic diversity, with the potential to expand the environmental landscape to which a variety is adapted. We hypothesize that the presence of two evolutionarily divergent genes with partially redundant function leads to a less-than-additive gene interaction, and introduce this as a subfunctionalization model of epistasis.

Subfunctionalization Epistasis

We generalize the duplicate factor model of epistasis from Hill *et al.* (2008), by introducing a subfunctionalization coefficient, s , that allows the interaction to shift between the duplicate factor and additive models. Let us consider an ancestral allele with an effect a . Through mutation, the effect of this locus is allowed to diverge from the ancestral allele to have effects a^* and \tilde{a} in the two descendant species. When the two divergent loci are brought back together in the same nucleus, the effect of combining these becomes $s(a^* + \tilde{a})$ (Figure 1).

Values of $s < 1$, indicate a less-than-additive epistasis (Eshed and Zamir 1996), in this case, resulting from redundant gene function. When $s = 1/2$, and $a^* = \tilde{a}$, the descendant alleles have maintained the same function and the duplicate factor model is obtained. As s exceeds $1/2$, the descendant alleles diverge in function (*i.e.*, subfunctionalization), until s reaches 1, implying that the two genes evolved completely nonredundant function (*i.e.*, neofunctionalization). At the point where $s = 1$, the effect becomes completely additive.

For values of $s > 1/2$, the benefit of multiple alleles is realized in a model analogous to overdominance in traditional hybrids. As alleles diverge they can pick up advantageous function under certain environmental conditions. The homeo-heterozygote then gains an advantage if it experiences conditions of both adapted homeoalleles. Values

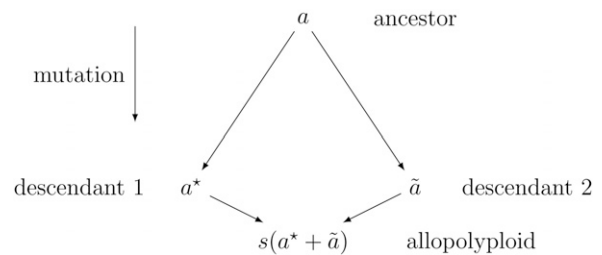


Figure 1 Diagram of subfunctionalization where a is the effect of a functional allele, a^* and \tilde{a} are the effects of the descendant alleles, and s is the subfunctionalization coefficient.

of $s < 1/2$ may indicate allelic interference (Herskowitz 1987), or genomic shock (McClintock 1984)—a phenomenon that has been observed in many newly formed allopolyploids (Comai *et al.* 2003). Allelic interference, also referred to as dominant negative mutation, can result from the formation of nonfunctional homeodimers, while homodimers from the same ancestor continue to function properly. This interference effectively reduces the number of active dimers by half (Herskowitz 1987; Veitia 2007).

Epistasis models

Let us consider the two locus model, with loci B and C . Using the notation of Hill *et al.* (2008), the expected phenotype, $E[y]$, is modeled as

$$E[y] = \mu + B\alpha_B + C\alpha_C + BC(\alpha\alpha)_{BC}. \quad (1)$$

where μ is the population mean, B and C are the marker allele scores, BC is the pairwise product of those scores, α_B and α_C are the additive effects of the B and C loci and $(\alpha\alpha)_{BC}$ is the interaction effect.

We revisit two epistatic models, the “Additive \times Additive Model without Dominance or Interactions Including Dominance” (henceforth called “Additive \times Additive”) and the “Duplicate Factor” considered by Hill *et al.* (2008) that are relevant for this discussion. Omitting the heterozygous classes and letting a be the effect on the phenotype, these models can be tabulated as follows.

Additive \times Additive		Duplicate Factor			
	CC	cc		CC	cc
BB	2a	0	BB	a	a
bb	0	2a	bb	a	0

We propose a generalized Duplicate Factor epistatic model to estimate the degree of gene functional redundancy, or subfunctionalization.

Subfunctionalization		
	CC	cc
BB	$s(a^* + \tilde{a})$	a^*
bb	\tilde{a}	0

When markers are coded $\{0, 1\}$ for presence of the functional allele, the deviation from the additive expectation, δ , is estimated by $(\alpha\alpha)_{BC}$. δ can then be used to calculate the subfunctionalization coefficient, $s = 1 + \frac{\delta}{a^* + \tilde{a}}$ (Figure 2). The least squares expectation of additive and epistatic effects is then

$$E \begin{bmatrix} \mu \\ B\alpha_B \\ C\alpha_C \\ BC(\alpha\alpha)_{BC} \end{bmatrix} = \begin{bmatrix} \mu \\ a^* \\ \tilde{a} \\ \delta \end{bmatrix} = \begin{bmatrix} \mu \\ a^* \\ \tilde{a} \\ (s-1)(a^* + \tilde{a}) \end{bmatrix}$$

Epistatic contrasts

Epistatic interaction predictors must be formed from marker scores in order to estimate interaction parameters. These interaction predictors are typically calculated as the pairwise product of the genotype scores for their respective loci. This can lead to ambiguity in the meaning of those interaction effects depending on how the marker scores are coded. Different marker parameterizations can center the problem at different reference points (*i.e.*, different intercepts), and can scale the predictors based on allele or genotype effects (*i.e.*, different slopes).

When loci B and C are coded as $\{-1, 1\}$ for inbred genotypes, including the product of the marker scores, BC , corresponds to the Additive \times Additive model (Table 1). Changing the reference allele at either locus does not change the magnitude of effect estimates but will change their signs. Using $\{0, 1\}$ coding, BC corresponds to the subfunctionalization model and estimates δ directly. For this coding scheme, the magnitude and sign can change depending on the reference allele at the two loci. This highlights one of the difficulties of effect interpretation, as it is not clear which marker orientations should be paired. That is, which allele should be B as opposed to b , and which should be C as opposed to c ? Marker alleles can be oriented to have either all positive or all negative additive effects, but the question remains: which direction should the more biologically active allele have on the phenotype?

Marker scores are typically assigned as either presence (or absence) of the reference, major, or minor allele, which may or may not be biologically relevant. While it has been noted that the two different marker encoding methods do not result in the same contrasts of genotypic classes (He *et al.* 2015; Martini *et al.* 2016, 2017), coding does not affect the least squares model fit (Zeng *et al.* 2005; Álvarez-Castro and Carlborg 2007). Álvarez-Castro and Carlborg (2007) show that there exists a linear transformation to shift between multiple parameterizations using a change-of-reference operation (see Appendix A). This is convenient because all marker orientation combinations can be generated easily by changing the effect signs of a single marker orientation fit for the $\{-1, 1\}$ marker coding. These effect estimates can subsequently be transformed to the $\{0, 1\}$ coding effect estimates using the change-of-reference operation for all marker orientation combinations.

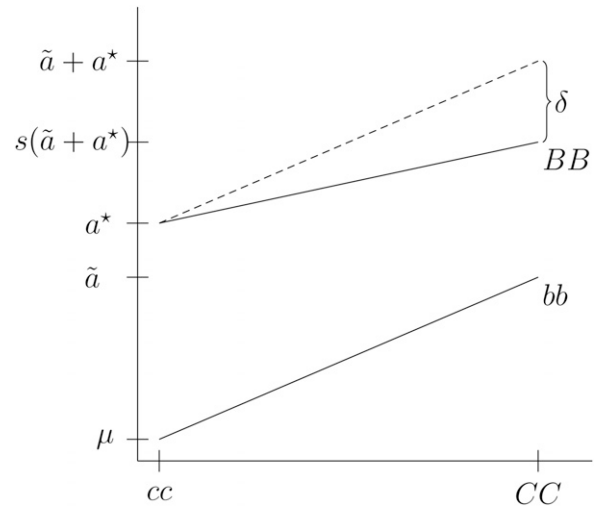


Figure 2 Epistatic interaction of two loci, B and C , with the expected effects for the $\{0, 1\}$ parameterization. δ indicates the deviation of the BCC genotype from an additive model for the $\{0, 1\}$ parameterization, where $s = 1 + \frac{\delta}{\tilde{a} + a^*}$. The dotted line indicates the expectation under the additive model.

This transformation does not hold when marker effects are considered random, where the interaction effect is subject to differential shrinkage depending on the marker coding and orientation (Martini *et al.* 2017, 2018). As such, orienting markers to capture functional allele relationships may be crucial for optimizing genomic prediction including epistasis. We make an attempt to orient markers based solely on estimated fixed marker additive effects, with the assumption that homeoalleles with similar additive effects are functionally similar. Other attempts at marker orientation have included orienting markers to maximize the interaction effect magnitude and including interaction predictors from all possible marker orientations (Martini *et al.* 2017). The former is biased toward selecting interaction predictors with a high joint frequency, whereas the latter suffers from a high degree of linear dependency.

Materials and Methods

Recombinant inbred line population

A biparental recombinant inbred line (RIL) population of 158 lines segregating for two dwarfing genes was used to illustrate an epistatic interaction between the well known homeologous genes on chromosomes 4B and 4D, $Rht-B1$ and $Rht-D1$, important in the Green Revolution (Allan *et al.* 1959; Gale *et al.* 1975; Gale and Marshall 1976; McVittie *et al.* 1978). Two genotyping by sequencing (GBS) markers linked to these genes were used to track the segregating mutant (b and d) and wildtype (B and D) alleles. Only one test for epistasis between these two markers was run. This homeologous marker pair was denoted RIL_Rht1. Details of the population can be found in Appendix B.

Table 1 Epistatic interaction score tables resulting from the products of marker scores using $\{-1, 1\}$ and $\{0, 1\}$ parameterizations for inbreds

	$\{-1, 1\}$		$\{0, 1\}$	
	CC	cc	CC	cc
BB	1	-1	1	0
bb	-1	1	0	0

CNLM population

The Cornell small grains soft winter wheat breeding population (CNLM) was used to investigate the importance of homeologous gene interactions in a large adapted breeding population. The dataset and a detailed description of the CNLM population can be found in Santantonio *et al.* (2019b). Briefly, the dataset consists of 1447 lines evaluated in 26 environments around Ithaca, NY. Because the data were collected from a breeding population, only 21% of the genotype/environment combinations were observed, totaling 8692 phenotypic records. Standardized phenotypes of four traits, GY, plant height (PH), heading date (HD), and test weight (TW) were recorded. All lines were genotyped with 11,604 GBS markers aligned to the International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0 wheat genome sequence of “Chinese Spring” (International Wheat Genome Sequencing Consortium *et al.* 2018), and subsequently imputed.

