Interchromosomal Linkage Disequilibrium Analysis Reveals Strong Indications of Sign Epistasis in Wheat Breeding Families

Augusto Tessele[1], Geoffrey Morris[2], Eduard Akhunov[3], Blaine Johnson[4], Marshall Clinesmith[1], Addison Carroll Hope[1], Allan Fritz[1]

[1]Department of Agronomy, Kansas State University, Manhattan, Kansas, USA, 66506-5500 [2]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA, 80523-1101 [3]Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA, 66506-5500 [4]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, Nebraska, USA, 68583-0915 *Corresponding author: Department of Agronomy, Kansas State University, Manhattan, Kansas, USA, 66506-5500 E-mail: tessele@ksu.edu

Abstract

Additive gene action is assumed to underly quantitative traits, but the eventual poor performance of elite wheat lines as parents suggests that epistasis could be the underlying genetic architecture. Sign epistasis is characterized by alleles having either a beneficial or detrimental effect depending on the genetic background, which can result in elite lines that fail as parents in certain parental combinations. Hence, the objective of this study were to test the existence of sign epistasis and examine its consequences to wheat breeding. The presence of sign epistasis is expected to distort the allele frequency distribution between two interacting genes compared to neutral sites, creating strong linkage disequilibrium (LD). To test this hypothesis, analysis of interchromosomal LD in breeding families was performed and detected 19 regions in strong disequilibrium, whose allele frequency distribution matched the sign epistasis prediction and falsified the competing hypothesis of additive selection. To validate these candidate interactions while avoiding the biases of a circular analysis and the confounding effects of genetic drift, two independent sets of populations were analyzed. Genetic drift was attributed to creating the sign epistasis patterns observed in eleven interactions, but there was not sufficient evidence to reject the sign epistasis hypothesis in eight interactions. Sign epistasis may explain the poor performance of elite lines as parents, as crossing lines with complementary allelic combination re-establishes epistatic variance in the offspring. Reduction in the effective population size in certain crosses may also occur when unfavorable sign epistatic combinations are deleterious. The potential existence of di-genic and higher order epistatic interactions in elite germplasm can tremendously impact breeding strategies as managing epistasis becomes imperative for success.

Introduction

The infinitesimal model (Fisher, 1918) proposes that additive effects play a primary role in genetic inheritance, while relegating dominance and epistasis to lesser importance (Turelli, 2017). However, the eventual occurrence of elite lines performing poorly as parents suggests that epistasis might be the underlying genetic architecture. Sign epistasis, a significant form of gene action in evolutionary genetics (Wright, 1932), is characterized by alleles exhibiting either advantageous or detrimental effects depending on the allelic state of the interacting gene (Weinreich et al., 2005; Poelwijk et al., 2007). Under the influence of sign epistasis, genotypes can adapt to environmental pressures in various combinations of alleles, often leading to similar adaptation (Wright, 1988). In breeding, it implies the existence of elite genotypes with different frameworks of favorable allelic combinations. Hence, crossing contrasting genotypes causes the favorable allelic combinations to segregate in the offspring and results in mean performance lower than the midparent value. This clash of epistatic interactions between parental lines may justify the eventual poor performance of elite lines as parents. For instance, the widely grown hard red winter wheat cultivars *Jagger* and *2137*, both released in 1994, had comparable agronomic performance, reaching 34.8% and 23.1% of the acres in Kansas in 2000, respectively, yet *Jagger* appears in the pedigree of 21 subsequent releases, whereas *2137* contributed to only one (USDA ARS n.d.).

Establishing the importance and extent of epistasis in wheat breeding by modeling epistatic variance has tremendous challenges because of the lack of orthogonality with additive effects (Tessele et al., 2024; Raffo et al., 2022; Vitezica et al., 2017). However, the existence of sign epistasis interactions under selection may leave detectable signatures in the genome. The influence of sign epistasis in each locus closely resembles the effects of balancing selection, which maintains genetic diversity at sites under selection and can augment diversity at linked loci (Lewontin and Hubby, 1966; Hudson and Kaplan, 1988; Charlesworth et al., 1997; Takahata and Satta, 1998). The polymorphism flanking the loci under balancing selection can span over large or short regions, depending on local recombination frequency and strength of LD (Charlesworth et al., 1997; Wiuf et al. 2004; Charlesworth, 2006), the age of the advantageous polymorphism and the selection intensity (Kreitman and Rienzo, 2004; Tian et al., 2002). Besides affecting each locus independently, sign epistasis also impacts the relationship between interacting genes because selection acts on allelic combinations rather than on individual alleles. In a two-way interaction, the preferential selection of favorable allelic combinations distorts the allele frequency distribution of interacting genes compared to two neutral sites, creating strong disequilibrium that can be regarded as a selection signature.

Although selection for sign epistasis may leave signatures in the genome, the mere existence of LD does not imply sign epistasis. For instance, short-range linkage disequilibrium

(SRLD) primarily arises due to random genetic drift or common ancestry of chromosomal blocks (Koch et al., 2013). SRLD are perpetuated under low recombination, making it virtually impossible to disentangle from an eventual sign epistasis signature. Long-range linkage disequilibrium, however, suggests that a countervailing force to recombination maintains strong LD between distant sites, which is expected to be even more prominent in interchromosomal sites because the independent assortment of chromosomes recombines physically unlinked loci in every meiotic event. If sign epistasis is pervasive in wheat, breeding programs may perpetuate interchromosomal LD by continuously selecting favorable interactions. Detecting true selection signatures of epistasis (Lewontin et al., 1960) is challenging as other confounding effects can also create long-range LD, as population admixture (Maccaferri et al., 2005; Chao et al., 2010), genetic drift (Koch et al., 2013), artificial selection (Flint-Garcia et al., 2005; Joukhadar et al., 2019), or structural variation in chromosomes (Zhao et al., 2022). Disentangling signatures of epistasis from genetic artifacts may be possible through examining the allele frequency of sites in strong LD, as additive selection is expected to create a different allele frequency distribution than epistasis. Population structure can be minimized by analyzing breeding families independently. Analyzing interactions independently from the base population used for discovery may reduce the bias steaming from genetic drift and prevent statistical inference on a circular analysis (Kriegeskorte et al., 2009).

The main hypothesis is that sign epistasis exists and has major implications to wheat breeding. The objectives were to: (i) validate the existence of loci in abnormally strong interchromosomal linkage disequilibrium; (ii) test if selection for sign epistasis creates the observed strong interchromosomal LD; (iii) verify the existence of and disentangle confounding genetic artifacts that could likewise create strong interchromosomal LD.

Material and Methods

Plant Material and Genomic Data

The dataset consisted of genomic information on experimental lines from the Kansas State Hard Red Winter Wheat breeding program for the years 2011 through 2021, which were categorized based on the last stage of the breeding pipeline reached by each experimental line. The four stages in the breeding pipeline were early, preliminary, advanced, and elite yield trials, named IPSR, PYN, AYN and EYN, respectively. All experimental lines were genotyped in the IPSR stage using genotype-by-sequencing (GBS) (Poland et al. 2012). As part of the genotyping pipeline of the K-State wheat breeding program, markers with a MAF <0.01, more than 20% missing values and more than 20% of heterozygotes were filtered, as described in Jarquin et al (2017). The R package 'rrBLUP '(Endelman, 2011) was used for marker imputation, where missing markers were

106 imputed with the mean value among all lines for that marker and resulted in a total of 55,148
107 SNPs.
108
109 *Validating the existence of strong interchromosomal linkage disequilibrium*
110 This exploratory stages sought to validate the existence of interchromosomal linkage
111 disequilibrium in breeding germplasm. To test this hypothesis, experimental lines from the EYN
112 stage of the breeding program from 2010 to 2021 were considered. A linkage disequilibrium
113 analysis (r^2) focusing only on interchromosomal interactions was conducted and identified
114 pairwise markers in strong linkage disequilibrium. Validating the existence of strong
115 interchromosomal LD in wheat enabled further investigation into the hypothesis of selection
116 signatures of sign epistasis (Figure 1).
117

