Received: 6 July 2022 | Accepted: 27 April 2023 | Published online: 2 June 2023

DOI: 10.1002/csc2.21018

Crop Science

ORIGINAL ARTICLE

Crop Breeding & Genetics

Reciprocal recurrent selection based on genetic complementation: An efficient way to build heterosis in diploids due to directional dominance

Giovanny Covarrubias-Pazaran[1,2] | **Christian Werner**[1,2] | **Dorcus Gemenet**[1,2]

[1]Excellence in Breeding Platform, Consultative Group of International Agricultural Research, Texcoco, Mexico

[2]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Correspondence
Giovanny Covarrubias-Pazaran, Excellence in Breeding Platform, Consultative Group of International Agricultural Research, Texcoco, Mexico
Email: g.covarrubias@irri.org

Assigned to Associate Editor Paulo Teodoro.

Funding information
Bill and Melinda Gates Foundation, Grant/Award Number: OPP1177070

Abstract

Depending on the trait architecture and reproduction system, selection strategies in plant breeding focus on the accumulation of additive, dominance effects, or both. Innovation in the exploitation of dominance-effect-based heterosis has been limited since the proposal of general combining ability (GCA)-based approaches. We propose the use of a new surrogate of genetic complementation between genetic pools to increase accumulation of dominance effects and heterosis. We simulated breeding programs to show how reciprocal recurrent selection (RRS) by genetic complementation would build the dominance-based heterosis cheaper than GCA-based approaches and used real phenotypic data from hybrid maize (*Zea mays*) to demonstrate the underlying concepts. We found RRS by genetic complementation to be an attractive and viable strategy to exploit dominance, build de novo heterotic pools, and enhance the current GCA-based approaches. If demonstrated in practice, we hypothesized that this approach would lower the cost of hybrid breeding drastically and contribute to food security.

1 | INTRODUCTION

When Shull (1952) referred to the term heterosis as "increased vigor, size, fruitfulness, speed of development, resistance to disease and to insect pests, or to climatic rigors of any kind, manifested by crossbred organisms as compared with corresponding inbreds, as the specific results of unlikeness in the constitutions of the uniting parental gametes" he mostly focused on the positive phenotypic effects but a clear genetic definition was not provided. The difference between crossbred organisms compared with corresponding inbreds occurs because inbreds do not leverage from the dominance interactions, whereas hybrid or non-inbred organisms exploit the immediate advantage of dominance interactions and epistasis. Single-locus dominance is the phenomenon where a heterozygote individual tends to reflect more the phenotype of one of the homozygote individuals, and in polygenic traits

List of abbreviations: BV, breeding value; D, genetic distance; GCA, general combining ability; G×E, genotype by environment; LD, linkage disequilibrium; meanDD, mean dominance degree; QTL, quantitative trait locus; QTN, quantitative trait nucleotide; RRS, reciprocal recurrent selection; RS, recurrent selection; SCA, specific combination ability; SE, standard error; SNP, single nucleotide polymorphism; VarD, dominance variance; VarG, genetic variance.

these multiple dominance interactions can add to a substantial portion of the phenotypic value (Crow, 1999; Falconer & Mackay, 1996). Throughout this paper, we assume dominance (defined as the phenomenon of one variant [allele] of a gene on a chromosome tending to mask or override the effect of a different variant of the same gene on the other copy of the chromosome) as the main contributor to heterosis. When considered at the individual level, heterosis is referred as mid-parent or best-parent heterosis (difference between the hybrid and the average of both parents, or the best parent, respectively) (Falconer & Mackay, 1996), whereas when considered at the population level, heterosis is referred as baseline (difference between the inbred and non-inbred population) and panmictic heterosis (difference between the interpopulation hybrids and the non-inbred populations from each subpopulation) (Lamkey & Edwards, 1999). In the absence of epistasis, the presence of a positive average difference in performance between hybrids and parents implies the presence of directional dominance which led past breeders to develop approaches to exploit it for commercial settings (Birchler et al., 2010; Hallauer et al., 2010; Lamkey & Edwards, 1999). There is overwhelming evidence that heterosis is a manifestation of dominance effects, as opposed to overdominance (see the review by Bingham, 1998). In addition, evidence indicates that heterosis is the opposite effect of inbreeding in which recessive-lethal alleles are unmasked in the phenotype in the absence of epistasis (Bernardo, 2002; Charlesworth & Willis, 2009; Davenport, 1908; East, 1936; Falconer & Mackay, 1996; Joshi et al., 2015; Lamkey & Edwards, 1999; Varona et al., 2018). Since heterosis is the reflect of the accumulation of many genes, dominance × dominance epistatic interactions can contribute but will not be the focus of this paper (Jiang et al., 2017; Lippman & Zamir, 2007).