Homeologous marker sets

Using the IWGSC RefSeq v1.0 (International Wheat Genome Sequencing Consortium *et al.* 2018), homeologous sets of genes were constructed by aligning the annotated coding sequences (v1.0) back onto themselves. The known 4A, 5A, and 7B translocation in wheat (Devos *et al.* 1995) was ignored for simplicity in this study, but could easily be accounted for by allowing homeologous pairs across these regions. The resulting 23,796 homeologous gene sets, comprised of 18,184 triplicate and 5612 duplicate gene sets, sampled roughly 59% of the gene space of hexaploid wheat. Additional details on homeologous gene alignment can be found in Appendix C. Each homeologous gene was then anchored to the nearest marker by physical distance (Supplemental Material, Figures S1 and S2), and used to build a homeologous set of markers for each homeologous gene set. Redundant marker sets due to homeologous genes anchored by the same markers were removed, resulting in 6142 triplicate and 3985 duplicate marker sets for a total of 10,127 unique homeologous marker sets. Of the 11,604 markers, 8396 were included in at least one homeologous marker set. These marker sets (denoted “Homeo”) were then used to calculate marker interaction scores as pairwise products of the marker score vectors. The absence of one genotype class in 7912 interaction

terms resulted in 20,641 testable interaction effects out of 28,553 total interaction terms.

As a control, two additional marker sets were produced by sampling the same number of duplicate and triplicate marker sets as the Homeo set. These markers sets were sampled either from chromosomes within a subgenome (denoted “Within”, *e.g.*, markers on 1A, 2A, and 3A), or across nonsyntenic chromosomes of different subgenomes (denoted “Across”, *e.g.*, markers on 1A, 2B, and 3D). Samples were taken to reflect the same marker distribution of the Homeo set with regard to their native genome, which has a larger proportion of D genome markers relative to their abundance. Note that three-way homeologous interactions have equal proportions of markers belonging to the A, B, and D genomes, whereas D genome markers only account for 13% of all markers in the CNLM population (Santantonio *et al.* 2019b).

Determining marker orientation

For each homeologous marker set, additive homeologous marker effects and their multiplicative interaction effects were estimated as fixed effects using the $\{-1, 1\}$ marker parameterization in the following linear mixed model while correcting for background additive and epistatic effects.

$$\mathbf{y} = \tilde{\mathbf{Z}}\mathbf{S}_{-11}\mathbf{E}_{-11} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_{G+I} + \boldsymbol{\varepsilon} \quad (2)$$

where \mathbf{X} is the design matrix, $\boldsymbol{\beta}$ is the vector of fixed environmental effects, and \mathbf{Z} is the line incidence matrix. \mathbf{S}_{-11} is the matrix of genotype marker scores and interactions for each genotype class, while \mathbf{E}_{-11} is the fixed additive and interaction effects that need estimated (Appendix A). $\tilde{\mathbf{Z}}$ is the incidence matrix for the two- or three-way genotype of each homeologous marker set. \mathbf{Z} and $\tilde{\mathbf{Z}}$ differ in that the former links observations to a specific line, whereas the latter links observations to one of the two- or three-way genotype classes for the homeologous marker set. The background genetic effects were assumed to be $\mathbf{g}_{G+I} \sim \mathcal{N}(0, \sigma_G^2\mathbf{K}_G + \sigma_I^2\mathbf{H}_I)$ with population parameters previously determined (Zhang *et al.* 2010). The additive and epistatic covariances, \mathbf{K}_G and \mathbf{H}_I , were calculated as described in VanRaden (2008, method I) and Jiang and Reif (2015, Equation 5), respectively.

A Wald test was used to obtain a *P*-value for marker additive and interaction effects. Two Bonferroni corrections were used to determine if additive effects [$0.05/8396 = -\log_{10}(6.0 \times 10^{-6}) = 5.2$] or interaction effects [$0.05/20,641 = -\log_{10}(2.4 \times 10^{-6}) = 5.6$] were significantly greater than zero. All marker orientation combinations were generated by changing the estimated effect signs, and then transformed to the $\{0, 1\}$ marker effect estimates using the change-of-reference operation (Álvarez-Castro and Carlborg 2007). Only marker orientations with all positive or all negative additive effects were considered. It should be noted that the marker orientation has no effect on the *P*-value, as they are linear transformations of one another.

Markers were oriented to minimize the difference (or variance for three-way sets) of the additive main effects while maximizing the mean of the absolute values of the additive main effects. This orientation, which we denote “low additive variance high additive effect” (LAVHAE), assumes that marker alleles with similar effects are functionally similar. Only additive effects were used to select the marker orientation to keep from systematically selecting marker orientations with a specific interaction pattern. Three other marker orientation schemes were also investigated by orienting markers to either have all positive (POS) effects, all negative (NEG) effects, or to maximize the variance of the additive and interactions effects (“high total effect variance,” HTEV).

Additive only simulated controls

Marker effect and interaction estimates using either $\{0, 1\}$ or $\{-1, 1\}$ marker parameterizations are not orthogonal, so care must be taken when interpreting the direction and magnitude of the effects estimates. The positive covariance between the marker scores and their interaction leads to a multicollinearity problem, and results in a negative relationship between additive and interaction effects if both additive effects are oriented in the same direction. To determine if the negative relationship between the additive and epistatic effects was greater than expected due to multicollinearity, a new phenotype with no epistatic effects was simulated from the data for each trait. The estimate of the marker variance was calculated from the additive genetic variance estimate as $\hat{\sigma}_m^2 = \hat{\sigma}_G^2(2\mathbf{p}^T(1-\mathbf{p}))^{-1}$, where \mathbf{p} is the vector of marker allele frequencies. Then, for each trait, a new additive phenotype was simulated as $\mathbf{y}_{sim} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{M}\mathbf{u}_{sim} + \boldsymbol{\varepsilon}_{sim}$ where the trial effect estimates from Santantonio *et al.* (2019b, Equation 2) were used for $\boldsymbol{\beta}$, \mathbf{M} is the matrix of marker scores, \mathbf{u}_{sim} was sampled from $\mathcal{N}(0, \hat{\sigma}_m^2)$, and $\boldsymbol{\varepsilon}$ was sampled from $\mathcal{N}(0, \hat{\sigma}^2)$ as estimated in Santantonio *et al.* (2019b, Equation 2). A Kolmogorov-Smirnov (KS) test was used to determine if the distribution of the estimated interaction effects from the actual data differed from the distribution of effects estimated from simulated data. An additional simulated phenotype was also produced by first permuting each column of \mathbf{M} to remove any effects due to linkage disequilibrium (LD) structure.

Genomic prediction

To determine the importance of epistatic interactions to the predictability of a genotype, a genomic prediction model was fit as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}_G + \mathbf{Z}\mathbf{g}_I + \boldsymbol{\varepsilon} \quad (3)$$

where $\mathbf{1}$ is a vector of ones, μ is the population mean. The random vectors of additive genotype, epistatic interactions, and errors were assumed to be distributed as $\mathbf{g}_G \sim \mathcal{N}(0, \sigma_G^2\mathbf{K})$, $\mathbf{g}_I \sim \mathcal{N}(0, \sigma_I^2\mathbf{H})$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$, respectively.

The additive covariance matrix, \mathbf{K} , was calculated using VanRaden (2008), method I. The epistatic covariance matrix \mathbf{H} was calculated either as defined by Jiang and Reif (2015, Equation 5) and Martini *et al.* (2016, Equation 9) to model all pairwise epistatic interactions using $\{-1, 1\}$ coding (Pairwise), or in a similar fashion as \mathbf{K} for oriented marker sets, where only unique products of marker variables were included instead of the marker variables. For the latter, the matrix was scaled with the sum of the joint marker variances as $(2\mathbf{q}^T(1-\mathbf{q}))^{-1}$, where \mathbf{q} is the joint frequency of individuals containing both the nonreference marker alleles. Three-way marker products were included if they were unique from the additive and pairwise product predictors.

A small coefficient of 0.01, was added to the diagonals of the covariance matrix to recover full rank lost in centering the matrix of scores prior to calculating the covariance. Five-fold cross-validation was performed by randomly assigning individuals to one of five folds for 10 replications. Four folds were used to train the model and predict the fifth fold for all five combinations. All models were fit to the same sampled folds so that models would be directly comparable to one another and not subject to sampling differences. Prediction accuracy was assessed by collecting genetic predictions for all five folds, then calculating the Pearson correlation coefficient between the predicted genetic values for all individuals and a “true” genetic value. The “true” genetic values were obtained by fitting a mixed model to all the data with fixed effects for environments and a random effect for genotypes, assuming genotype independence with a genetic covariance \mathbf{I} (Santantonio *et al.* 2019b).

Increase in genomic prediction accuracy from the additive model was used as a proxy to assess the relative importance of oriented marker interaction sets. To determine the proportion of nonadditive genetic signal attributable to each interaction set, the ratio of the prediction accuracy increase from the additive model using the interaction set (Homeo, Within, and Across) to the prediction accuracy increase from the additive model modeling all pairwise epistatic interactions (Pairwise) was used for comparison of models. The percentage of non-additive predictability was calculated as follows for each interaction set.

$$\frac{\text{accuracy(Interaction Set)} - \text{accuracy(Additive)}}{\text{accuracy(Pairwise)} - \text{accuracy(Additive)}}. \quad (4)$$

We recognize that because marker orientation was conducted on the full data set and not within each fold, prediction accuracy may be influenced by this orientation step. However, no selection of predictors was made, only a change of reference based on the estimated additive effect coefficients.

Software

ASReml-R (V3; Gilmour 1997; Butler 2009) was used to fit all mixed models. BLAST+ (Camacho *et al.* 2009) was used for

Table 2 Marker and epistatic effect estimates for *Rht-1D* and *Rht-1B* linked GBS markers for PH (cm) in 158 RIL lines derived from NY91017-8080 × Caledonia

Marker coding	Effect orientation	Intercept	<i>Rht-1B</i>	<i>Rht-1D</i>	<i>Rht-1B</i> × <i>Rht-1D</i>	s^a
{0, 1}	+	69.9	23.4	22.2	−12.2	0.73
{0, 1}	−	103.3	−11.2	−10.0	−12.2	1.58
{−1, 1}	+	89.7	8.6	8.0	−3.0	
{−1, 1}	−	89.7	−8.6	−8.0	−3.0	

Least squares effect estimates are for markers coded either using {0, 1} coding or {−1, 1}, and then oriented such that the two marker main effects are either both positive (+) or both negative (−).

^a The subfunctionalization coefficient calculated from the additive and interaction effects is shown for the {0, 1} marker coding.

coding sequence alignment. All additional computation, analyses, and figures were made using base R (R Core Team 2015) implemented in the Microsoft Open R environment 3.3.2 (Microsoft 2017) unless noted otherwise. Figure 1 and Figure 2 were created using the “tikz” package (Tantau 2018) for LaTeX. Figure 4 was made with the “circlize” R package (Gu *et al.* 2014). The R package “xtable” (Dahl 2016) was used to generate LaTeX tables in R.