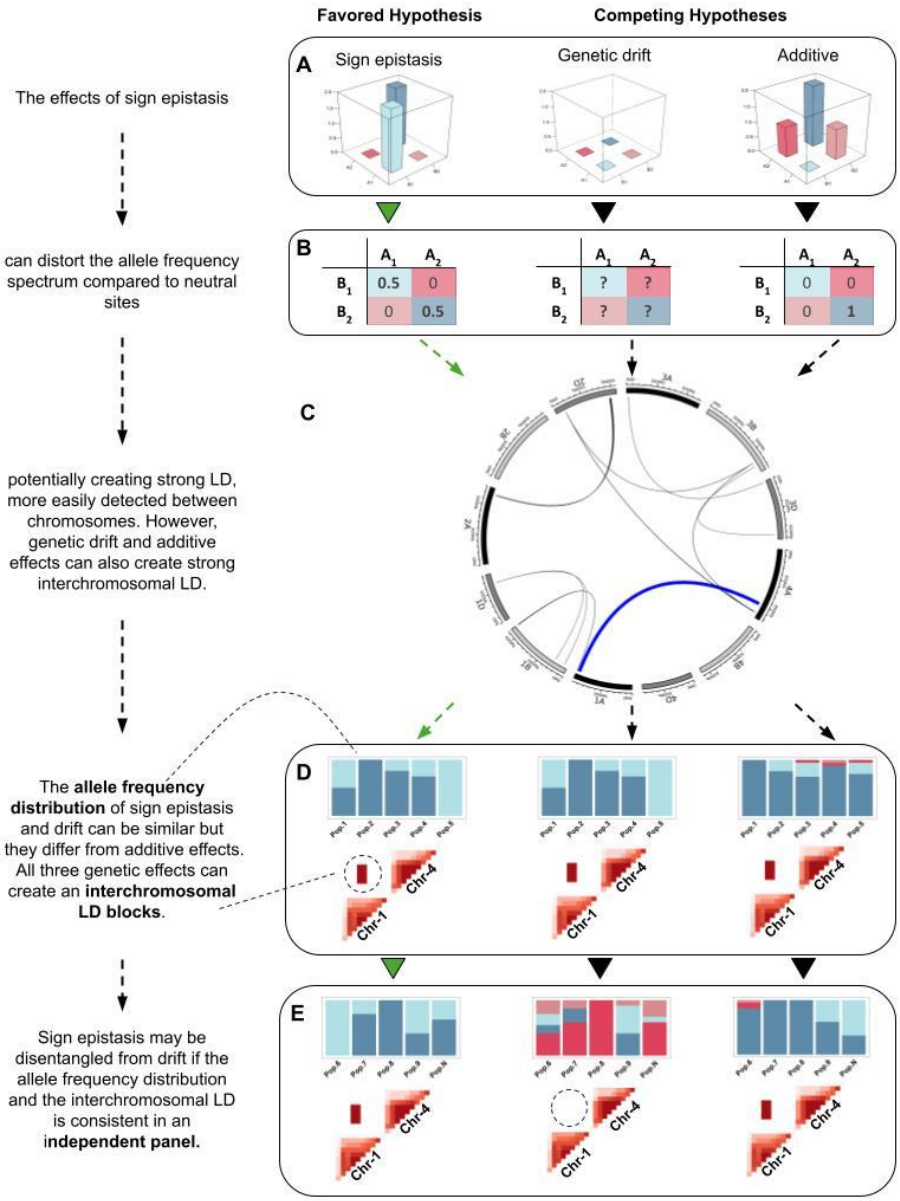


Figure 1 Selection of sign epistasis interactions can create interchromosomal LD. **A** depicts the levels of adaptation of two genes (A and B) under the effects of sign epistasis, drift and additive. **B** shows the expected allele frequency distribution under an ideal scenario for each genetic effect. **C** illustrates strong linkage disequilibrium between pairwise interchromosomal sites, where inner links connect sites in strong LD. **D** represents the allele frequency distribution in five populations for two genes in strong LD and the respective heatmap of intra- and inter-chromosomal LD for each genetic effect. **E** represents the allele frequency and LD values in an independent panel from D, which highlights the consistency of sign epistasis contrasting to the randomness of genetic drift.

Minimizing the influence of population structure and genetic drift

130 Historical data from 25 breeding families was used for the discovery of selection
131 signatures indicative of sign epistasis (Figure 1C). These were large families tested in the early
132 yield trial stage of the breeding pipeline, with sizes ranging from 20 to 45 genotypes. As a data
133 quality measure, a principal component analysis was conducted for each of the families to identify
134 and exclude any outlier genotypes detected in the two-dimensional plot defined by the first two
135 principal components to mitigate the potential introduction of population structure resulting
136 from genotyping or seed processing errors.
137 An interchromosomal linkage disequilibrium (r^2) analysis was conducted only on markers
138 with MAF>0.3 to reduce the total number of interactions to screen for and to increase the
139 likelihood of detecting interactions with allele frequency characteristic of sign epistasis (Figure 1A
140 and B). To circumvent the effect of population structure, the interchromosomal LD analysis was
141 conducted in each family individually and all interactions presenting $r^2 > 0.9$ were retained.
142 Although this approach controls population structure, it was highly influenced by genetic drift
143 within families. To address the genetic drift effect, the combined data of the 25 families was used
144 to run another interchromosomal LD analysis, and all interactions presenting $r^2>0.8$ were
145 selected. Interactions selected in both analyses and observed in a minimum of two families were
146 deemed worthy of further investigation. Supplemental Table 2 demonstrates this theoretical
147 framework for this approach.
148
149 *The Balancing Selection Nature of Sign Epistasis and Interchromosomal Linkage Blocks*
150 A sign epistatic interaction under selection, besides the characteristic intrachromosomal
151 haplotype block, may also create an "interchromosomal block" (Figure 1D). To investigate this
152 hypothesis, 40 markers flanking each candidate sign epistatic marker were used to run a linkage
153 disequilibrium analysis, and the resulting intra and interchromosomal LD values were used to
154 construct heatmaps of 2D LD plots.
155
156 *Allelic Frequencies to Differentiate Additive and Sign Epistasis*
157 Selection for sign epistasis is expected to change allele frequencies differently than
158 additive selection (Figure 1B). To empirically test this hypothesis, the allele frequencies of each
159 candidate interaction was computed, and the patterns were analyzed (Figure 1D). Since these
160 families were genotyped in the $F_{5:6}$ generation using GBS, heterozygous loci were still observed.
161
162 *Independent Family and Breeding Program Validation*
163 To minimize the effects of genetic drift and to avoid a circular analysis, 15 distinct
164 validation families were selected to reanalyze the heatmaps of 2D LD plots and the distribution

of allele frequencies in order to validate candidate interactions (Figure 1E). Although independent from the discovery set, the total number of families analyzed was relatively small and prone to the effects of genetic drift. To diminish the confounding effects of drift in small populations, data from a time span of 10 years in the wheat breeding program was also utilized to further validate the consistency of candidate interactions. The breeding program dataset was divided sequentially from early to elite yield trials and allele frequencies were calculated for each stage. In addition, the combined data of the breeding program was used to falsify candidate interactions if the opposite combination (relative to the most common allelic combination) had less than 10% of the total allelic combinations.

Simulation to Estimate Theoretical Distribution of Interchromosomal LD Values

To evaluate interchromosomal linkage disequilibrium (LD) attributable solely to genetic drift, we simulated 100 breeding crosses using random lines from the Kansas State University wheat breeding program as parents. Both single- and three-way crosses were generated, and the resulting populations (15 individuals each) were selfed to the fifth generation. Interchromosomal LD was assessed within each family individually and across all families combined. The same thresholds used to detect candidate interactions in the empirical discovery families were also applied to the simulated families and the results were compared with those from empirical discovery families. Simulations were performed using the R package AlphaRSim (Faux et al., 2016).

Software and Packages

The linkage disequilibrium analysis was performed using the software PLINK (Purcell et al., 2007). The R software packages “factoextra” (Kassambara et al., 2017), “circlize” (Gu and Gu, 2022), and “gaston” were utilized to perform the principal component analysis, to plot the SNPs presenting strong interchromosomal LD, and to plot the heatmap of 2D LD plots, respectively. The package “ggplot2” was used to plot allele frequency counts (Wickham, 2016). In addition, built-in functions in R were used during data management and to calculate allele frequencies.