The idea that genes can have different modes of action—that is, an additive, a dominance, or an epistatic effect on the final phenotype—led scientists like Comstock and Robinson, among others, to develop sophisticated mating designs to understand the gene action of important and complex traits such as grain yield in maize (Bernardo, 2002; Comstock & Robinson et al., 1949, 1952; Jinks & Jones, 1957). Better understanding of the inheritance theory led to the development of recurrent and reciprocal recurrent selection (RRS) as the predominant method to improve hybrid populations, which over the years have been enhanced by additional mating designs and introgression steps to develop products capable of increasing the performance in the agricultural fields in the 20th and 21st century (Hallauer et al., 2010). Although the Green Revolution had an enormous impact in developing countries using major genes (e.g., dwarfing genes) (Evenson & Gollin, 2003; Hedden, 2003), hybrid breeding based on the quantitative genetics theory of dominance was one of the main drivers together with improved agronomic management of the massive yield increases in developed countries in North America and Europe in crops like maize (Hill, 2010). The RRS based on recycling parents using general combining ability (GCA) as a surrogate of value and specific combining ability (SCA) as the additional value to identify products have become the foundation of hybrid breeding until the present (Hallauer et al., 2010). Unfortunately, RRS compared to single-pool inbred breeding (e.g., line breeding) and single-pool non-inbred breeding (e.g., clonally propagated crops) tends to be more expensive in terms of time and money due to the need of additional crossing among subpopulations and evaluation of the resulting hybrids depending on the mating strategy (e.g., testcrossing). In summary, since the proposal of GCA-based approaches, few strategies to harness dominance and epistasis addressing time and resources constraints of GCA-based approaches have been proposed (de Boer & Hoeschele, 1993; Hallauer et al., 2010; Mrode, 2014; Werner et al., 2023).

Here, we want to highlight that the availability of genomic information in the form of genetic markers and our knowledge of the modes of inheritance (especially directional dominance) provide us with a unique opportunity to breed for the accumulation of dominance effects to exploit heterosis in a fast and cheap way to boost RRS approaches that improve populations based on GCA. We have developed a method, a new variant of RRS, which we call breeding dominance by genetic complementation which leverages from the idea that dominance alleles mask deleterious alleles to create pools that highly complement each other and display high levels of dominance-based heterosis in a controlled manner. This can be thought of as a controlled genetic distance method. We show with simulations and some real datasets the implementation of the method and the implications. This could drastically reduce the cost of a breeding program depending on the levels of dominance found in the species of interest and serve to create de novo heterotic pools fast and cheap.

Core Ideas

- Heterotic patterns can be developed quickly through the genetic complementation metric proposed.
- Complementation can produce high-performance heterotic pools at a low cost in diploid species displaying dominance.
- Complementation can enhance general combining ability-based recycling approaches in hybrid breeding.