Data availability

Phenotypes and genotypes for the CNLM population can be found in Santantonio *et al.* (2019b). A list of homeologous genes can be found in supplemental file “homeoGeneList.txt.” The supplemental file “HomeoMarkerSet.txt” contains nonunique marker sets anchored to each homeologous gene set. Unique marker sets used can be found in “uniqueHomeoMarkerSet.txt,” “WithinMarkerSet.txt,” “AcrossMarkerSet.txt” for the Homeo, Within, and Across marker sets. Marker and marker interaction effect estimates and *P*-values for the Homeo set can be found in “twoWayInteractions.txt” and “threeWayInteractions.txt” for two- and three-way marker interactions, respectively. Phenotypes and genotypes used in the RIL population are included in the “NY8080Cal.txt” file. KASP marker scores of *Rht-1B* and *Rht-1D* for 1259 CNLM lines can be found in “Rht1.txt.” Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6913253>.

Results and Discussion

Rht-1

RIL population: The markers linked to the *Rht-1B* and *Rht-1D* genes both had significant additive effects ($P < 10^{-10}$) and explained 19.6 and 20.5% of the variation in the height of the RIL population (Table S1). The test for a homeoallelic epistatic interaction between these *Rht-1* linked loci was also significant ($P = 0.0025$), but explained only 3.5% of the variance after accounting for the additive effects. Had we tested all pairwise marker interactions in this population, this test would not have passed a Bonferroni corrected significance threshold.

Effect estimates for the *Rht-1* markers and their epistatic interaction are shown in Table 2, for {0, 1} and {−1, 1} marker parameterizations, and for orientations where the marker main effects are both positive or both negative. The {0, 1} parameterization is arguably more intuitive, as effects

correspond directly to differences in genotype values (Figure 3). They both contain the same information and are equivalent for prediction using ordinary least squares, but the interpretation of the {−1, 1} marker coding is less obvious because the slopes are deviations from the expected double heterozygote (assuming no dominance), which does not exist in an inbred population. The {0, 1} parameterization uses the double dwarf as the reference point, where the effects α_B and α_C are the two semidwarf genotypic values. The tall genotype is the sum of the semidwarf allele effects plus the deviation coefficient, δ , which corresponds to $(\alpha\alpha)_{BC}$.

The estimated *s* parameter of 0.73 indicates a significant degree of redundancy between the wild-type *Rht-1* homeoalleles. This suggests that either the gene products maintain partial redundancy in function, or the expression of the two homeoalleles is somewhat redundant. The latter is less likely given that the two functional wild-type genes have comparable additive effects relative to the double dwarf. If the two genes were expressed at different times or in different tissues based on their native subgenome, the additive effects would be likely to differ in magnitude. This demonstrates a functional change between homeoalleles that has been exploited for a specific goal: semidwarfism.

When the markers are oriented in the opposite direction, to indicate the GA insensitive mutant allele as opposed to the GA sensitive wild-type allele, the interpretation of the interaction effect changes. The additive effect estimates become indicators of the reduction in height by adding a GA insensitive mutant allele. The interaction effect becomes the additional height reduction from the additive expectation of having both GA insensitive mutant alleles, resulting in a *s* parameter of 1.58. The same interpretation can be made, but must be done so with care. Losing wild-type function at both alleles results in a more drastic reduction in height than expected because there is redundancy in the system. Therefore, the *s* parameter is most easily interpreted when the functional direction of the alleles is known. Simply put, when function is added on top of function, little is gained, but when all function is removed, catastrophe ensues.

CNLM population: For the CNLM population, the markers with the lowest additive effect *P*-values associated with PH on the short arms of 4B and 4D did not show a significant interaction with their respective assigned homeologous

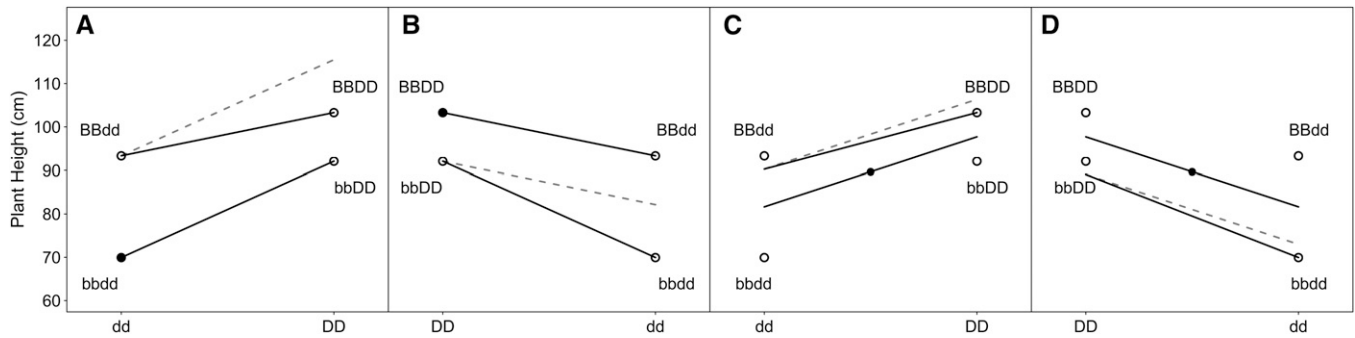


Figure 3 Epistasis plot of effects for *Rht-1B* and *Rht-1D* linked markers on PH in 158 RIL lines derived from NY91017-8080 × Caledonia. The filled circles indicate the intercept (*i.e.*, reference point) for each model parameterization while open circles indicate genotype class means. The solid lines indicate the marker effect estimates including the interaction term, while the dotted line indicates the expectation based on the additive model. (A) {0, 1} marker coding with positive marker effect orientation. (B) {0, 1} marker coding with negative marker effect orientation. (C) {-1, 1} marker coding with positive marker effect orientation, (D) {-1, 1} marker coding with negative marker effect orientation.

marker in homeologous sets H4.16516 and H4.23244. A new homeologous marker set, CNLM_Rht1, was constructed with the SNPs on 4BS and 4DS with the lowest *P*-values mentioned above. The additive effects of markers S4B_PART1_38624956 and S4D_PART1_10982050 had *P*-values of 5.5×10^{-4} and 3.7×10^{-8} , respectively, while the interaction had a *P*-value of 0.015. This set was oriented in the same direction as the RIL_Rht1 set using the LAVHAE orientation method. While the magnitude of these effects was reduced (7.13, 7.09, and -4.56 cm for the 4D, 4B, and 4B × 4D effects, respectively), the CNLM_Rht1 set had a *s* parameter value of 0.68, similar to that of RIL_Rht1. Had this set alone been tested, we would have concluded that this was a significant homeologous interaction.

To verify these results, we genotyped 1259 individuals of the CNLM population with two “perfect” markers designed to track the *Rht-1B* and *Rht-1D* alleles (Ellis *et al.* 2002). When correcting for population structure, effect estimates were 19.93 cm ($P = 3.3 \times 10^{-2}$), 23.08 cm ($P < 2 \times 10^{-16}$) and -12.28 cm ($P = 6.4 \times 10^{-6}$) for the *Rht-1B*, *Rht-1D* and *Rht-1B* × *Rht-1D* terms, respectively, resulting in an *s* value of 0.71. The relatively high *P*-value for the *Rht-1B* is likely due to correction for population structure, where the *Rht-1Db* dwarfing allele is the predominant source of semidwarfism in the breeding population (Table S2). Ignoring population structure produced *P*-values of $P < 10^{-19}$ for both additive effects and $P = 5.7 \times 10^{-5}$ for the interaction.

Significant homeoallelic interactions: Few homeoallelic interactions were significant at the trait-wise Bonferroni cut-off (Figure 4). Significant homeoallelic interactions for PH were identified between 4AL and 4DS, as well as 4BL and 4DL. Both of these locations were likely too far away from the *Rht-1* alleles to be tagging these genes directly, but they may be regulatory sites for these genes. Another set of interacting sites between the homeologous chromosome arms 3AS, 3BS, and 3DS was also identified for PH, but the additive effects were not significant. Two interacting regions on homeolog 1, between 1AS and 1DS and

between 1AL and 1DL, and three interacting regions on homeolog 5 also appeared to be influencing HD. One region on the distal end of homeolog 7 affected both HD and TW, with significant two-way and three-way interactions. Although they were tagged with different marker sets for the two traits, these epistatic regions appeared to colocalize within 2 Mbp.

No significant additive or interaction effects were detected for GY, highlighting the highly polygenic nature of the GY trait. In several cases, one of the additive effects was significant but the other was not, and it is not clear if this is influencing the detection of interactions. It may be that the significant marker is simply in higher LD with the functional mutation conditional on the presence of the other marker, allowing the interaction to pick up the additional signal from the functional mutation (Wood *et al.* 2014). However, if this were the case, the interaction would be expected to be in the same direction as the additive effect, which was not generally observed.

We did not detect an interaction between the two significant additive regions on 2B and 2D for the HD trait. While these two markers were not grouped as a homeologous set, they were tested as such based on their proximity to the well described Photoperiod-1 genes, *Ppd-B1* and *Ppd-D1*, on chromosomes 2B and 2D, respectively. These genes are known to influence photoperiod sensitivity, and therefore transition to flowering and HD (Welsh *et al.* 1973; Law *et al.* 1978; Scarth and Law 1983). Certain allele pairs at these genes have been shown to exhibit a high degree of epistasis (Wang *et al.* 2019) in a biparental family. It is unclear why no interaction was observed in this population.