Results

Simulated versus empirical interchromosomal linkage disequilibrium

The simulated families, generated under the sole effects of random genetic drift, displayed numerous pairwise interactions exhibiting some degree of interchromosomal linkage disequilibrium (LD) in at least two families, but only a few with R^2 values approaching 0.9 (Figure 2). In contrast, the empirical discovery families exhibited a much larger number of pairwise interactions exceeding these thresholds (Figure 2).

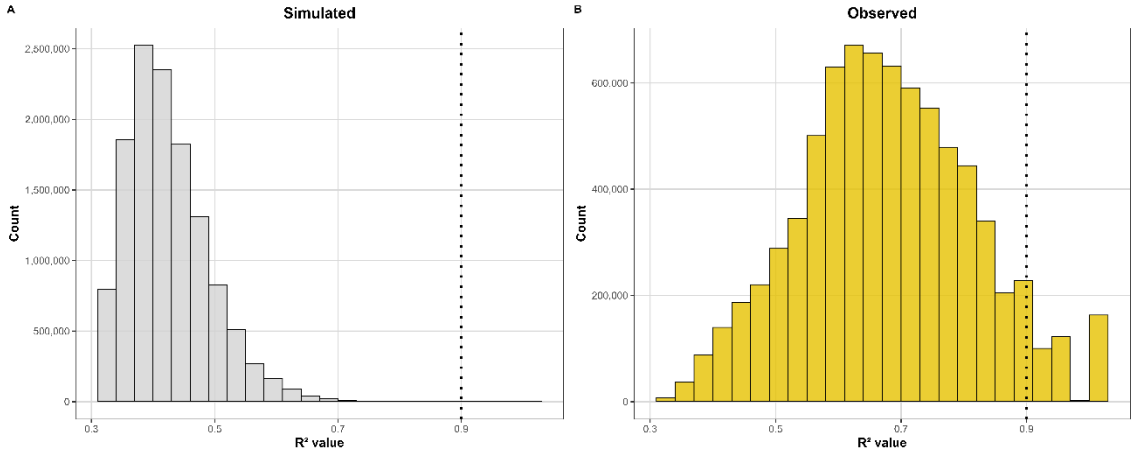


Figure 2. Simulated and observed distribution of interchromosomal R^2 values found in at least two families. (A) Simulated distribution of interchromosomal LD R^2 values under the sole influence of genetic drift. (B) Observed distribution of interchromosomal LD R^2 values in empirical discovery panel. The vertical dashed line marks the first threshold used to select pairwise interactions for further analysis.

When marker data from all populations were combined, the simulated families showed very few interactions with R^2 values near or above 0.8 (Figure 3), whereas the discovery families displayed a far greater number of pairwise interactions surpassing this threshold. Importantly, none of the simulated interactions met the criteria used to designate candidate interactions—namely, an average $R^2 > 0.9$ in at least two families individually and $R^2 > 0.8$ when analyzed across all populations combined.

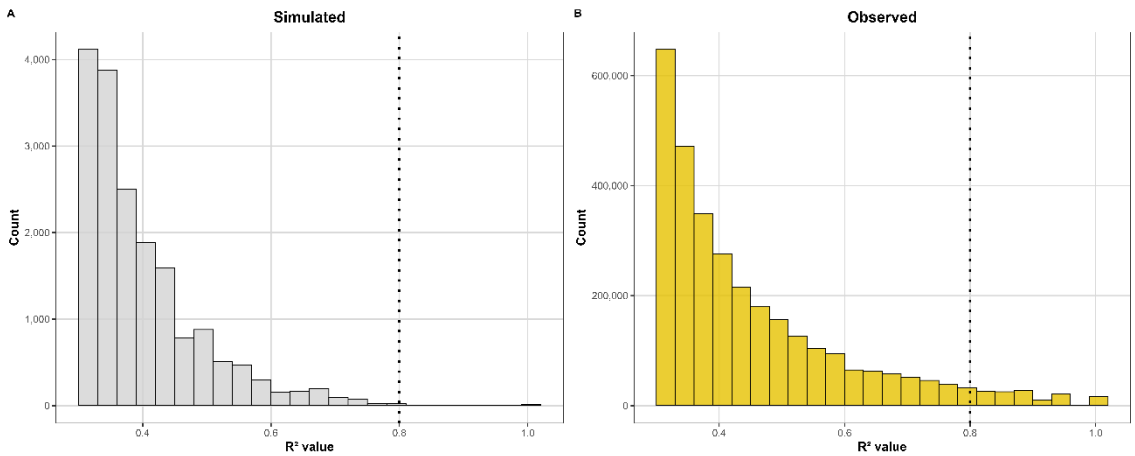


Figure 3. Simulated and observed distribution of interchromosomal R^2 values when analyzing populations collectively. (A) Simulated distribution of interchromosomal LD R^2 values under the sole influence of genetic drift. (B) Observed distribution of interchromosomal LD R^2 values in

empirical discovery panel. The vertical dashed line marks the second threshold used to select pairwise interactions for further examination.

Detection of loci in strong interchromosomal linkage disequilibrium

The goal of the discovery stage was to pinpoint a manageable number of interactions in strong LD while minimizing the likelihood of identifying artifacts of population structure, genetic drift, or additive selection. In total, 269 pairwise SNPs, which represent 19 interacting genomic regions, passed the set thresholds, and were considered candidate selection signatures (Figure 4). Wheat subgenome B had the most candidate interactions with 15, followed by A and D with 13 and 10, respectively. The chromosomes with the most candidate interactions were 1B and 6B, with four interactions each. In total, 14 out of the 19 interactions were intergenomic, and most involved subgenomes B and D. Subgenome A and B had candidate intragenomic interactions, which were absent in subgenome D. In addition, a candidate of a three-way selection signature was also observed in chromosomes 2D, 3B and 4A (Figure 4). A total of seven candidate interactions had opposed allelic combinations fixed in different families, and 12 were fixed for one allelic combination while the opposed allelic combination was observed in segregating families (Supplemental Table 1).

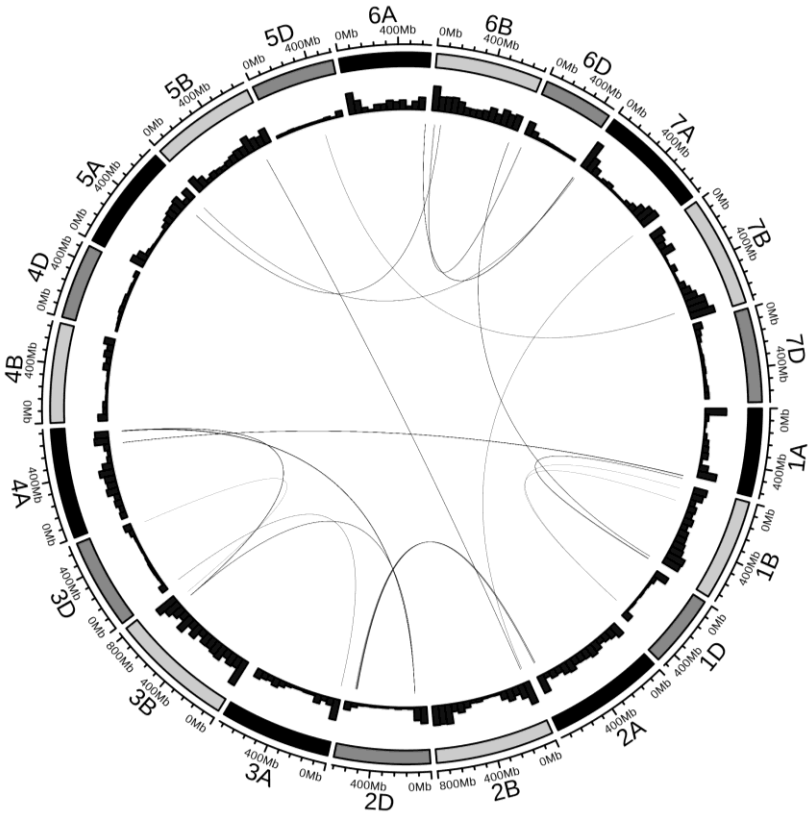


Figure 4 Candidate selection signatures of sign epistasis. The inner links represent regions (SNPs)