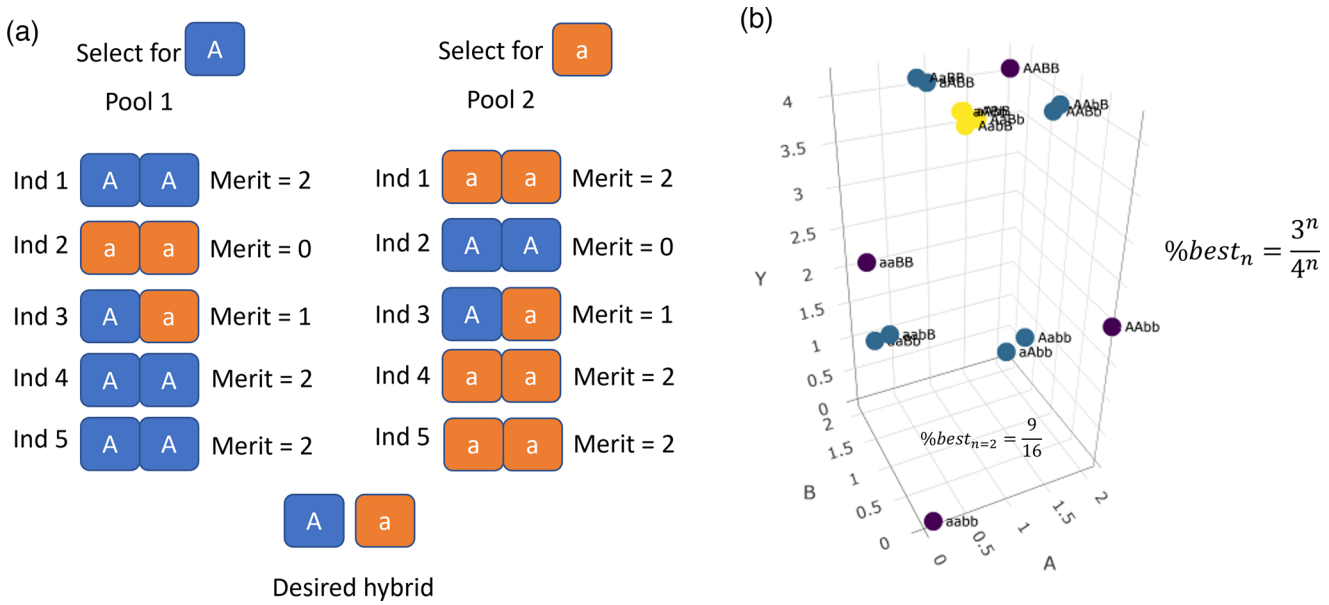

FIGURE 1 Graphical description of the principles behind reciprocal recurrent selection by genetic complementation under equal allele frequencies and biallelic loci. (a) Pool 1 assigns more merit to the selection of one allele whereas the other pool selects for the opposite allele. Both pools are bred to fix the desired allele (or haplotype if reasoning genome wide). Under the directional-dominance model, over time hybrids between these two populations are expected to better complement each other and produce higher-performing hybrids that display strong heterosis. At complete dominance, the best individual is the one that combines all loci at heterozygote state or all the loci at homozygous state for the positive allele. At lower levels of dominance, the individual that combines all loci at heterozygote state is not as good as the individual comprising all the loci at homozygous state for the positive allele. (b) The genetic value (Y) as a function of two quantitative trait loci (QTLs) (A and B) is shown, displaying the relationship to find the best hybrids depending on the number of QTLs under complete dominance ($3^n/4^n$).

2 | MATERIALS AND METHODS

2.1 | The complementation approach and computation of surrogates

Under the complete-dominance-based heterosis hypothesis, the best individuals are those which accumulate (A) all heterozygote interactions or (B) all homozygote interaction for the positive allele, genome wide for the quantitative trait loci (QTLs) underlying a trait (e.g., yield). In a diploid species, that implies accumulating numerous genotypes of the form A_iA_j for multiple QTLs ($i \neq j$) or A_iA_i (considering A_i the positive allele). For the sake of illustrating the concept, under equal allele frequencies in biallelic loci, selection of individuals for having the maximum performance under complete dominance for a trait under the control of "n" QTLs is given by the probability of $3^n/4^n$ (Figure 1). Under complete dominance, there is a probability of finding one individual per million individuals for 50 QTLs. It is accepted that for complex traits there are hundreds if not thousands of underlying QTLs with varying levels of dominance, from no dominance (fully additive) to full dominance. Assuming—only for demonstration purposes—complete dominance, the only practical way to achieve a fully heterozygote individual for hundreds and thousands of QTLs is to create two selection streams, one that selects for fixation of the allele A_i in population 1 and another that selects for allele A_{ij} in population 2 (where $i \neq j$). Although it will require multiple cycles (recurrent selection) to achieve, this remains more affordable than cultivating billions of individuals and attempting to phenotype them accurately to identify the best individual.

Following the complete-dominance assumption for now (we know that complete dominance does not hold for all loci and different levels of dominance and directionality of the dominance exist across the genome), to ensure the creation of that idealized hybrid, we need to ensure the definition of the idealized or desired genotype and haplotype for each of the two subpopulation/pools (in the case of diploids). The foundational step of a RRS approach by genetic complementation is that two populations will be created with the specific purpose to complement each other (Figure 1).

We then need to define first a desired/idealized genotype under the idea of genomic complementation and complete-dominance assumptions (to be a haplotype in its final state when fixation is reached) as the accumulation of homozygous allelic states for all the QTLs behind a trait of interest in a population, and to be opposite/complementary to another population. Notice the statement, under the complete-dominance assumption. There are two challenges to implement this: (1) unknown location of the QTLs behind the trait of interest (ideally, we would only focus on the actual QTLs), and (2) unknown coupling-repulsion phases present in the

populations (this could potentially slow down the recurrent selection using the complementation concept if we try to break many linkages in repulsion phase). To face these challenges, we propose that under the infinitesimal model of complex traits, the use of a genetic-marker chip with thousands of genetic markers is enough to target most QTLs of interest by linkage disequilibrium (challenge 1), and by identifying the highest-frequency alleles in a population (using genetic markers) we can identify what are the alleles in coupling phase or genome-wide population haplotype in order to come up with the desired genotypes for each pool (instead of picking desired alleles at random, which may lead to repulsion-phase linkages in both populations that are difficult to break) (challenge 2).