Jiang *et al.* (2017) also investigated the presence of homeologous interactions, but found little evidence in a large population of hybrid wheat. They did not attempt to tag homeologous loci, but instead considered interactions across any markers on homeologous chromosomes to be syntenic. Interactions at homologous and nonhomeologous loci may have largely outweighed interactions

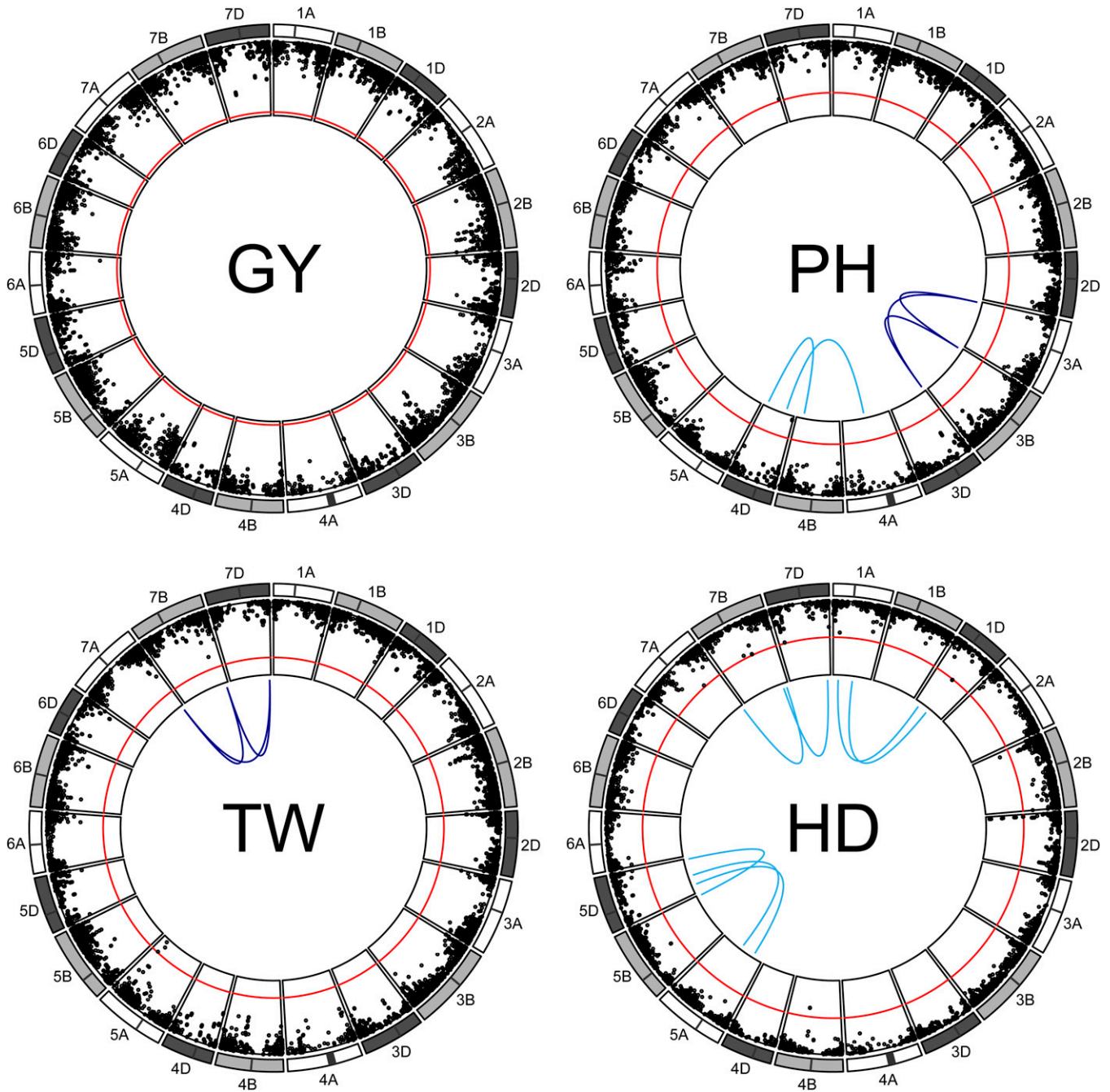


Figure 4 Manhattan plot of homeoallelic marker sets for each of the 21 chromosomes of wheat, where black circles indicate the $-\log(p\text{-value})$ of additive effect tests. The red line indicates a trait-wise Bonferroni significance threshold of 5.2 for those additive effects. Blue lines indicate significant two-way (light blue) and three-way (dark blue) homeoallelic marker interactions that exceeded a Bonferroni threshold of 5.6 (not shown) for all testable interaction effects.

across homeologous loci in that population, given it was constructed from highly divergent parents and that progeny were not inbred. Additionally, they tested all pairwise marker combinations, resulting in a strict significance threshold that may have missed small effect homeologous interactions.

Homeologous interactions make up relatively few of the potential two-way interactions within an allopolyploid genome. Given a subgenome with k genes and allopolyploidy level

p (i.e., the number of subgenomes), there are $k \binom{p}{2}$ two-way homeologous interactions vs. $\binom{kp}{2} - k \binom{p}{2}$ potential two-way nonhomeologous gene interactions. For a subgenome size of 30,000 genes, this represents 0.02 and 0.006% of the possible two-way gene interactions for an allotetraploid and an allohexaploid, respectively. That said, homeoallelic interactions should be far more likely to have a true biological interaction than random pairs of genes because

Table 3 Estimates of s coefficients for marker sets where both additive and the two-way interaction effects were significant at $P < 0.05$, combined for all four traits using {0, 1} marker coding

Marker set	$s < 0.5$	$0.5 \leq s < 1$	$s > 1$	Total ^a
Homeo	8	14	8	30***
Simulated additive	1	1	4	6
Across	9	7	1	17*
Within	6	3	4	13

The expected number of nonzero additive and two-way interactions effects based on a 0.05 significance threshold by chance is 11 (i.e., 4 traits \times 22,411 two-way interactions \times 0.05³). Coefficients have been grouped by categories related to the potential mode of epistasis, where $s < 0.5$ indicates a highly negative interaction, $0.5 \leq s < 1$ a less-than-additive interaction indicative of subfunctionalization for homeologous genes, and $s > 1$ which indicates positive, or greater-than-additive, epistasis. Three marker sets are shown, either across all homeologous loci (Homeo), sampled sets within (Within) and across (Across) nonsyntenic subgenome regions. An additional phenotype was simulated to contain no epistasis, and fit with the Homeo marker set (Simulated Additive).

^a *, **, *** indicate significantly greater than the expected number of significant sets at $P = 0.05$, 0.01 and 10^{-6} based the binomial distribution with 89,644 trials and a probability of 0.05³.

they should belong to the same or similar biochemical pathways.

Estimates of the subfunctionalization coefficient

There were few cases where at least two additive effects and their corresponding interaction effect were all significantly different from zero. This may be due to the difficulty of assigning functional homeologous gene sets using single SNPs, as well as a lack of statistical power owing to low minor allele frequencies (Hill *et al.* 2008). The lack of a large number of significant interactions is not surprising given that allele frequencies near 0.5 are uncommon in both natural and breeding populations.

To determine whether more homeologous marker sets were displaying a pattern indicative of subfunctionalization than would be expected by chance, marker sets where both additive and two-way interaction effects were significant at a threshold of $\alpha = 0.05$ were examined (Table 3). The expected number of two-way marker sets with significant additive and interaction effects is ~ 11 (i.e., four traits \times 22,411 two-way interactions \times 0.05³), assuming independence of loci and true additive and interaction effects of zero. Only the Homeo and Across marker sets had significantly more than expected. When broken down by trait, these appeared to be driven by interactions for PH and TW in the Homeo set (Table S3). The homeologous marker set had a larger proportion of s coefficients estimated between 0.5 and 1 relative to the strictly additive simulated phenotypes as well as the other nonhomeologous marker sets, suggesting that homeologous loci exhibit a pattern indicative of subfunctionalization more so than other marker sets tested. The Across set showed the highest proportion of $s < 0.5$, suggestive of gene pathway interference. Because the power to detect significant effects diminishes as more tests are accomplished, it may be prudent to look at global trends between homeologous additive effects and their interactions, regardless of statistical significance.

Evidence of subfunctionalization

A strong negative relationship between additive and interaction effects was observed when using the {0, 1} marker parameterization (Figure 5A). This negative relationship was also observed in the phenotypes simulated to be strictly additive (Figure S3). The multicollinearity of the additive and epistatic predictors at least partially drives this relationship, where positively correlated additive and epistatic predictors will tend to have effect estimates in opposing directions.

To determine if the interaction effects were greater in magnitude than expected by chance, the ordered interaction effects from the true and simulated phenotypes were plotted against one another to form a quantile–quantile plot (Figure 5B). The interaction effects were multiplied by the sign of the corresponding additive effects to highlight the direction of interaction effect relative to the additive effect. Interaction effect distributions were significantly different between the observed and strictly additive simulated data as determined by the KS test ($P < 0.05$) for all traits.

HD showed a pattern consistent with a subfunctionalization model, with a low dropping tail for interaction effects in the opposite direction than that of the corresponding additive effects. This indicates that the less-than-additive effects of some estimated interactions are greater than expected by additivity alone. PH showed some evidence of this pattern, but also demonstrated a greater-than-additive effect for positively related interaction effects. The LAVHAE orientation scheme may have selected the wrong marker coding for those marker sets, resulting in an s parameter > 1 , or there are true greater-than-additive interaction responses for positive effect alleles. Greater than additive responses would be indicative of overdominance across homeologous loci. GY and TW showed little evidence of the less-than-additive pattern, yet TW did show this trend when the HTEV marker orientation was used (Figures S4 and S5). These relationships were more pronounced when the markers were permuted to remove LD before simulating the data (Figure S6). High LD between homeologous marker sets may result in dampening of the epistatic signal due to unbalanced or missing genotype classes.

These findings are further supported by comparing the homeologous interactions to the Within and Across interaction effect estimates. The Homeo marker set showed more severe less-than-additive epistasis than both Within and Across for HD but not the other traits (Figures S7 and S8). The Within set had more severe less-than-additive interaction effects than the Homeo set for TW (Figure S7), and the Across had more severe less-than-additive effects for PH (Figure S8). Large or moderate effect negative epistasis is expected across subgenomes in allopolyploids, but it is unclear why this was also observed for the Within marker set for TW.