239 in strong interchromosomal LD following the discovery approach previously described. The outer
240 histogram plots represent the marker density in each chromosome. The outermost bar
241 represents the 7 chromosomes and 3 subgenomes of wheat.
242
243 *Putative balancing selection associated with sign epistasis*
244 The analysis of linkage drag was performed for each of the 19 candidate selection
245 signatures of sign epistasis. The results for the candidate interaction involving chromosomes 1A
246 and 4A reveal the presence of "interchromosomal linkage blocks" associated with the candidate
247 selection signature (Figure 5B). These blocks are primarily attributed to a sizable linkage block
248 located on chromosome 1A and a smaller block situated on chromosome 4A. The control
249 interaction, however, displayed some level of intrachromosomal LD on chromosome 2B but no
250 “interchromosomal LD block” was observed (Figure 5B).
251 The remaining interactions presented “interchromosomal linkage blocks” of variable
252 sizes, as the number of flanking SNPs in strong LD with the candidate interaction was variable
253 (Supplemental Figures 1-18). For example, candidate interactions 5D-7B, 2D-3B, 1BS-1D and 1BL-
254 1D were in intrachromosomal LD with a few SNPs, resulting in very small “interchromosomal LD
255 blocks”.
256

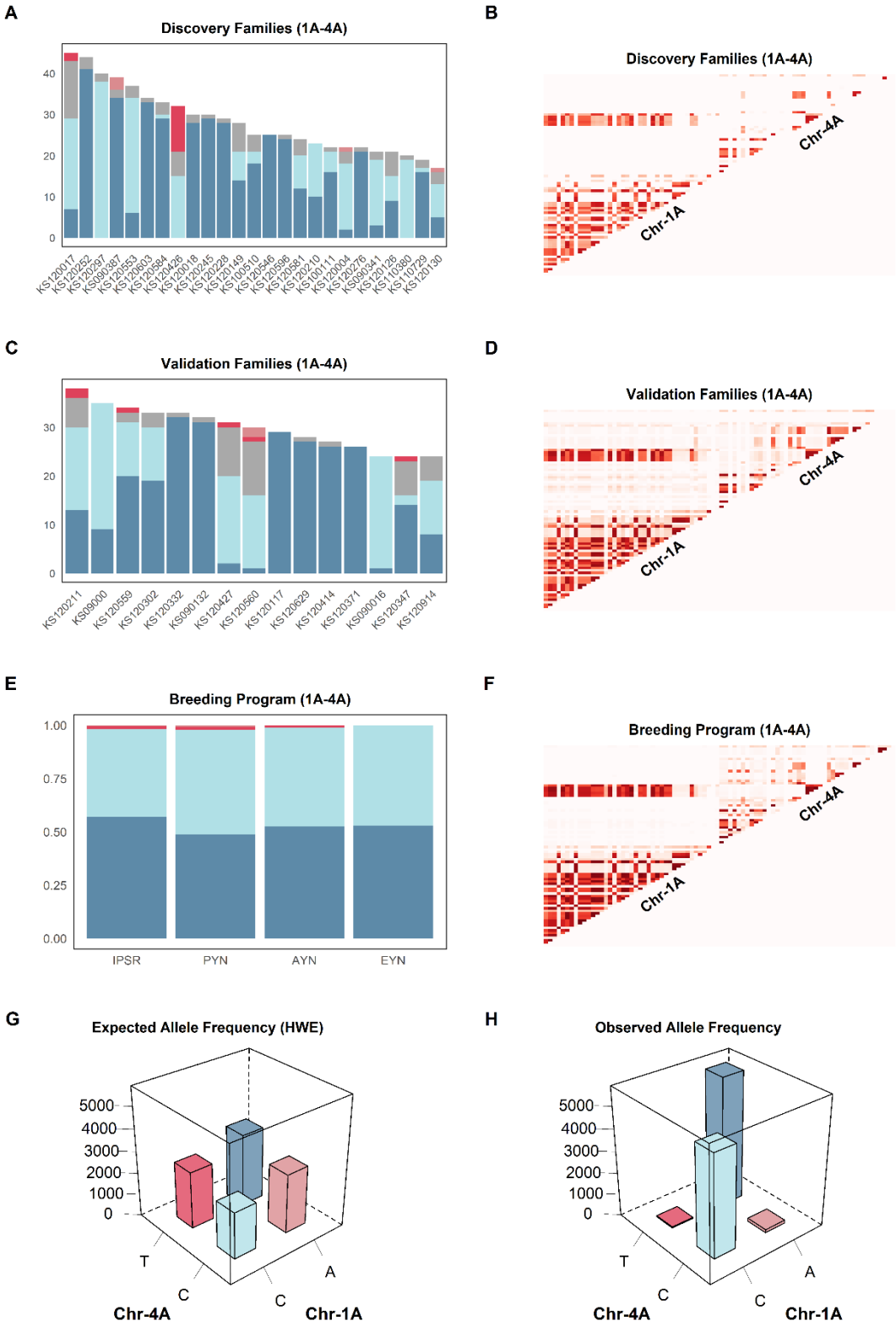


Figure 5. Testing independent breeding panels for sign epistasis. Allele frequency distribution (left column) and heatmaps of 2D LD plots (right column) of the candidate epistatic interaction involving segments of chromosomes 1A and 4A. **A**, **C** and **E** display the allele frequency distribution from the discovery and validation panels, and from the last four stages of the breeding pipeline, respectively. **B**, **D** and **F** shows the 2D LD plot from discovery and validation

263 panels, and from the last four stages of the breeding pipeline, respectively. G displays the
264 expected allele frequency distribution in the entire breeding program under Hardy-Weinberg
265 Equilibrium and H shows the observed allele frequency distribution.
266
267 *Allele frequencies disentangle epistasis from additive selection*
268 The allele frequencies of candidate interaction 1A-4A in the discovery stage reveals many
269 families nearly fixed for the A/T and C/C allelic combinations (Figure 5A). Families segregating for
270 these two alleles exhibited a high frequency of opposed allelic combinations, suggesting that
271 selection is actively favoring both the A/T and C/C allelic combinations. Off-combinations
272 consistently presented very low frequency in most families. The absence of a predominant allelic
273 combination in high frequency among segregating populations weakens the competing
274 hypothesis of additive selection.
275 The candidate interactions 2B-5B, 3B-3D, and 6B-6Da had a similar distribution of
276 opposing allelic combinations as interaction 1A-4A (Supplemental Figures 5:7-A). Candidate
277 interactions 2D-4A, 2A-2D, 2D-3B, 3B-4A and 5A-7A had a common pattern where one of the
278 contrasting combinations predominated in most families, followed by the opposed combination
279 as the second most frequent, and followed by heterozygotes (Supplemental Figures 1:4-A, 14-A).
280 However, for interactions 1B-6B, 5D-7B, 3A-3D, 1A-1B, 5A-6B, 6A-7A, 1BL-1D, 1BS-1D, 2B-7A and
281 6BL-6D, the most common allelic combination was found at a very high frequency in most
282 families, and the opposing combination was observed in only a few families (Supplemental Figures
283 8:13-A, 15:18-A). In all cases, the off-combinations were observed at very low frequencies or
284 absent in many families.
285 The allele frequencies of the candidate three-way epistatic interaction indicates the
286 predominance of the C/C/C alleles, which was fixed in many families (Supplemental Figure 19-A).
287 In segregating families, the opposing allelic combination (A/T/T) was found in high frequency,
288 suggesting that both combinations could be under positive selection. Off-combinations were less
289 common than heterozygotes in most families.
290
291 *Validation Families Support Sign Epistasis Hypothesis*
292 The patterns of LD in the validation families remained consistent with the discovery
293 families for interaction 1A-4A, where a long linkage block on chromosome 1A and a small block
294 on 4A created the characteristic "interchromosomal linkage block" (Figure 5-D). The frequency of
295 the allelic combinations in the validation families presented a similar pattern as the discovery
296 families (Figure 5-C). Six populations are fixed for the A/T combination, and the remaining 11 are
297 segregating with both A/T and C/C as the most frequent combinations. The frequency of off-

298 combinations was also very low.