Assuming a population of individuals genotyped with biallelic markers with genotypes coded as 0, 1, 2 (corresponding to A_iA_i, A_iA_j, A_jA_j) stored in a vector named "x" with dimensions $n \times 1$ for n individuals in a column vector, and with genotype frequencies (f_1, f_2, f_3), the desired/idealized genotype for the ith marker in pool 1 is the genotype with the highest frequency in the population ignoring the heterozygote genotype and its frequency:

$$d_{1_i} = x_{\max(f_1, f_3)}$$

Over the entire genome, the vector d_1 of desired alleles for pool 1 is the column binding of all individual marker $d_{1,i}$'s. The vector d_2 of desired alleles for pool 2 and complementary to pool 1 is just $d_2 = |d_1 - 2|$ where $|\ |$ is the entry-wise absolute value operator. So we capture the opposite alleles to d_1. These vectors represent the ideal genotype for all individuals in their respective pool in the scale 0, 1, 2. We then center these two vectors around zero for further analysis by subtracting the value of 1 to the vectors. Using the highest-frequency allele to capture the desired alleles in pool 1 is not a bullet-proof approach but this aims to capture the alleles from different markers more often found in coupling phase at the population level, and put them in the same pool instead of assigning them at random which would make it more difficult to break haplotypes to accumulate the desired alleles defined in the d vectors.

Once the desired/idealized genotypes in the vectors d_1 and d_2 have been defined, the next step is to define the surrogate of complementation. This will reflect how close a genotype is to the desired/idealized genotype to complement the opposite pool as χ and is calculated as follows:

$$\chi = \left| \frac{M_o + M_d}{2} \right| w$$

where M_o is the matrix of observed genotypes coded as -1, 0, and 1 for the presence of a reference allele, M_d is the matrix of desired genotypes composed by the d vector (of the respective pool) copied and row bound as many times as individuals (rows) in the M_o matrix, so M_o and M_d have the same row and column dimensions. The vector w refers to the weights to be applied to each marker in case we would like to weight by allele frequencies and $|\ |$ is the entry-wise absolute value operator, otherwise it is a vector of ones. The vector denoting the complementation value for each individual to the opposite pool is named χ. The higher the value of χ, the better parent the individual will be for the next generation to complement the opposite pool moving forward. Although the complete-dominance assumption is unrealistic, for any genome-wide average value of dominance degree ≥ 0, this approach can help harness that level of heterosis.

2.2 | Validating the complementation model in a generic simulation and real datasets

Under the proposed genetic model, a collateral effect on hybrid performance based on levels of dominance is that dominance effects can be predicted with reasonable accuracy by the complementation surrogate (χ) in different forms (Bernardo, 1992, 2002). We selected parental lines using two variations of the χ metric to predict the hybrid performance in the current generations and keeping track of the correlation between the modified χ metric and the hybrid performance:

1. The complementation of individual i from population A to the desired haplotype of population B (implying only one side contributes to the prediction, the other side is constant for all individuals).
2. The complementation between individuals i and j from pools A and B, respectively (implying that both sides contribute to the prediction).

We kept track of the correlation of hybrid performance and the complementation surrogate at different levels of dominance and levels of genetic variance across 10 cycles of recurrent selection in the case of the simulations. The appropriate correlation is expected to be calculated between the dominance effect and the complementation surrogate, but the known difficulties to separate dominance from additive effects motivated us to use the total hybrid performance instead of trying to calculate/separate only the dominance effects which is expected to lower the expected correlation. The initial heterotic pools were created from a single population randomly split into two pools. A summary of simulation features for the genome and phenotypes can be found in Table 1.

Then, we took the hybrid-maize (*Zea mays*) dataset from Kadam et al. (2016) which includes marker data and yield performance for 312 hybrids coming from two heterotic pools (46 lines in the female pool and 172 lines in the male

TABLE 1 Summary of simulation features for the genome and phenotypes.