Homeologous model fit

Comparing variance component estimates across different unstructured covariance matrices can be misleading as variance

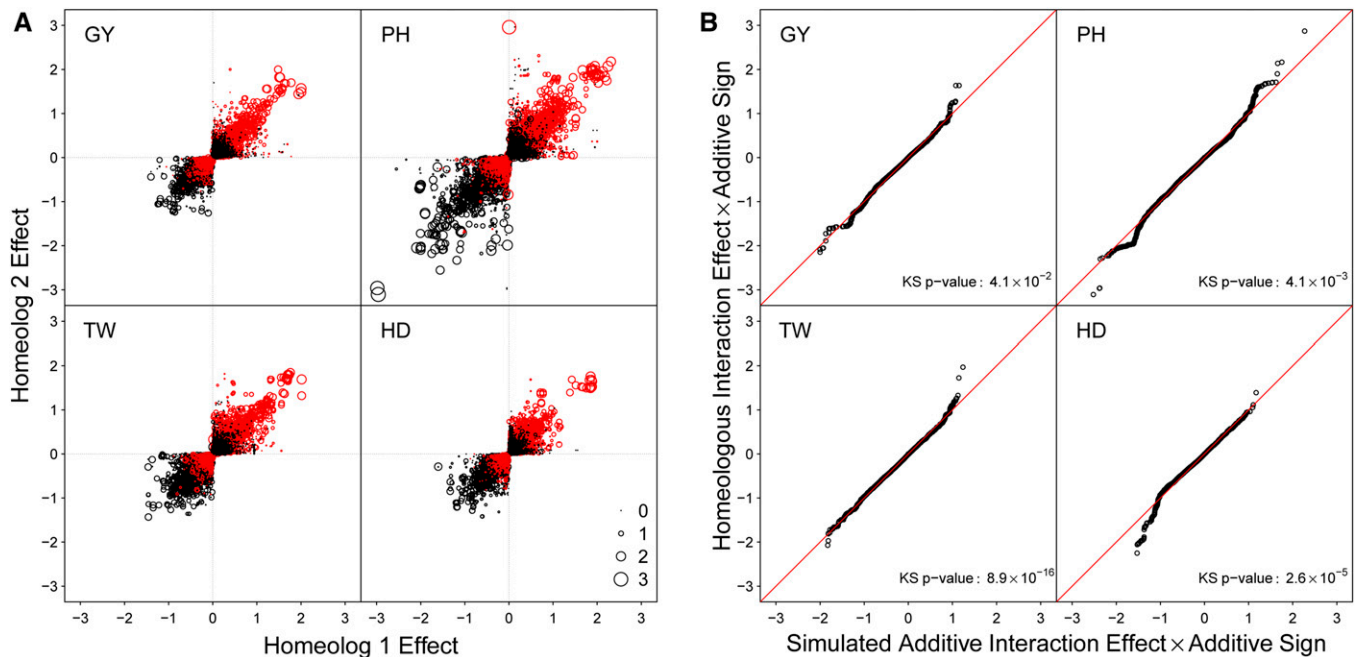


Figure 5 (A) LAVHAE oriented homeologous marker pair additive effects for four traits, GY, PH, TW, and HD using the $\{0, 1\}$ marker parameterization. Point size represents the magnitude of the two-way homeologous interaction effect while color denotes the direction of the interaction effect, where black is positive and red is negative. (B) Quantile–quantile plot of the ordered homeologous interaction effect estimates plotted against those from a simulated phenotype sampled to obtain no epistatic interactions. Interaction effects have been multiplied by the effect sign of the corresponding additive effects to emphasize the relationship between the additive and interaction effects. The lower left quadrant indicates a less-than-additive interaction, whereas the upper right quadrant indicates a greater-than-additive interaction. The P -value from a KS test is reported to test if the distributions of actual and simulated interaction effect estimates are the same. A deviation below the line on the bottom left of each graph (*i.e.*, a low dropping tail) should indicate a less-than-additive epistatic pattern of subfunctionalization, whereas a deviation above the line in the upper right (*i.e.*, a high rising head) should indicate a greater-than-additive epistasis pattern of homeologous overdominance.

components can be scaled by pulling a constant out of the covariance matrix. Additionally, variance partitioning is only reliable when the covariance matrices are truly independent (Huang and Mackay 2016; Jiang *et al.* 2017; Vitezica *et al.* 2017). Therefore, we do not make an attempt to discern meaning from the variance components *per se*, and instead focus the discussion on model fit diagnostics, as well as prediction accuracy from cross validation to determine the value of the predictive information included in the model.

All epistatic models using the $\{-1, 1\}$ marker parameterization provided a superior fit to the additive only model based on Akaike's information criterion (AIC) for all traits (Table S4). These results were confirmed by a likelihood ratio test to determine if the epistatic variance component was zero for all traits. With the exception of the GY trait, all of the epistatic models using the $\{0, 1\}$ marker parameterization also had nonzero variance components (Table S5), but did not result in a better fit for any models or traits. The LAVHAE method outperformed all other marker orientation schemes (Tables S6–S9). The Pairwise, Within, and Across epistatic models outperformed the Homeo marker interaction set for all traits.

Genomic prediction

All epistatic models resulted in higher prediction accuracies for all traits other than GY, where only marginal increases

were seen for certain marker interaction sets and parameterizations (Table 4). The $\{-1, 1\}$ marker coding resulted in higher prediction accuracies with a mean increase of 0.045 over the $\{0, 1\}$ coding, and ranged from 0.007 to 0.084 higher accuracy. This increase may be due to choosing the wrong orientation using the $\{0, 1\}$ marker coding effects. While these two codings are equivalent for prediction when marker effects are fixed, this is not the case for the mixed model genomic prediction environment (Martini *et al.* 2017, 2018). The discrepancy lies in shrinkage of interaction effects, where the $\{0, 1\}$ marker coding should result in greater shrinkage than the $\{-1, 1\}$ marker coding. This can be seen from a simple example with one observation of each genotypic class in $\{bbcc, bbCC, BBcc, BBCC\}$. The $\{-1, 1\}$ coding would have an interaction predictor of $\{1, -1, -1, 1\}$, whereas the $\{0, 1\}$ coding would have an interaction predictor of $\{0, 0, 0, 1\}$. This results in different numbers of observations per interaction class, with the $\{0, 1\}$ coding contrasting 3 and 1, versus 2 and 2 for the $\{-1, 1\}$ coding. Therefore the shrinkage of the $\{0, 1\}$ coding should be greater than for the $\{-1, 1\}$ coding. Martini *et al.* (2017), also noted that the $\{-1, 1\}$ marker coding has a 50% chance of choosing the wrong marker orientation if chosen at random, whereas the $\{0, 1\}$ marker coding has a 75% chance of choosing the wrong orientation.

Table 4 Prediction accuracies of whole genome Additive and Pairwise epistasis using $\{-1, 1\}$ coding, along with the Homeo, Within, and Across genome marker sets for both $\{-1, 1\}$ and $\{0, 1\}$ marker coding using the LAVHAE marker orientation

LAVHAE	Additive	Pairwise	Homeo ₋₁₁	Homeo ₀₁	Within ₋₁₁	Within ₀₁	Across ₋₁₁	Across ₀₁
GY	0.601 ^a	0.604	0.606 (167%) ^b	0.599 (-67%)	0.627 (867%)	0.600 (-33%)	0.630 (967%)	0.604 (100%)
PH	0.559	0.637	0.606 (60%)	0.580 (27%)	0.652 (119%)	0.570 (14%)	0.650 (117%)	0.584 (32%)
TW	0.515	0.576	0.560 (74%)	0.516 (2%)	0.596 (133%)	0.514 (-2%)	0.581 (108%)	0.525 (16%)
HD	0.664	0.712	0.692 (58%)	0.682 (38%)	0.710 (96%)	0.674 (21%)	0.722 (121%)	0.682 (38%)

^a Mean Pearson correlation between predicted and observed genetic values across 10 random 5-fold cross-validation replications.

^b The percentage of the nonadditive genetic predictability as relative to the Pairwise model is shown in parentheses (Equation 4).

The LAVHAE marker orientation scheme was superior for prediction of all traits and marker sets for the $\{-1, 1\}$ coding, but had little effect on the $\{0, 1\}$ marker coding (Tables S10–S12). This suggests that information can be gained from orienting markers relative to one another; however, it is still unclear what strategy should be used to orient pairs of markers. In this report, marker additive effects were forced to be either all positive or all negative to model the homeologous subfunctionalization hypothesis, but there may be more biologically relevant orientations not explored here. Martini *et al.* (2017) used a categorical interaction that included a predictor for each pairwise genotype. That model was shown to be less predictive than the $\{-1, 1\}$ multiplicative model, perhaps due to more linearly dependent predictors assumed to have nonzero effects. Feature selection may be useful for selecting the most informative interactions from this population of linearly dependent predictors. How an optimal set of orientations might be obtained without losing biological meaning of the orientation warrants further investigation.

The proportion of nonadditive genetic signal attributable to homeologous gene interaction was determined by taking the ratio of the percent increase in prediction accuracy of the Homeo, Within or Across prediction models from the additive model to the increase in prediction accuracy due to all pairwise interactions (Equation 4). All three marker sets resulted in higher genomic prediction accuracy than the additive only GBLUP model (*G*) when the $\{-1, 1\}$ marker coding was used. The homeologous marker interaction set explained between 58% and 167% of the additional genetic signal from the additive model. This result supports the idea that homeologous interactions are an important feature in the wheat genome. Conversely, Within and Across epistatic marker sets always resulted in a higher increase in genomic prediction accuracy relative to the Homeo marker set for all traits. This may suggest that the homeologous marker interactions are the least important relative to other epistatic interactions within and across the subgenomes, but could also be due to the paucity of these interactions relative to all possible two-way interactions, as previously discussed.

Another explanation might be provided by the relatively higher degree of LD across Homeo marker sets than found for the Within or Across marker sets. Homeologous marker sets were selected next to one another along syntenic regions of homeologous chromosome, and more often shared two of the three homeoallelic markers (Figure S9 and S10). The Within

and Across sets appear to have sampled the entire genome better than selecting only homeologous loci, as they track more unique pairs of genomic regions. Two additional samples of each Within and Across sets showed very similar outcomes to the samples shown here (see Figures S11–S14 and Tables S13–S15).

Homeologous LD

Homeologous marker sets had a much higher tendency to be coinherited together, as seen by relatively higher standardized LD values, *D'* (Lewontin 1964), than observed for either Within (KS test *P*-value = 1.1×10^{-6}) or Across (KS test *P*-value = 2.3×10^{-13}) marker sets (Figure 6). The greater fixation of allele pairs at homeologous regions may explain the lack of increased prediction accuracy of the Homeo marker set, but this may not diminish the importance of homeologous interactions. As sets of interactions are fixed within the population, the epistatic variance becomes additive (Hill *et al.* 2008). The higher degree of LD, *per se*, may indicate the importance of homeologous interactions.

The Green Revolution dwarfing genes are an excellent example of how pairs of homeoalleles may develop a tendency for coinheritance under selection or become fixed. In this example, the desirable phenotype is a semidwarf, due to its resistance to lodging. Therefore, wild-type *Rht-1B* alleles will usually be paired with a GA-insensitive *Rht-1D* dwarfing allele, while wild-type *Rht-1D* alleles will usually be found with a GA-insensitive *Rht-1B* dwarfing allele to confer the desirable semidwarf phenotype. The “perfect” *Rht-1* markers had a large standardized *D'* value of 0.89, indicating that pairs of alleles were being fixed in the population.

We recognize that it is also possible that the higher degree of LD observed between homeologous marker pairs could be due to misalignment of markers to the wrong subgenome. Markers assigned to the wrong homeolog would appear in high LD simply because they are physically located near their assigned homeologous partner on the same chromosome. We used strict filtering parameters to reduce the likelihood of misalignment. This included a threshold on observed heterozygosity in the population, which could indicate alignment to more than one subgenome.