299 Interaction 2B-5B (Figure 5-C) had a similar pattern of allelic combination as interaction
300 1A-4A, where the validation families had an allele frequency pattern similar to the discovery
301 families. The interchromosomal linkage analysis for interactions 1B-6B, and 5D-7B did not present
302 the characteristic "interchromosomal LD block" (Supplemental Figure 8-D and 9-D) because most
303 families were fixed for the major allelic combination (Supplemental Figure 8-C and 9-D).

304 On the allele frequency analysis, the candidate interactions 5A-7A, 1BL-1D, 1BS-1D, 2B-
305 7A and 6B-6Db (Supplemental Figures 14:18-C) had at least one family with an off-combination
306 as the most frequent allelic combination. This suggests that these loci are not under selection,
307 and the sign epistasis hypothesis was rejected in these candidate interactions. For interaction 6B-
308 6D*a*, family KS120427 had an off-combination as the most frequent allelic combination. However,
309 the abnormal high frequency of heterozygotes, which could have a biological meaning or simply
310 be the consequence of genotyping errors, prevents drawing any definitive conclusions about the
311 interaction (Supplemental Figure 7-C). Similarly, for interaction 3B-3D, the family KS090132 was
312 fixed for an off-combination but the remaining families presented allelic frequencies matching
313 the sign epistasis hypothesis (Supplemental Figure 6-C). It is probable that the parental lines of
314 population KS090132 were fixed for the off-combination, which prevents observing the favorable
315 candidate combinations. Thus, this interaction was kept for further analysis.

316 For the remaining interactions, a consistent pattern emerged where one allelic
317 combination predominated in the majority of families, either fixed or at a very high frequency,
318 while the opposing allelic combination was observed in a few other families at varying frequencies
319 (Supplemental Figure 1:4-C, 10:13-C). The candidate three-way interaction was segregating in few
320 populations, with the allelic combination C/C/C as the most frequent, followed by the opposed
321 allelic combination (A/T/T) and a small occurrence of off-combinations (Supplemental Figure 19-
322 C). The remaining families were fixed for the C/C/C allelic combination.

323

324 *Breeding Program Validation*

325 To further test if the 14 non-falsified interactions are not the result of strong genetic drift
326 effects in small population sizes, data from 10 years of the breeding program was used to
327 calculate the frequency of allelic combinations. The results for interaction 1A-4A (Figure 5-E)
328 reveal that off-combinations were present at low frequencies in the early stages of yield trials
329 (IPSR). The frequency of off-combination declined with successive generations of selection,
330 ultimately leading to complete elimination of off-combinations in the EYN.

331 The same pattern was observed for interactions 6B-6Da, 5A-6B, and 6A-7A (Supplemental
332 Fig 7-E, 12-E, and 13-E), which exhibited no or very few off-combinations at the EYN stage.

333 However, the latter two candidate interactions had the opposite combination with less than 10%
334 of the total allelic combinations, suggesting that the detection of these interactions may have
335 arisen due to genetic drift. Consequently, the sign epistasis hypothesis for the 5A-6B and 6A-7A
336 interactions was rejected. Similar allele frequencies were observed for interactions 3A-3D and
337 1A-1B, leading to the rejection of these combinations as well (Supplemental Figure 10-E and 11-
338 E). Interactions 1B-6B and 5D-7B presented one off-combination with similar frequency than the
339 favored, opposed allelic combination and were also discarded (Supplemental Figure 8-E and 9-E).
340 The six remaining interactions (2A-2D, 2D-4A, 2D-3B, 3B-4A, 2B-5B, and 3B-3D) displayed
341 a high frequency of both opposing allelic combinations but showed minor variations in off-
342 combinations across the breeding stages (Supplemental Figures 1:6-E) and were not falsified due
343 to allele frequency. However, interactions 2D-3B, 3B-4A, 2B-5B, 3B-3D, and 6B-6Da had at least
344 one of the two interacting regions involving a single marker, where a single marker alignment
345 error in the GBS pipeline could result in the observed LD and allele frequencies. For such reason,
346 these interactions were discarded. There was not enough evidence to discard interactions 2D-4A
347 and 2A-2D.
348 Interactions 3B-3D, 2B-5B, and 2A-2D had an average of 94% of genotypes with the two
349 favorable contrasting allelic combinations, at an average ratio of 4:1 between the most common
350 allelic combination and its opposed combination. The interactions 2D-3B, 2D-4A, and 3B-4A
351 comprised the candidate three-way sign epistasis interaction and had the two favorable
352 interactions in very high frequencies across the four breeding stages (Figure 19-E). The C/C/C
353 combination added up to almost 88% of the genotypes across the breeding stages while the A/T/T
354 represented a little over 10%. The remaining 6 off-combinations summed up to less than 2% of
355 the interactions. Although the allele frequencies were consistent with expectations under an
356 epistatic model, the fact that interaction 2D–3B included only a single marker on chromosome
357 segment 3B also led us to reject this potential three-way hypothesis as well.
358
359 Discussion
360 *Random Genetic Drift Alone Does Not Create Strong and Consistent Interchromosomal LD*
361 Strong linkage disequilibrium (LD) between pairs of markers—particularly when
362 physically distant—can arise through random genetic drift, an effect that is more pronounced in
363 small populations (Charlesworth et al., 2003; Charlesworth, 2009; Kliman et al., 2008; Rothammer
364 et al., 2013). In this study, comparisons between theoretical and empirical R^2 distributions
365 showed that families simulated under the sole influence of drift consistently exhibited lower
366 maximum interchromosomal LD values than empirical families, both when considering pairwise
367 interactions present in at least two families (Figure 2) and when pooling all families for analysis

368 (Figure 3). Importantly, even under conditions that favored the chance appearance of strong LD
369 — smaller simulated family size (n = 15) and a fourfold increase in the total number of simulated
370 families— the simulations produced no pairwise interactions that met the thresholds used to
371 identify candidate interactions empirically. These findings indicate that random genetic drift
372 alone is unlikely to create strong and consistent interchromosomal LD patterns across
373 independent families.
374
375 *Strong interchromosomal LD could be created from selection of epistatic interactions*
376 The poor performance of elite lines as parents is hypothesized to be related to epistatic
377 effects. Although estimating epistatic variance is difficult because of non-orthogonality with
378 additive effects (Tessele et al., 2024; Raffo et al., 2022; Vitezica et al., 2017), selection signatures
379 of epistasis are expected to exist in breeding programs. In this study, 25 breeding families were
380 utilized to identify 19 regions in strong interchromosomal LD which are hypothesized to be
381 selection signatures of epistasis. Interchromosomal LD patterns were also observed in tetraploid
382 wheat (Laido et al 2014) and common beans (Rossi et al., 2009; Diniz et al., 2018), but no
383 inferences regarding the genetic mechanism leading to interchromosomal LD were drawn.
384 The allele frequency distribution of candidate interactions resembled that expected
385 under the influence of sign epistasis, which falsifies the additive hypothesis. Similar results were
386 also observed in wild barley, where the cross of plants based on spatial distance resulted in co-
387 adapted gene complexes fixed in locally abundant genotypes (Volis et al., 2011). In two diverse
388 populations of barley, several generations of natural selection resulted in two out of the 16
389 potential combinations of gametes (4^2) as the most frequent, which presented perfectly
390 complementary gametic types (Clegg et al. 1972), suggesting natural selection acted on
391 coadapted (Dobzhansky, 1970) multilocus units.
392 As an attempt of disentangling epistasis from drift, an independent set of 15 validation
393 families was utilized to further test the consistency of the candidate interactions, which falsified
394 seven interactions. As the strength of genetic drift effects is expected to be especially
395 predominant in smaller populations (Charlesworth et al., 2003; Charlesworth, 2009; Kliman et al.,
396 2008; Rothammer et al. 2013), the assembly of allelic combinations resembling selection for sign
397 epistasis in the discovery stage could still have been a consequence of genetic drift. Even though
398 our simulation results suggested otherwise, random genetic drift has been attributed to the
399 emergence of apparent selection signatures between pairs of SNPs in simulation studies (Id-
400 Lahoucine et al., 2019; Vilas et al., 2012).
401 To mitigate the impact of genetic drift, a significantly larger dataset spanning a decade of
402 breeding records was analyzed, leading to the rejection of four interactions. Among the remaining