Simulation features			
Burn-in		Generic genetic model	Inbred and inbred-hybrid programs
	Genome sequence	100,000 generations of evolution	100,000 generations of evolution
		1 chromosome pair	18 chromosome pairs
		1.43 Morgans per chromosome	1.43 Morgans per chromosome
		8×10^8 base pairs per chromosome	8×10^8 base pairs per chromosome
		2×10^{-9} mutation rate	2×10^{-9} mutation rate
	Founder genotypes	100 inbred founders per pool	24 inbred founders per pool
		1, 10, 50, 100, and 1000 QTNs (additive + dominance effects)	1000 QTNs (additive+ dominance effects)
		Normally distributed QTN effects with meanDD values of 0, 0.5, and 0.95 and varDD = 0.2	Normally distributed QTN effects with meanDD value of 0.5 and varDD = 0.2
		0 cycles of breeding	20 years of breeding
		Inbred individuals	Inbred individuals
		Conventional breeding	Conventional line and hybrid breeding
Evaluation	Future breeding	20 cycles of breeding	40 years of breeding
		Testing alternative allocation of resources	Testing alternative allocation of resources
		Equal cost programs	Variable cost programs
		Conventional breeding and alternative treatments	Conventional breeding and alternative treatments

pool) tested in 5environments and calculated the one-to-one complementation metrics χ_{ii} and χ_{ip} (individual-to-individual and individual-to-population respectively), calculating the correlation between these metrics with the hybrid yield performance in each of the 5 environments. In addition, we took the maize dataset made available by Technow et al. (2014) that includes genetic-marker data and adjusted means across environments data for 1254 hybrids coming from crosses between the flint (86 lines) and dent (123 lines) maize pools and 35,478 SNP markers, calculating the one-to-one complementation metrics χ_{ii} and χ_{ip} and the correlation between these metrics with the hybrid yield performance across environments.

2.3 | Simulating the effect of updating testers to build the dominance

Using the previous simulation and the baseline treatment of RRS based on GCA (TWO_POOL_GCA_PHENO), we tested different strategies to update the testers to understand the role of testers in the accumulation of dominance in the classical GCA approach. We simulated the following treatments: RRS without updating the testers (TWO_POOL_GCA_UPDATE0), and RRS updating all the testers every 1, 4, and 8 cycles (TWO_POOL_GCA_UPDATE_x, with $x = \{1, 4, 8\}$), and as control treatments: RRS using true GCA updating all the testers every cycle (TWO_POOL_GCA_TRUE (positive control), and RRS selecting parents at random (NEGATIVE CONTROL). Four levels of dominance were considered (meanDD equal to 0, 0.2, 0.5, and 0.9). Additive, dominance, and total-genetic value were recorded as in the previous simulation. All testers were selected based on their GCA value. The initial heterotic pools were created from a single population randomly split into two pools. A summary of simulation features for the genome and phenotypes can be found in Table 1 and breeding schemes can be found in Figure 2.

2.4 | Simulating the implementation of the complementation approach in a running inbred program

To understand how a program that currently improves inbred lines would transition into a hybrid program under either the conventional strategy recycling based on GCA or the proposed complementation strategy to build initial pools and then move to recycle based on GCA, we simulated the same treatments as before but for the genome structure of a crop similar to maize, and a strategy that reflects how a program working on recycling and delivering inbreds would transition to the complementation approach. The initial heterotic pools were created from a single population randomly split into two pools. Specific details of the simulations like burn-in, treatments, and quantitative genetic models behind can be found in File S1.