Further Considerations

Wagner (2005) suggested that there are two potential drivers of less-than-additive (Eshed and Zamir 1996) or synergistic

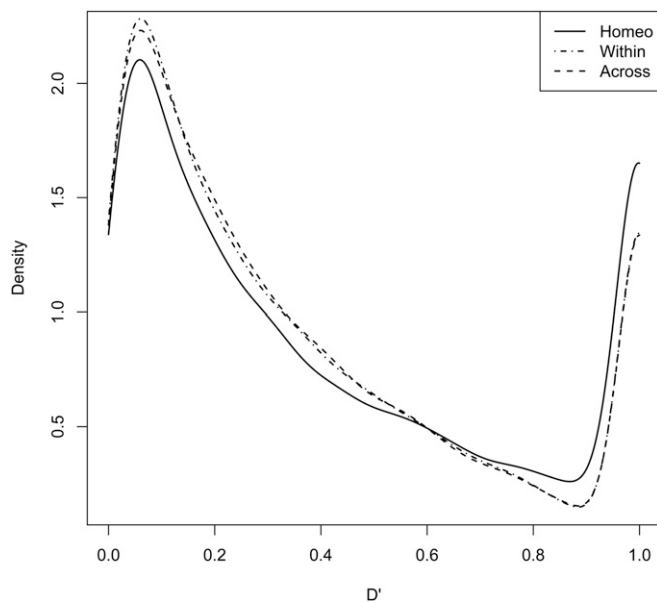


Figure 6 Smoothed densities of standardized D' statistics of linkage disequilibrium for expected and observed joint allele frequencies for Homeo, Within, and Across marker sets. KS tests were used to determine if the distribution of LD differed between Homeo and Within (KS P -value = 1.1×10^{-6}) or Across (KS P -value = 2.3×10^{-13}) marker sets.

(Segrè *et al.* 2005) epistasis. These drivers are (i) functional redundancy, as might be expected across homeologous loci; and (ii) distributed robustness of function, in which there can be are many pathways that can achieve the same outcome. Our observation that most epistasis is not due to homeologous interactions is supported by the findings of Jannink *et al.* (2009), who found the synergistic epistasis signal in a wheat dataset to be indicative of Wagner's distributed hypothesis, and not of the redundancy hypothesis.

It may be that there are few differences in protein function or expression across the three subgenomes, although this seems unlikely given mounting evidence that homeologous copies are differentially expressed in time, tissue and environment (Adams *et al.* 2003; Liu and Adams 2007; Chaudhary *et al.* 2009; Liu *et al.* 2011, 2015; Pfeifer *et al.* 2014; Zhang *et al.* 2016; Mutti *et al.* 2017). We were unable to assign homeologous pairs to all genes within the genome, suggesting that many of these potential sites for interacting loci were lost during polyploidization. Rapid loss of genetic material due to genome shock (McClintock 1984) is common in newly synthesized allopolyploids (Comai *et al.* 2003; Chen and Ni 2006), as has been shown in synthetic allopolyploid wheat (Ozkan *et al.* 2001; Kashkush *et al.* 2002). Other interacting loci may have undergone epigenetic (Comai 2000; Lee and Chen 2001; Comai *et al.* 2003) or transposon induced silencing of one or more homeoalleles (Kashkush *et al.* 2003; Wang *et al.* 2004).

The large portions of duplicated gene retainment across subgenomes suggest there is a benefit to their maintenance. Duplicate copies may be important contributors to differential genotype performance in contrasting environments.

Unfortunately, the CNLM dataset lacks sufficient genotype by environment variation to properly ask this question (data not shown). Experiments designed to explicitly model the phenotypic effect of differential homeologous gene expression across contrasting environments will be necessary to provide a satisfactory answer.

One of the challenges of using diverse panels of individuals is that marker proximity to a functional mutation is not necessarily indicative of high LD between the two sites. Significantly older or newer marker mutations may be in weak LD with a functional mutation despite close physical proximity, at least until a genetic bottleneck brings them back into high LD, such as in a biparental population (Flint-Garcia *et al.* 2003; Weir 2008). Other strategies to determine functional homeologous regions relax which sets of markers are considered homeologous. This has been accomplished by allowing pairwise relationships with all markers across entire subgenomes (Santantonio *et al.* 2019b) or on syntenic chromosome arms (Santantonio *et al.* 2019a), with mixed success. The construction of smaller haplotypes in a manner similar to Gao *et al.* (2017) may also improve functional pairing of homeologous alleles. Higher depth sequencing and advances in marker imputation may also aid in detection of homeologous epistasis.

The TILLING population developed by Krasileva *et al.* (2017) could be a useful resource for future investigation into homeoallelic gene interactions. Lines with complementary loss of function homeologous genes could be used to develop biparental mapping populations to test the degree of subfunctionalization with the high statistical power afforded by allele frequencies of 0.5. So called "synthetic" wheat populations formed by crossing common wheat with newly synthesized allohexaploids containing durum A and B genomes coupled to an *Ae. tauschii* D genome (*e.g.*, Sorrells *et al.* 2011) may prove powerful for detection of interactions between the common wheat homeologs and their durum and *Ae. tauschii* ancestors.

Conclusion

While much epistasis is partitioned to additive variance, it has been shown to be prevalent (Forsberg *et al.* 2017), and is important for maintaining long-term selection (Carlborg *et al.* 2006; Paixão and Barton 2016). Our results indicate that homeologous interactions contribute to the total genetic variance of the CNLM population. However, sampling interactions across nonsyntenic regions was superior for all traits examined, suggesting that homeologous epistasis make up a minority of the nonadditive genetic variance. The biological state of allopolyploids, along with the suggestive evidence presented here, demonstrate that there is value in further investigation of homeologous interactions.

The most important trait, GY, showed little to no evidence of homeologous subfunctionalization. This may be due to the highly polygenic nature of the trait, where essentially all functional genetic differences in the population should contribute to GY. Modern plant breeding has likely driven large

effect homeologous allele pair interactions to fixation in elite wheat genotypes. The implementation of the semidwarf phenotype provides perhaps the most important example where fixation of specific pairs of homeoalleles resulted in the single largest increase in wheat grain production in modern agriculture.

Prediction of unobserved homeologous allele pairs may prove difficult, as it currently is in diploid hybrids. However, large populations may be used to identify beneficial homeologous combinations that may subsequently be used for selection of unobserved lines before intensive field trials are conducted.

Treating the genome as consisting of purely additive gene action assumes that genes are independent machines, whose products sum to the final value of an individual. While convenient for selection, this is almost certainly not true when we consider the molecular mechanisms of biological organisms. Instead, genes work in concert to produce an observable phenotype. To this day, breeders of allopolyploid crops have treated allopolyploids as diploids for simplicity, but we now have the technical ability to view and start to breed these organisms as the ancient immortal hybrids that they are.

Acknowledgments

We are grateful to Jesse Poland's research group at Kansas State University for their contribution to genotyping of CNLM materials. We thank Gina Brown-Guedira at the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) Plant Science Research in the Department of Crop Science at North Carolina State University for genotyping the CNLM population for the *Rht-1B* and *Rht-1D* loci. The authors give special thanks the International Wheat Genome Sequencing Consortium for prepublication access to IWGSC RefSeq v1.0. The authors would also like to acknowledge Roberto Lonzano Gonzalez for the suggestion of using the coding sequences to identify homeologous genes. Finally, we would like to acknowledge the Cornell small grains staff, particularly David Benscher and James Tanaka, who were vital in implementing, collecting and processing the materials used to build the CNLM dataset. Funding of this research was provided by the USDA National Needs Fellowship for N.S., in partial fulfillment of the requirements for a Ph.D. in Plant Breeding and Genetics at Cornell University. The field trials comprising the phenotypic data for the CNLM population were funded in part by the Hatch Project # 149-447. Genotyping was funded by the Wheat Coordinated Agricultural Project (WheatCAP).

Note added in proof: See Santantonio *et al.* 2019 (pp. 675-684 and 685-695) in the G3 March 2019 issue for related work.

Literature Cited

Adams, K. L., and J. F. Wendel, 2005 Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8: 135-141. <https://doi.org/10.1016/j.pbi.2005.01.001>

Adams, K. L., R. Cronn, R. Percifield, and J. F. Wendel, 2003 Genes duplicated by polyploidy show unequal contributions to the

transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* 100: 4649-4654. <https://doi.org/10.1073/pnas.0630618100>

Akhunova, A. R., R. T. Matniyazov, H. Liang, and E. D. Akhunov, 2010 Homeolog-specific transcriptional bias in allopolyploid wheat. *BMC Genomics* 11: 505. <https://doi.org/10.1186/1471-2164-11-505>

Allan, R., and O. Vogel, 1965 Monosomic analysis of red seed color in wheat 1. *Crop Sci.* 5: 474-475. <https://doi.org/10.2135/cropsci1965.0011183X000500050030x>

Allan, R., O. Vogel, and J. Craddock, 1959 Comparative response to gibberellic acid of dwarf, semidwarf, and standard short and tall winter wheat varieties 1. *Agron. J.* 51: 737-740. <https://doi.org/10.2134/agronj1959.00021962005100120013x>

Allard, R. W., and A. Bradshaw, 1964 Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4: 503-508. <https://doi.org/10.2135/cropsci1964.0011183X000400050021x>

Álvarez-Castro, J. M., and Ö. Carlborg, 2007 A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* 176: 1151-1167. <https://doi.org/10.1534/genetics.106.067348>

Assis, R., and D. Bachtrog, 2013 Neofunctionalization of young duplicate genes in drosophila. *Proc. Natl. Acad. Sci. USA* 110: 17409-17414. <https://doi.org/10.1073/pnas.1313759110>

Bingham, E., R. Groose, D. Woodfield, and K. Kidwell, 1994 Complementary gene interactions in alfalfa are greater in autotetraploids than diploids. *Crop Sci.* 34: 823-829. <https://doi.org/10.2135/cropsci1994.0011183X003400040001x>

Birchler, J. A., H. Yao, S. Chudalayandi, D. Vaiman, and R. A. Veitia, 2010 Heterosis. *Plant Cell* 22: 2105-2112. <https://doi.org/10.1105/tpc.110.076133>

Blanc, G., and K. H. Wolfe, 2004 Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* 16: 1679-1691. <https://doi.org/10.1105/tpc.021410>

Börner, A., J. Plaschke, V. Korzun, and A. J. Worland, 1996 The relationships between the dwarfing genes of wheat and rye. *Euphytica* 89: 69-75. <https://doi.org/10.1007/BF00015721>