403 eight interactions, 2A-2D, 2D-3B, 2D-4A, and 3B-4A had a single allelic combination representing
404 close to 90% of the total allelic combinations, contradicting the assumption of comparable
405 adaptability between opposing allelic combinations (Wight, 1932). However, this observed
406 phenomenon may relate to founder effects (James, 1971), where a restriction in allele frequency
407 within the founder germplasm results in the spread of one of allelic combination in the breeding
408 program. In that sense, the balanced frequency of opposed allelic combinations in interactions
409 1A-4A, 6B-6D, 3B-3D, and 2B-5B within an elite breeding program is surprising, considering that
410 recycling lines and extensive utilization of good parental lines is expected to favor one epistatic
411 combination over generations. For this reason, it is expected that the number and extent of sign
412 epistatic interactions should be much larger when considering different breeding programs, as a
413 function of founder effects, aka founder germplasm.
414 Even though the allele frequencies and interchromosomal linkage blocks observed for
415 interactions 2D–3B, 3B–4A, 2B–5B, 3B–3D and 6B-6Da were consistent with expectations under
416 the epistasis hypothesis, these interactions were not considered reliable for the purposes of this
417 study. The polyploid nature of wheat and its large genome (~17 Gb) enriched in repetitive DNA
418 (Alipour et al., 2019) increase the likelihood of alignment errors during the GBS pipeline. Because
419 each of these candidate interactions involved at least one genomic segment represented by a
420 single marker, it is possible that the apparent strong linkage arose from misalignment rather than
421 true biological signal. By contrast, we found no evidence to falsify interactions 1A–4A, 2D–4A, and
422 2D–3B, which therefore remain strong candidates for epistatic selection.
423
424 *Selection Against Unfavorable Epistatic Combinations and Its Implications for Wheat Breeding*
425 The near-complete recovery of opposed allelic combinations as observed for interaction
426 1A-4A suggests artificial selection alone may not be the primary driver of distorted allele
427 frequencies. Instead, it is possible that natural selection is responsible for recovering favorable
428 allelic combinations, as observed in barley and wild barley populations (Volis et al., 2011; Clegg
429 et al. 1972), through a nearly lethal response to off-combinations, similar to hybrid necrosis in
430 wheat (Tsunewaki, 1960) or through embryo lethality, as reported in wheat-rye hybrids (Tikhenko
431 et al., 2005; Tikhenko et al., 2010; Tikhenko et al., 2017). Anecdotal evidence on this hypothesis
432 has been observed in single seed descent families in the Kansas State wheat breeding program.
433 In this speed breeding pipeline, the quantity of seed recovered after three generations of
434 inbreeding is often significantly lower than expected based solely on poor levels of germination.
435 This could be an indication that a varying numbers of epistatic interactions, as interactions 1A-4A,
436 6B-6D, and 3B-3D, segregate in the families and plants homozygotes to off-combinations may
437 experience a severe setback on performance which eventually leads to premature death or

438 reproductive incapability. Populations with a higher number of segregating epistatic interactions,
439 if additionally segregating to interactions 2B-5B, 2A-2D, and 2D-3B, for instance, may
440 consequently end up with fewer viable plants.

441 A practical parallel can be drawn with the well-known *Rht-1* genes in wheat, which confer
442 reduced responsiveness to gibberellic acid and lead to a semi-dwarf phenotype (Gale and
443 Marshall, 1973; Pinthus et al., 1989). Although most modern cultivars carry either *Rht-B1b* or *Rht-*
444 *D1b* (Evans, 1998), the combination of both alleles results in a dwarf phenotype that is highly
445 undesirable. Similarly, when breeding programs combine genotypes that achieve a target
446 phenotype through distinct genetic mechanisms, unfavorable allelic interactions can emerge,
447 effectively reducing the viable population size and constraining breeding progress, and reduce
448 the rate of genetic gain in the long term (Fall et al., 2025).

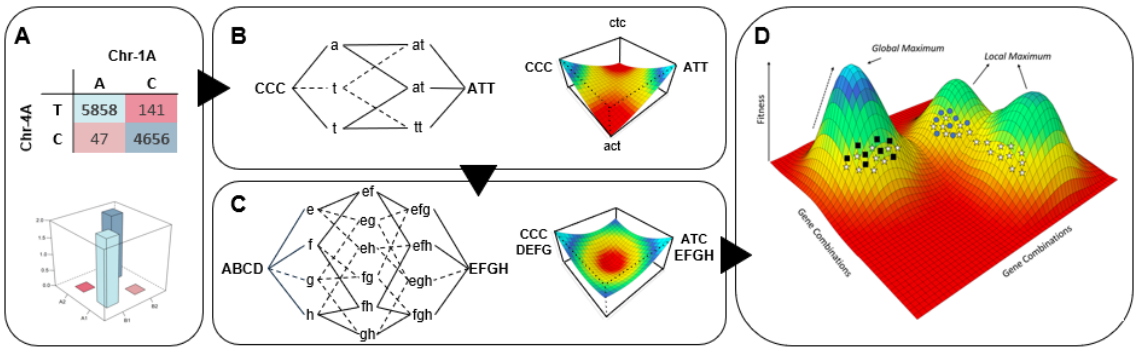
449 One alternative to overcome this challenge is to constrain the genetic diversity of the
450 breeding program with the fixation of one of the interacting alleles, converting epistatic variance
451 into additive variance (Hill et al., 2008; Technow et al., 2021). In the candidate interaction 1A-4A,
452 fixing the A allele in chromosome 1A causes the allele in chromosome 4A to behave additively,
453 where the T allele is always positive, and C is always negative (Figure 5 A, C, and E). This ensures
454 that the selection pressure is constant for the alleles in the interacting locus and increases the
455 probability of having one favorable combination fixed.

456

457 *Sign Epistasis Creates a Rugged Surface on the Fitness Landscape*

458 The possible existence of di-genic sign epistasis, as observed in interaction 1A-4A (Figure
459 5A), implies that transitioning from one superior combination to another requires two allelic
460 substitutions, passing through a transitional state of inferior performance (Wright, 1988; Wade,
461 2002). The fitness landscape of the potential tri-genic sign epistasis interaction reported in this
462 study has a valley between the two optimum allelic combinations (Figure 6B). And, if extended to
463 higher order epistasis and more gene interactions (Figure 6C), it indicates that different
464 combinations of alleles can present similar levels of adaptation to the same environmental
465 conditions (Wright, 1988). The entire field of allelic combinations and respective levels of
466 adaptation can be greatly simplified through the fitness landscape metaphor (Figure 6D), as
467 proposed in the shifting balance theory (Wright, 1932), where hills and valleys represent high and
468 low levels of adaptation, respectively. The fitness landscape, as a metaphor, should not be taken
469 literally (Wade, 2002), as it is inadequate to represent higher order epistatic interactions (Wright,
470 1932). It does, however, provides a powerful simplified representation of complex genetic
471 systems and how populations navigate the rugged surface during evolution. The concept of
472 evolution under the shifting balance theory has striking similarities with the structure of breeding

473 programs (Technow et a., 2021) and could be leveraged to better understand epistasis in the face
474 of managing breeding germplasm pools.
475



476
477 **Figure 6. Rugged fitness landscapes are formed from sign epistasis interactions. A** The allele
478 frequency of interaction 1A-4A across the breeding program indicates a di-genic sign epistasis
479 interaction. **B** The potential three-gene sign epistasis interaction involving chromosomes 2D, 3B
480 and 4A creates a valley of adaptation between the two optimum allelic combinations. **C**
481 Combining a hypothetical four- and the potential three-gene sign epistasis interaction creates a
482 fitness landscape with regions of high, intermediate, and low adaptation. Factual representation
483 of the fitness landscape is difficult because of the elevated complexity of the genetic system. **D**
484 Rugged fitness landscape formed by several sign epistasis interactions with variable levels of gene
485 interactions. This is an extreme simplification of a highly complex, multi-dimensional genetic
486 system. Adaptations from Wight (1932).