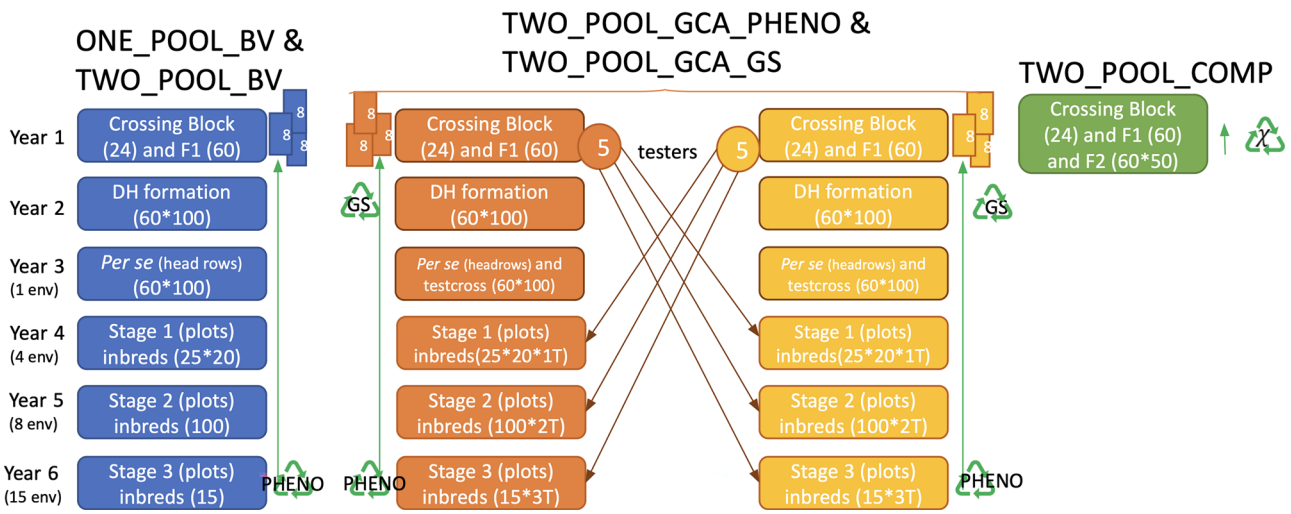

FIGURE 2 Summary of treatments (breeding strategies) compared to build dominance-based heterosis. Each treatment uses a different surrogate of merit to select parents in each pool.

2.5 | Statistical analysis

Stochastic simulations executed with AlphaSimR included 20 Monte-Carlo replications (Chen et al., 2009; Gaynor et al., 2021). For the exploration of the practical implementation of genetic complementation in inbred and hybrid programs, standard errors (SE) were computed for treatments across years as:

$$SE = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation from the 20 Monte-Carlo replicates and n is the number of replicates (20). Standard errors were plotted as shadowed lines for the genetic trends in all figures to declare differences between treatments.

3 | RESULTS

3.1 | Validating the complementation model in a generic simulation and real datasets

Given the unavailability of data from programs that have executed the complementation approach for several cycles as proposed, we had to incorporate a set of the side results from the complementation approach. Under the proposed genetic model, a collateral effect is that the hybrid performance can also be predicted with more or less accuracy by the complementation surrogate (χ) depending on the levels of dominance (Table 2). In theory, the complementation surrogate (χ) is highly correlated with the dominance effect observed in hybrids, but we had to focus on hybrid performance given the difficulty to estimate properly or orthogonally the dominance effects (Nishio & Satoh, 2014).

We kept track of the correlation of hybrid performance and the complementation surrogates at different levels of dominance and levels of genetic variance across 10 cycles in the simulation of a generic breeding program (Table 2). The χ metric is a value for the line merit instead of hybrid performance, so to produce a χ metric for the hybrids themselves we developed the individual-to-population χ_{ip} and individual-to-individual χ_{ii} metrics for a cross. We found that at intermediate and high levels of dominance (meanDD of 0.5 and 0.95, respectively) the correlation between the complementation surrogate and the total-genetic value of the hybrids were intermediate to high (Table 2). At no dominance (meanDD = 0 or purely additive trait) the correlations were around zero at any cycle of selection (Table 2). At intermediate levels of dominance (meanDD = 0.5) the levels of correlation between hybrid performance and χ_{ip} ranged from 0.4 to 0.6 in the first 7 cycles of selection, decreasing at a constant rate as dominance variance gets depleted (Table 2). For almost complete dominance (meanDD = 0.95) the levels of correlation between hybrid performance and χ_{ip} ranged from 0.4 to 0.8 in the first 7 cycles of selection, decreasing at a constant rate as dominance variance is depleted (Table 2). The higher correlation observed is between the complementation surrogate and the dominance effects.

Using the hybrid-maize dataset from Kadam et al. (2016), which includes marker data and yield performance for 312 hybrids coming from two heterotic pools (46 lines in the female pool and 172 lines in the male pool tested in five environments), we calculated the one-to-one complementation metrics χ_{ip} and χ_{ii}, and the correlation between these metrics

TABLE 2 Genetic correlation and VarD/VarG between the complementation surrogate χ_{ip} (individual-to-population complementation) and total-genetic (additive + dominance) value of the hybrid performance at different levels of dominance degree in a simulated program. Genetic correlation is shown in cells and VarD/VarG is shown inside parenthesis.

meanDD/cycle	1	4	7	10
0	−0.06 (0.12)	−0.03 (0.18)	−0.07 (0.14)	0.18 (0.26)
0.5	0.62 (0.84)	0.58 (0.55)	0.43 (0.43)	0.06 (0.4)
0.95	0.77 (0.89)	0.83 (0.86)	0.44 (0.81)	0.3 (0.7)