Butler, D., 2009 asreml: asreml() fits the linear mixed model. R package version 3.0. Available at: <https://asreml.kb.vsnr.co.uk/wp-content/uploads/sites/3/2018/03/ASReML-R-Models-Butler-et-al.pdf>. Accessed: October 2015.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 Blast+: architecture and applications. *BMC Bioinformatics* 10: 421. <https://doi.org/10.1186/1471-2105-10-421>

Carlborg, Ö., L. Jacobsson, P. Åhgren, P. Siegel, and L. Andersson, 2006 Epistasis and the release of genetic variation during long-term selection. *Nat. Genet.* 38: 418-420. <https://doi.org/10.1038/ng1761>

Chaudhary, B., L. Flagel, R. M. Stupar, J. A. Udall, N. Verma *et al.*, 2009 Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 182: 503-517. <https://doi.org/10.1534/genetics.109.102608>

Chen, Z. J., 2010 Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15: 57-71. <https://doi.org/10.1016/j.tplants.2009.12.003>

Chen, Z. J., 2013 Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14: 471-482. <https://doi.org/10.1038/nrg3503>

Chen, Z. J., and Z. Ni, 2006 Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28: 240-252. <https://doi.org/10.1002/bies.20374>

Comai, L., 2000 Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* 43: 387-399. <https://doi.org/10.1023/A:1006480722854>

- Comai, L., A. Madlung, C. Josefsson, and A. Tyagi, 2003 Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358: 1149–1155. <https://doi.org/10.1098/rstb.2003.1305>
- Dahl, D. B., 2016 xtable: export Tables to LaTeX or HTML. R package version 1.8–2. Available at: <https://CRAN.R-project.org/package=xtable>. Accessed: September 2017.
- Devos, K., J. Dubcovsky, J. Dvořák, C. Chinoy, and M. Gale, 1995 Structural evolution of wheat chromosomes 4a, 5a, and 7b and its impact on recombination. *Theor. Appl. Genet.* 91: 282–288. <https://doi.org/10.1007/BF00220890>
- Duarte, J. M., L. Cui, P. K. Wall, Q. Zhang, X. Zhang *et al.*, 2005 Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of arabidopsis. *Mol. Biol. Evol.* 23: 469–478. <https://doi.org/10.1093/molbev/msj051>
- Dubcovsky, J., and J. Dvořák, 2007 Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316: 1862–1866. <https://doi.org/10.1126/science.1143986>
- Ellis, M., W. Spielmeier, K. Gale, G. Rebetzke, and R. Richards, 2002 "perfect" markers for the rht-b1b and rht-d1b dwarfing genes in wheat. *Theor. Appl. Genet.* 105: 1038–1042. <https://doi.org/10.1007/s00122-002-1048-4>
- Ellstrand, N. C., and K. A. Schierenbeck, 2000 Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc. Natl. Acad. Sci. USA* 97: 7043–7050. <https://doi.org/10.1073/pnas.97.13.7043>
- Eshed, Y., and D. Zamir, 1996 Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* 143: 1807–1817.
- Feldman, M., and A. A. Levy, 2012 Genome evolution due to allopolyploidization in wheat. *Genetics* 192: 763–774. <https://doi.org/10.1534/genetics.112.146316>
- Feldman, M., A. A. Levy, T. Fahima, and A. Korol, 2012 Genomic asymmetry in allopolyploid plants: wheat as a model. *J. Exp. Bot.* 63: 5045–5059. <https://doi.org/10.1093/jxb/ers192>
- Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler, IV, 2003 Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54: 357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-I. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545.
- Forsberg, S. K., J. S. Bloom, M. J. Sadhu, L. Kruglyak, and Ö. Carlborg, 2017 Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* 49: 497–503. <https://doi.org/10.1038/ng.3800>
- Gale, M., and G. A. Marshall, 1976 The chromosomal location of gai 1 and rht 1, genes for gibberellin insensitivity and semi-dwarfism, in a derivative of norin 10 wheat. *Heredity* 37: 283–289. <https://doi.org/10.1038/hdy.1976.88>
- Gale, M. D., C. Law, and A. Worland, 1975 The chromosomal location of a major dwarfing gene from norin 10 in new British semi-dwarf wheats. *Heredity* 35: 417–421. <https://doi.org/10.1038/hdy.1975.112>
- Gao, N., J. W. Martini, Z. Zhang, X. Yuan, H. Zhang *et al.*, 2017 Incorporating gene annotation into genomic prediction of complex phenotypes. *Genetics* 207: 489–501. <https://doi.org/10.1534/genetics.117.300198>
- Gault, C. M., K. A. Kremling, and E. S. Buckler, 2018 *Tripsacum* de novo transcriptome assemblies reveal parallel gene evolution with maize after ancient polyploidy. *Plant Genome* 11. <https://doi.org/10.3835/plantgenome2018.02.0012>
- Gilmour, A., 1997 Asreml for testing fixed effects and estimating multiple trait variance components. *Proc. Assoc. Adv. Anim. Breed. Genet.* 12: 386–390.
- Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors, 2014 circlize implements and enhances circular visualization in R. *Bioinformatics* 30: 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>
- Haldane, J., 1933 The part played by recurrent mutation in evolution. *Am. Nat.* 67: 5–19. <https://doi.org/10.1086/280465>
- He, D., Z. Wang, and L. Parida, 2015 Data-driven encoding for quantitative genetic trait prediction. *BMC Bioinformatics* 16: S10. <https://doi.org/10.1186/1471-2105-16-S1-S10>
- Herskowitz, L., 1987 Functional inactivation of genes by dominant negative mutations. *Nature* 329: 219–222. <https://doi.org/10.1038/329219a0>
- Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4: e1000008. <https://doi.org/10.1371/journal.pgen.1000008>
- Huang, W., and T. F. Mackay, 2016 The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genet.* 12: e1006421. <https://doi.org/10.1371/journal.pgen.1006421>
- International Wheat Genome Sequencing Consortium (IWGSC) IWGSC RefSeq principal investigators Appels, R., K. Eversole, C. Feuillet *et al.*, 2018 Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361: eaar7191. <https://doi.org/10.1126/science.aar7191>
- Jannink, J.-L., L. Moreau, G. Charmet, and A. Charcosset, 2009 Overview of qtl detection in plants and tests for synergistic epistatic interactions. *Genetica* 136: 225–236. <https://doi.org/10.1007/s10709-008-9306-2>
- Jiang, Y., and J. C. Reif, 2015 Modeling epistasis in genomic selection. *Genetics* 201: 759–768. <https://doi.org/10.1534/genetics.115.177907>
- Jiang, Y., R. H. Schmidt, Y. Zhao, and J. C. Reif, 2017 A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49: 1741–1746. <https://doi.org/10.1038/ng.3974>
- Kashkush, K., M. Feldman, and A. A. Levy, 2002 Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160: 1651–1659.
- Kashkush, K., M. Feldman, and A. A. Levy, 2003 Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33: 102–106 [corrigenda: *Nat. Genet.* 47: 1099 (2015)]. <https://doi.org/10.1038/ng1063>
- Krasileva, K. V., H. A. Vasquez-Gross, T. Howell, P. Bailey, F. Paraiso *et al.*, 2017 Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. USA* 114: E913–E921. <https://doi.org/10.1073/pnas.1619268114>
- Law, C., J. Sutka, and A. Worland, 1978 A genetic study of day-length response in wheat. *Heredity* 41: 185–191. <https://doi.org/10.1038/hdy.1978.87>
- Lee, H.-S., and Z. J. Chen, 2001 Protein-coding genes are epigenetically regulated in arabidopsis polyploids. *Proc. Natl. Acad. Sci. USA* 98: 6753–6758. <https://doi.org/10.1073/pnas.121064698>
- Lewontin, R., 1964 The interaction of selection and linkage. I. general considerations; heterotic models. *Genetics* 49: 49–67.
- Liu, S.-L., G. J. Baute, and K. L. Adams, 2011 Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in arabidopsis thaliana. *Genome Biol. Evol.* 3: 1419–1436. <https://doi.org/10.1093/gbe/evr114>
- Liu, Z., and K. L. Adams, 2007 Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr. Biol.* 17: 1669–1674. <https://doi.org/10.1016/j.cub.2007.08.030>
- Liu, Z., M. Xin, J. Qin, H. Peng, Z. Ni *et al.*, 2015 Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*triticum aestivum* L.). *BMC Plant Biol.* 15: 152. <https://doi.org/10.1186/s12870-015-0511-8>

- Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155. <https://doi.org/10.1126/science.290.5494.1151>
- Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.
- Mac Key, J., 1970 Significance of mating systems for chromosomes and gametes in polyploids. *Hereditas* 66: 165–176.
- Marcussen, T., S. R. Sandve, L. Heier, M. Spannagl, M. Pfeifer *et al.*, 2014 Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345: 1250092. <https://doi.org/10.1126/science.1250092>
- Martini, J. W., V. Wimmer, M. Erbe, and H. Simianer, 2016 Epistasis and covariance: how gene interaction translates into genomic relationship. *Theor. Appl. Genet.* 129: 963–976. <https://doi.org/10.1007/s00122-016-2675-5>
- Martini, J. W., N. Gao, D. F. Cardoso, V. Wimmer, M. Erbe *et al.*, 2017 Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended gblup and properties of the categorical epistasis model (ce). *BMC Bioinformatics* 18: 3. <https://doi.org/10.1186/s12859-016-1439-1>
- Martini, J. W., F. Rosales, N.-T. Ha, T. Kneib, J. Heise *et al.*, 2018 Lost in translation: on the impact of data coding on penalized regression with interactions. arXiv:1806.03729.
- McClintock, B., 1984 The significance of responses of the genome to challenge. *Science* 266: 792–801. <https://doi.org/10.1126/science.15739260>
- McVittie, J. A., M. D. Gale, G. A. Marshall, and B. Westcott, 1978 The intra-chromosomal mapping of the norin 10 and tom thumb dwarfing genes. *Heredity* 40: 67–70. <https://doi.org/10.1038/hdy.1978.8>
- Metzger, R., and B. Silbaugh, 1970 Location of genes for seed coat color in hexaploid wheat, *triticum aestivum* l. 1. *Crop Sci.* 10: 495–496. <https://doi.org/10.2135/cropsci1970.0011183X001000050012x>
- Microsoft, 2017 *Microsoft R Open*. Microsoft, Redmond, WA.
- Mutti, J. S., R. K. Bhullar, and K. S. Gill, 2017 Evolution of gene expression balance among homeologs of natural polyploids. *G3 (Bethesda)* 7: 1225–1237. <https://doi.org/10.1534/g3.116.038711>
- Nagasaki, H., M. Arita, T. Nishizawa, M. Suwa, and O. Gotoh, 2005 Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. *Gene* 364: 53–62. <https://doi.org/10.1016/j.gene.2005.07.027>
- Ohno, S., 1970 *Evolution by Gene Duplication*. Springer, New York. <https://doi.org/10.1007/978-3-642-86659-3>
- Ohta, T., 1987 Simulating evolution by gene duplication. *Genetics* 115: 207–213.
- Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19: 141–147. [https://doi.org/10.1016/S0168-9525\(03\)00015-5](https://doi.org/10.1016/S0168-9525(03)00015-5)
- Ozkan, H., A. A. Levy, and M. Feldman, 2001 Allopolyploidy-induced rapid genome evolution in the wheat (*aegilops-triticum*) group. *Plant Cell* 13: 1735–1747. <https://doi.org/10.1105/tpc.13.8.1735>
- Paixão, T., and N. H. Barton, 2016 The effect of gene interactions on the long-term response to selection. *Proc. Natl. Acad. Sci. USA* 113: 4422–4427. <https://doi.org/10.1073/pnas.1518830113>
- Pfeifer, M., K. G. Kugler, S. R. Sandve, B. Zhan, H. Rudi *et al.*, 2014 Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345: 1250091. <https://doi.org/10.1126/science.1250091>
- Pumphrey, M., J. Bai, D. Laudencia-Chingcuanco, O. Anderson, and B. S. Gill, 2009 Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. *Genetics* 181: 1147–1157. <https://doi.org/10.1534/genetics.108.096941>
- Rastogi, S., and D. A. Liberles, 2005 Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* 5: 28. <https://doi.org/10.1186/1471-2148-5-28>
- R Core Team, 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Santantonio, N., J.-L. Jannink, and M. E. Sorrells, 2019a A low resolution epistasis mapping approach to identify chromosome arm interactions in allohexaploid wheat. *G3 (Bethesda)* 9: 675–684. <https://doi.org/10.1534/g3.118.200646>
- Santantonio, N., J.-L. Jannink, and M. E. Sorrells, 2019b Prediction of subgenome additive and interaction effects in allohexaploid wheat. *G3 (Bethesda)* 9: 685–695. <https://doi.org/10.1534/g3.118.200613>
- Scarath, R., and C. Law, 1983 The location of the photoperiod gene, *ppd2* and an additional genetic factor for ear-emergence time on chromosome 2b of wheat. *Heredity* 51: 607–619. <https://doi.org/10.1038/hdy.1983.73>
- Segovia-Lerma, A., L. Murray, M. Townsend, and I. Ray, 2004 Population-based diallel analyses among nine historically recognized alfalfa germplasms. *Theor. Appl. Genet.* 109: 1568–1575. <https://doi.org/10.1007/s00122-004-1784-8>
- Segrè, D., A. Deluna, G. M. Church, and R. Kishony, 2005 Modular epistasis in yeast metabolism. *Nat. Genet.* 37: 77–83. <https://doi.org/10.1038/ng1489>
- Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson *et al.*, 2009 Polyploidy and angiosperm diversification. *Am. J. Bot.* 96: 336–348. <https://doi.org/10.3732/ajb.0800079>
- Soltis, P. S., and D. E. Soltis, 2009 The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60: 561–588. <https://doi.org/10.1146/annurev.arplant.043008.092039>
- Sorrells, M. E., J. P. Gustafson, D. Somers, S. Chao, D. Benscher *et al.*, 2011 Reconstruction of the synthetic w7984 x Opata M85 wheat reference population. *Genome* 54: 875–882. <https://doi.org/10.1139/g11-054>
- Stoltzfus, A., 1999 On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49: 169–181. <https://doi.org/10.1007/PL00006540>
- Tantau, T., 2018 The TikZ and PGF Packages. Available at: <https://www.bu.edu/math/files/2013/08/tikzpgfmanual.pdf>. Accessed: April 2017.
- te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda *et al.*, 2012 The more the better? the role of polyploidy in facilitating plant invasions. *Ann. Bot.* 109: 19–45. <https://doi.org/10.1093/aob/mcr277>
- VanRaden, P., 2008 Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Veitia, R. A., 2007 Exploring the molecular etiology of dominant-negative mutations. *Plant Cell* 19: 3843–3851. <https://doi.org/10.1105/tpc.107.055053>
- Vitezica, Z. G., A. Legarra, M. A. Toro, and L. Varona, 2017 Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics* 206: 1297–1307. <https://doi.org/10.1534/genetics.116.199406>
- Wagner, A., 2005 Distributed robustness vs. redundancy as causes of mutational robustness. *BioEssays* 27: 176–188. <https://doi.org/10.1002/bies.20170>
- Walsh, J. B., 1995 How often do duplicated genes evolve new functions? *Genetics* 139: 421–428.
- Wang, J., L. Tian, A. Madlung, H.-S. Lee, M. Chen *et al.*, 2004 Stochastic and epigenetic changes of gene expression in arabidopsis polyploids. *Genetics* 167: 1961–1973. <https://doi.org/10.1534/genetics.104.027896>
- Wang, X., H. Xuan, B. Evers, S. Shrestha, R. Pless *et al.*, 2019 High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat. bioRxiv: 527911.

- Weir, B., 2008 Linkage disequilibrium and association mapping. *Annu. Rev. Genomics Hum. Genet.* 9: 129–142. <https://doi.org/10.1146/annurev.genom.9.081307.164347>
- Welsh, J. R., D. L. Kein, B. Pirasteh, and R. D. Richards, 1973 Genetic control of photoperiod response in wheat, pp. 879–884 in *Proceedings of the fourth International Wheat Genetics Symposium*, edited by E. R. Sears, and L. M. S. Sears. University of Missouri, Columbia, MO.
- Wendel, J. F., 2000 Genome evolution in polyploids, pp. 225–249 in *Plant Molecular Evolution*, Springer, New York. https://doi.org/10.1007/978-94-011-4221-2_12
- Wood, A. R., M. A. Tuke, M. A. Nalls, D. G. Hernandez, S. Bandinelli *et al.*, 2014 Another explanation for apparent epistasis. *Nature* 514: E3–E5. <https://doi.org/10.1038/nature13691>
- Zeng, Z.-B., and C. C. Cockerham, 1993 Mutation models and quantitative genetic variation. *Genetics* 133: 729–736.
- Zeng, Z.-B., T. Wang, and W. Zou, 2005 Modeling quantitative trait loci and interpretation of models. *Genetics* 169: 1711–1725. <https://doi.org/10.1534/genetics.104.035857>
- Zhang, Y., Z. Liu, A. A. Khan, Q. Lin, Y. Han *et al.*, 2016 Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*triticum aestivum* l.). *Sci. Rep.* 6: 21476. <https://doi.org/10.1038/srep21476>
- Zhang, Z., E. Ersoz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari *et al.*, 2010 Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42: 355–360. <https://doi.org/10.1038/ng.546>

Communicating editor: J. Birchler

Appendix

Appendix A: Change of Reference

Following Álvarez-Castro and Carlborg (2007), we demonstrate the change-of-reference operation simplified for inbred populations. For $\{0, 1\}$ marker coding and allowing G_1 to be the reference genotype, the genotypic values at a single locus can be represented as

$$\mathbf{G} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} = \mathbf{S}_{01}\mathbf{E}_{01} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ a \end{bmatrix}, \quad (5)$$

where \mathbf{S}_{01} is the marker score matrix using the $\{0, 1\}$ marker parameterization, and \mathbf{E}_{01} is the vector of expected values. For the two locus epistasis model, the four genotypic values are then

$$\mathbf{G} = \begin{bmatrix} G_{11} \\ G_{12} \\ G_{21} \\ G_{22} \end{bmatrix} = (\mathbf{S}_{01} \otimes \mathbf{S}_{01})\mathbf{E}_{01} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ a_1 \\ a_2 \\ a_1a_2 \end{bmatrix}. \quad (6)$$

The three locus interaction is extended by

$$\mathbf{G} = (\mathbf{S}_{01} \otimes \mathbf{S}_{01} \otimes \mathbf{S}_{01})[\mu \ a_1 \ a_2 \ a_1a_2 \ a_3 \ a_1a_3 \ a_2a_3 \ a_1a_2a_3]^T. \quad (7)$$

To shift from $\{-1, 1\}$ coding estimates, $\boldsymbol{\beta}_{-11}$, to $\{0, 1\}$ coding estimates, $\boldsymbol{\beta}_{01}$ the following transformation exists (Álvarez-Castro and Carlborg 2007). Let \mathbf{S}_{-11} indicate the $\{-1, 1\}$ marker parameterization

$$\mathbf{S}_{-11} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

then $\mathbf{E}_{01} = (\mathbf{S}_{01}^{-1} \otimes \mathbf{S}_{01}^{-1})(\mathbf{S}_{-11} \otimes \mathbf{S}_{-11})\mathbf{E}_{-11}$.

Appendix B: RIL Population

The population was formed from a cross between two Cornell soft winter wheat lines, NY91017-8080 and Caledonia. Caledonia contains a GA-insensitive 4D allele, d , and a wildtype 4B allele, B , while NY91017-8080 has a GA-insensitive 4B allele, b , and the wild type 4D allele, D . The population consisting of 192 individuals was planted in single row plots in Ithaca NY and measured for PH in 2008. The population was screened for loci influencing PH on chromosomes 4B and 4D using GBS markers. The markers with the lowest P -value on the short arms of 4B and 4D were used to indicate the *Rht-1* gene in this study. Only individuals with homozygous genotype calls for both loci were included to test for epistasis. This resulted in 19 double dwarfs (*bbdd*), 51 D genome semi-dwarfs (*BBdd*), 35 B genome semi-dwarfs (*bbDD*), and 53 tall (*BBDD*), for a total of 158 individuals. It appears that the Caledonia parent plant used in the cross was heterozygous for the D genome dwarfing allele, resulting in the 1:2 segregation ratio for the $d : D$ alleles, and was confirmed by the genotype call for that plant.

Appendix C: Coding Sequence Alignment

Alignments of coding sequences were accomplished with BLAST+, allowing up to 10 alignments with an e -value cutoff of $1e-5$. Alignments were considered only if they aligned to 80% or more of the query gene. Of the 110,790 coding sequences, 13,111 triplicate sets with one gene on each homeologous chromosome (representing 39,333 genes) were identified with no other alignments meeting the criterion. An additional 5073 triplicates (representing 15,219 genes) were added by selecting the top two alignments if they were on the corresponding homeologous chromosomes. Duplicate sets were also included if there was not a third alignment to one of the three subgenomes, adding an additional 5612 duplicates. The coding sequences for which we did not identify homeologous genes either appeared to be singletons (24,695 coding sequences) that did not have a good alignment to a gene on a homeologous chromosome, or had many alignments across the genome, making it impossible to determine with certainty which alignments were truly homeologous (20,319 coding sequences).