487

488 *Breeding Strategies on a Rugged Fitness Landscape*

489 A rugged surface on the fitness landscape has implications for breeding programs as the
490 strategy used to manage the germplasm pool can lead to higher or lower levels of epistatic
491 variance in breeding families. The operation of a closed breeding program inevitably reduces the
492 genetic diversity in the germplasm pool and, under the influence of drift and selection, leads to
493 the fixation of alleles involved in sign epistatic interactions. This process converts epistatic
494 variance into additive variance and increases the potential of adaptation of the breeding program
495 (populations) to local environments (Wade and Goodnight, 1998), as less complex genetic
496 systems have higher heritability and respond better to selection. This can be visualized in Figure
497 6D as the *black squares* or *blue circles* genotypes, located in independent hills and assumed to
498 comprise separate breeding programs. Alternatively, operating a breeding program with a large
499 genetic base may reintroduce epistatic variance when recycling lines from within the breeding
500 program, which can be depicted as the *white star* genotypes in Figure 6D. The cross of parental

501 lines that belong to different adaptive peaks leads to a phenomenon called F_2 breakdown
502 (Whitlock et al., 1995), where the mean performance of the population is lower than the
503 midparent value. This is a consequence of disrupting, in the offspring, the contrasting epistatic
504 interactions that were homozygous in each parental line. Breeding families with high levels of
505 epistatic variance present a low correlation between early and late generations (Upadhyaya and
506 Nigam, 1998; Humphrey et al., 1969), indicating a limited response to early generation selection.
507 If late generation selection is adopted, the quick pace of inbreeding and the commonly small
508 population size of breeding families causes both favorable and unfavorable combinations to be
509 fixed, reducing the probability of assembling optimal allelic combinations.

510 One common feature of breeding programs is the exchange of germplasm, where elite
511 genotypes are anticipated to transmit superior allelic combinations from one breeding pool to
512 another (Wright, 1932; Technow et al., 2021). The success of introducing germplasm on a rugged
513 fitness landscape scenario depends on having both parental lines on the same adaptive peak, that
514 is, sharing the same founding allelic combinations, otherwise, F_2 breakdown may also be observed
515 in breeding families. For navigating on multiple adaptive peaks, broad genetic base programs may
516 be more suitable for introducing external germplasm as there is more than one set of founding
517 allelic combinations that could match the germplasm to be introduced. Although normally used
518 as a strategy to introduce a few alleles in the background of a recurrent line (Hospital, 2005),
519 backcrossing can be leveraged by closed breeding programs to recover the epistatic framework
520 of the recurrent line. Backcrossing constraints epistatic variance, increasing additive variance and
521 the response to selection in breeding populations. For a comprehensive discussion of the
522 implications of sign epistasis from the perspective of the fitness landscape metaphor, view
523 Tessele et al. (2024) and Holland (2001).

524

525 ## Conclusion

526 This study investigated whether selection signatures consistent with sign epistasis exist
527 within wheat breeding families, aiming to expand the understanding of how epistasis may
528 underlie the occasional poor performance of elite lines as parents. A genome-wide analysis
529 detected 19 candidate interchromosomal interactions with abnormally strong linkage
530 disequilibrium whose allelic patterns were consistent with sign epistasis and refuted the
531 competing additive hypothesis. To avoid circular inference and minimize bias from genetic drift,
532 two independent validation sets were analyzed. Eleven interactions were falsified, and five were
533 excluded because they involved a single marker on one of the interacting chromosomes, where
534 strong LD and allele frequency distortions may have arisen from genotyping-by-sequencing (GBS)
535 alignment artifacts. There was not sufficient evidence to reject the sign epistasis hypothesis for

three candidate interactions with the current dataset and analysis. The possibility existence of sign epistasis could be responsible for a reduction in the effective population size in crosses with contrasting sign epistasis combinations, and for the eventual failure to recycle elite lines in the breeding program as epistatic variance is reestablished in the offspring. The potential presence of a rugged fitness landscape underscores the importance of adequately managing epistatic variance within breeding programs for efficiently developing elite germplasm.

Conflict of Interest

None declared.

Funding

We acknowledge the Kansas Wheat Alliance and the Kansas Wheat Commission for their essential financial support, which made this research possible. G.P.M.'s contribution was supported by National Science Foundation Enabling Discovery through GEnomics (EDGE) award #2421208.

Data Availability

Marker information of discovery and validation populations in HapMap format is included in files 'Discovery-Populations.hmp.txt' and 'Validation-Populations.hmp.txt'. The code to run the PCA analysis to eliminate outliers within families is found in 'PCA.R'. The code used to calculate linkage disequilibrium is found in 'C-Linkage_Disequilibrium_Loops.bach'. The code used to calculate the 2D-LD plots in discovery and validation families and the control plot is found in the `*_2D_LDplot.R' files. The 'Circleze.R' file has the code used to plot interchromosomal LD with chromosomes displayed circularly. The file ‘Simulation.R’ contains the R code used to simulate families under the sole influence of drift. Supplemental material is available at Figshare. The raw genotypic data, available at Figshare, will be deposited at NCBI if/upon the manuscript is accepted for publication.

References

Acosta-Pech, R., Crossa, J., de Los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., & Pérez-Rodríguez, P. (2017). Genomic models with genotype× environment interaction for predicting hybrid performance: an application in maize hybrids. Theoretical and Applied Genetics, 130, 1431-1440.

Alipour, H., Bai, G., Zhang, G., Bihamta, M. R., Mohammadi, V., & Peyghambari, S. A. (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. *PLoS One*, *14*(1), e0208614.

Alvarez-Castro JM, Carlborg O (2007) A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. Genetics 176:1151–1167.

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. Nature Reviews Genetics, 4(2), 99-110.

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M. C., Baenziger, S. P., Matnyazov, R., ... & Akhunov, E. D. (2010). Population-and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.). BMC genomics, 11(1), 1-17.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics, 10(3), 195-205.

Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology, Evolution, and Systematics, 34(1), 99-125.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetics Research, 70(2), 155-174.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. PLoS genetics, 2(4), e64.

Clegg, M. T., Allard, R. W., & Kahler, A. L. (1972). Is the gene the unit of selection? Evidence from two experimental plant populations. Proceedings of the National Academy of Sciences, 69(9), 2474-2478.

Cooper, M., Podlich, D. W., Micallef, K. P., Smith, O. S., Jensen, N. M., Chapman, S. C., & Kruger, N. L. (2002). Complexity, quantitative traits and plant breeding: a role for simulation modelling in the genetic improvement of crops. In Quantitative genetics, genomics and plant breeding (pp. 143-166). Wallingford UK: CABI Publishing.

Diniz, A. L., Giordani, W., Costa, Z. P., Margarido, G. R., Perseguini, J. M. K., Benchimol-Reis, L. L., Chiorato, A.F., Garcia, A.A.F., & Vieira, M. L. C. (2018). Evidence for strong kinship

influence on the extent of linkage disequilibrium in cultivated common beans. Genes, 10(1), 5.

Dobzhansky, T. (1970). Genetics of the evolutionary process (Vol. 139). Columbia University Press. James, J. W. (1970). The founder effect and response to artificial selection. Genetics Research, 16(3), 241-250.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The plant genome, 4(3).

Evans, L. T. (1998). *Feeding the ten billion: plants and population growth*. Cambridge University Press.

Fall, S. T., Kena, A., Rice, B. R., Kanfany, G., Diatta, C., Kane, N. A., ... & Morris, G. P. (2025). Genomic approaches to build de novo elite breeding gene pools from locally-adapted landraces. *bioRxiv*, 2025-06.

Faux, A. M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., ... & Hickey, J. M. (2016). AlphaSim: software for breeding program simulation. *The plant genome*, *9*(3), plantgenome2016-02.

Federer, W. T., & Raghavarao, D. (1975). On augmented designs. Biometrics, 29-35.

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants. Annual review of plant biology, 54(1), 357-374.

Forsberg, S. K., & Carlborg, Ö. (2017). On the relationship between epistasis and genetic variance heterogeneity. Journal of experimental botany, 68(20), 5431-5438.

Gale, M. D., & Marshall, G. A. (1973). Insensitivity to gibberellin in dwarf wheats. *Annals of Botany*, *37*(4), 729-735.

Gu, Z., Gu, M. Z., & GlobalOptions, I. (2022). Package 'circlize'.

He, S., Reif, J. C., Korzun, V., Bothe, R., Ebmeyer, E., & Jiang, Y. (2017). Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. Theoretical and Applied Genetics, 130, 635-647.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. PLoS genetics, 4(2), e1000008.