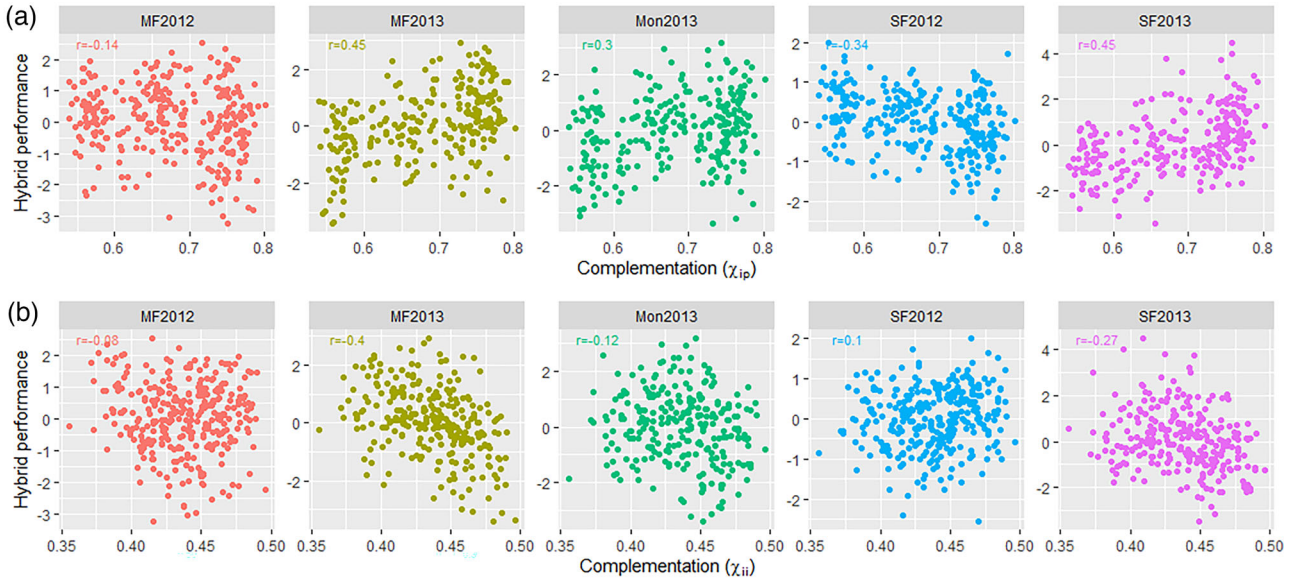

FIGURE 3 Genetic correlation between the complementation surrogates (individual-to-population and individual-to-individual genetic complementation) and hybrid performance at five different environments. Columns represent the five different environments where hybrids were tested. (a) The individual- to-population genetic complementation, (b) the individual-to-individual genetic complementation. Correlation legend is shown in the upper left corner of each plot. Data comes from Kadam et al. (2016) and comprises a real dataset of hybrids between the dent and flint heterotic pools.

with the hybrid yield performance in each of the five environments (Figure 3). We found the correlations between hybrid performance and χ_{ii} (individual-to-individual complementation) to range between −0.2 and 0.1 (Figure 3a), whereas the correlations between hybrid performance and χ_{ip} (individual-to-population complementation) ranged between −0.3 and 0.45, being more consistent and oscillating between positive and negative depending on the average performance for the environment (Figure 3b). The direction of the correlation changed but was just reflecting the GxE found in the dataset since the complementation metric is the same for any environment since it is based purely on markers. In addition, using the maize dataset made available by Technow et al. (2014), the correlation between hybrid performance and χ_{ii} to be 0.003, whereas the correlation between hybrid performance and χ_{ip} was 0.41 (Figure S1).

3.2 | The influence of dominance degree and number of QTLs in total-genetic value

To increase our understanding of how the complementation χ metric can increase genetic gain for dominance, we expanded the simulation not only to different levels of dominance (no dominance to complete dominance), but also to different trait complexities (i.e., varying levels of quantitative trait nucleotides [QTNs] behind the trait) and compared it to the classical GCA approach (Figure 2). When looking exclusively at dominance gain we found that, independently of the number of QTLs behind the trait, the complementation (TWO_POOL_COMP and TWO_POOL_COMP_TO_GCA) and the RRS strategies (TWO_POOL_GCA_PHENO and TWO_POOL_GCA_TRUE) strategies were able to build dominance effects efficiently, whereas the negative