Holland, J. B. (2010). Epistasis and plant breeding. *Plant Breeding Reviews: John Wiley & Sons, Inc*, 27-92.

Hudson, R. R., & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. Genetics, 120(3), 831-840.

Id-Lahoucine, S., Molina, A., Cánovas, A., & Casellas, J. (2019). Screening for epistatic selection signatures: A simulation study. Scientific reports, 9(1), 1-5.

James, J. W. (1970). The founder effect and response to artificial selection. Genetics Research, 16(3), 241-250.

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., ... & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype× environment interactions in Kansas wheat. The plant genome, 10(2), plantgenome2016-12.

Joshi, R., Meuwissen, T. H., Woolliams, J. A., & Gjøen, H. M. (2020). Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia. Genetics Selection Evolution, 52(1), 1-13.

Joukhadar, R., Daetwyler, H. D., Gendall, A. R., & Hayden, M. J. (2019). Artificial selection causes significant linkage disequilibrium among multiple unlinked genes in Australian wheat. Evolutionary Applications, 12(8), 1610-1625.

Kassambara, A., & Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses, 76(2).

Kliman, R., B. Sheehy & J. Schultz (2008) Genetic drift and effective population size. Nature Education 1(3).

Koch, E., Ristroph, M., & Kirkpatrick, M. (2013). Long range linkage disequilibrium across the human genome. PloS one, 8(12), e80754.

Kreitman, M., & Di Rienzo, A. (2004). Balancing claims for balancing selection. TRENDS in Genetics, 20(7), 300-304.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature neuroscience, 12(5), 535-540.

Kristensen, P. S., Jahoor, A., Andersen, J. R., Cericola, F., Orabi, J., Janss, L. L., & Jensen, J. (2018). Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. Frontiers in plant science, 9, 69.

Lehermeier, C., de Los Campos, G., Wimmer, V., & Schön, C. C. (2017). Genomic variance estimates: With or without disequilibrium covariances?. Journal of Animal Breeding and Genetics, 134(3), 232-241.

Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. Genetics 54, 595–609 (1966).

Lewontin, R. C., & Kojima, K. I. (1960). The evolutionary dynamics of complex polymorphisms. Evolution, 458-472.

Lin, C. S., & Poushinsky, G. R. E. G. (1985). A modified augmented design (type 2) for rectangular plots. Canadian journal of plant science, 65(3), 743-749.

Liu, G., Zhao, Y., Gowda, M., Longin, C. F. H., Reif, J. C., & Mette, M. F. (2016). Predicting hybrid performances for quality traits through genomic-assisted approaches in Central European

wheat. PLoS One, 11(7), e0158635.

Maccaferri, M., Sanguineti, M. C., Noli, E., & Tuberosa, R. (2005). Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Molecular Breeding, 15, 271-290.

Moran, P., & Smith, C. (1918). The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh, 52, 899-438.

Pérez, P., & de los Campos, G. (2014). BGLR: a statistical package for whole genome regression and prediction. Genetics, 198(2), 483-495.

Pinthus, M. J., Gale, M. D., Appleford, N. E., & Lenton, J. R. (1989). Effect of temperature on gibberellin (GA) responsiveness and on endogenous GA1 content of tall and dwarf wheat genotypes. *Plant Physiology*, *90*(3), 854-859.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), 559-575.

Raffo, M. A., Sarup, P., Guo, X., Liu, H., Andersen, J. R., Orabi, J., Jahoor, A., & Jensen, J. (2022). Improvement of genomic prediction in advanced wheat breeding lines by including additive-by-additive epistasis. Theoretical and Applied Genetics, 135(3), 965-978.

Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., & Papa, R. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. Evolutionary Applications, 2(4), 504-522.

Rothammer, S., Seichter, D., Förster, M., & Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC genomics, 14(1), 1-17.

Sign Epistasis Reduces the Effective Population Size of Breeding Populations.

Sorensen, D., Fernando, R., & Gianola, D. (2001). Inferring the trajectory of genetic variance in the course of artificial selection. Genetical Research, 77, 83–94.

Takahata, N., & Satta, Y. (1998). Footprints of intragenic recombination at HLA loci. Immunogenetics, 47, 430-441.

Technow, F., Podlich, D., & Cooper, M. (2021). Back to the future: implications of genetic complexity for the structure of hybrid breeding programs. G3, 11(7), jkab153.

Tessele, A., González-Diéguez, D. O., Crossa, J., Johnson, B. E., Morris, G. P., & Fritz, A. K. (2025). Improving genomic selection in hexaploid wheat with sub-genome additive and epistatic models. *G3: Genes, Genomes, Genetics*, *15*(4), jkaf031.

Tian, D., Araki, H., Stahl, E., Bergelson, J., & Kreitman, M. (2002). Signature of balancing selection in Arabidopsis. Proceedings of the National Academy of Sciences, 99(17), 11525-11530.

Tikhenko, N. D., Tsvetkova, N. V., & Voylokov, A. V. (2005). Genetic control of embryo lethality in crosses between common wheat and rye. *Russian Journal of Genetics*, *41*(8), 877-884.

Tikhenko, N., Poursarebani, N., Rutten, T., Schnurbusch, T., & Börner, A. (2017). Embryo lethality in wheat-rye hybrids: dosage effect and deletion bin mapping of the responsible wheat locus. Biologia plantarum, 61(2), 342-348.

Tikhenko, N., Tsvetkova, N., Voylokov, A., Dobrovolskaya, O., Zaynali Nezhad, K., Röder, M. S., & Börner, A. (2010). Embryo lethality in wheat× rye hybrids—mode of inheritance and the identification of a complementary gene in wheat. Euphytica, 176(2), 191-198.

Tsunewaki K (1960) Monosomic and conventional analyses in common wheat. III. Lethality. Jpn J Genet 35:71–75

Turelli, M. (2017). Commentary: Fisher's infinitesimal model: A story for the ages. Theoretical population biology, 118, 46-49.

U.S. Department of Agriculture, Agricultural Research Service. n.d. Hard Winter Wheat Regional Nursery Program — Research. USDA ARS. Accessed [Month Day, Year]. https://www.ars.usda.gov/plains-area/lincoln-ne/wheat-sorghum-and-forage-research/docs/hard-winter-wheat-regional-nursery-program/research/

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of dairy science, 91(11), 4414-4423.

Vazquez, M. D., Zemetra, R., Peterson, C. J., Chen, X. M., Heesacker, A., & Mundt, C. C. (2015). Multi-location wheat stripe rust QTL analysis: genetic background and epistatic interactions. Theoretical and Applied Genetics, 128(7), 1307-1318.

Vilas, A., Pérez-Figueroa, A., & Caballero, A. (2012). A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology*, *25*(7), 1364-1376.

Vitezica ZG, Legarra A, Toro MA, Varona L (2017) Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. Genetics 206:1297–1307.

Volis, S., Shulgina, I., Zaretsky, M., & Koren, O. (2011). Epistasis in natural populations of a predominantly selfing plant. Heredity, 106(2), 300-309.

Wade, M. J. (2002). A gene's eye view of epistasis, selection and speciation. Journal of Evolutionary Biology, 15(3), 337-346.

Wade, M. J., & Goodnight, C. J. (1998). Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. Evolution, 52(6), 1537-1553.

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: sign epistasis and genetic costraint on evolutionary trajectories. Evolution, 59(6), 1165-1174.

Whitlock, M. C., Phillips, P. C., Moore, F. B. G., & Tonsor, S. J. (1995). Multiple fitness peaks and epistasis. Annual review of ecology and systematics, 601-629.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wiuf, C., Zhao, K., Innan, H., & Nordborg, M. (2004). The probability and chromosomal extent of trans-specific polymorphism. Genetics, 168(4), 2363-2372.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution.

Wright, S. (1988). Surfaces of selective value revisited. The American Naturalist, 131(1), 115-123.

Zhao, J., Zheng, X., Qiao, L., Yang, C., Wu, B., He, Z., ... & Qi, Z. (2022). Genome-wide association study reveals structural chromosome variations with phenotypic effects in wheat (Triticum aestivum L.). The Plant Journal, 112(6), 1447-1461.