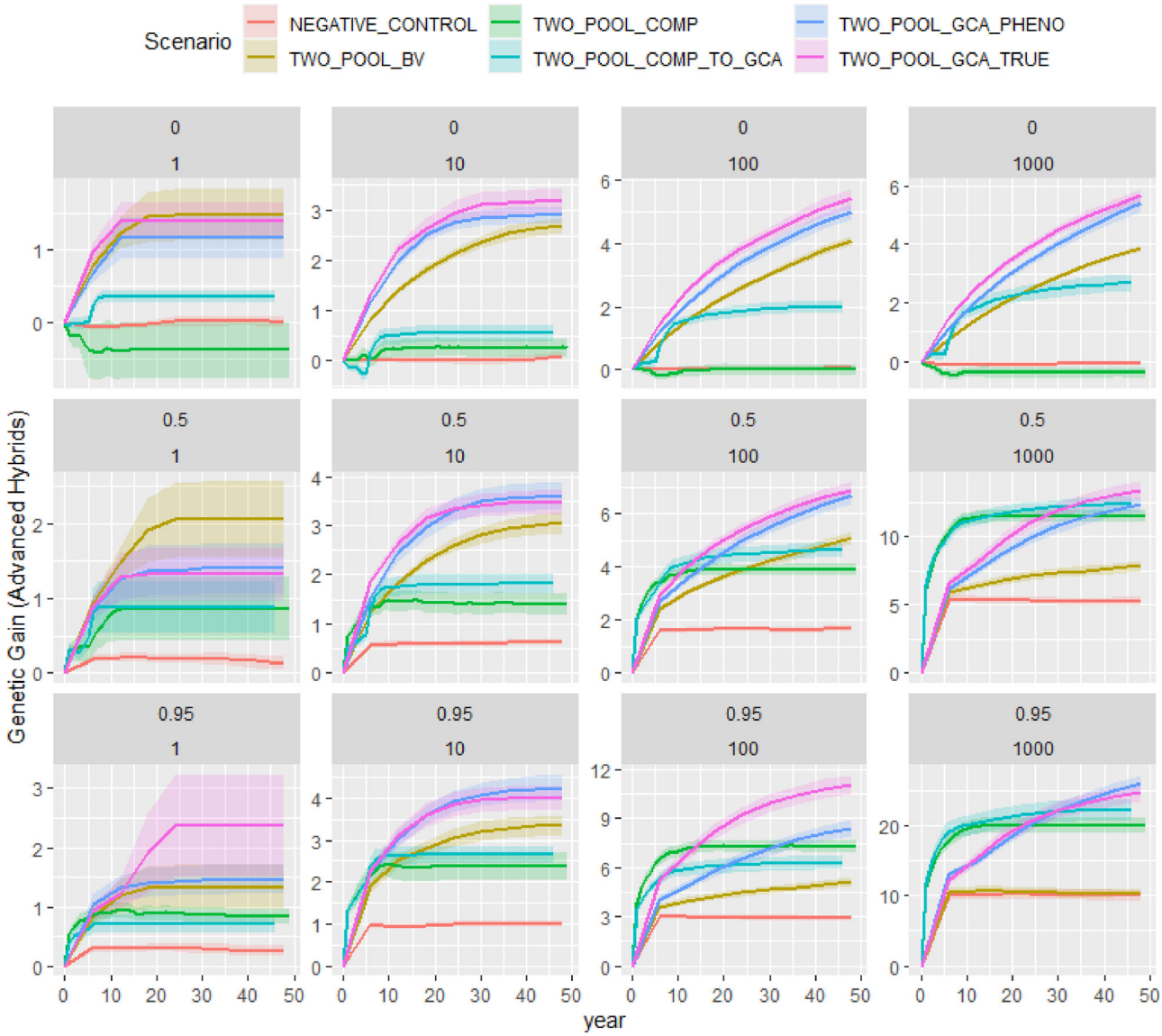

FIGURE 4 Total-genetic gain for a trait in on-farm hybrids under different mean dominance degrees and number of quantitative trait loci (QTLs) for different simulated breeding strategies. The columns represent the comparison of strategies where the simulated trait has different number of QTLs behind (1, 10, 50, 100, and 1000 QTLs). The rows represent the comparison of strategies where the simulated trait has different levels of mean dominance degree (0 implies a completely additive trait, 0.5 partial dominance, and 0.95 represents a trait with almost complete dominance). A value of varDD = 0.2 was used across all scenarios. In the *x* axis, the number of selection cycles (20) are indicated whereas in the *y* axis the gain is shown. The different lines represent the different selection strategies: (1) reciprocal recurrent selection program that selects and recycles parents based on general combining ability (GCA) (TWO_POOL_GCA_PHENO), (2) reciprocal recurrent selection program that selects and recycles parents based on genetic complementation (TWO_POOL_COMP), (3) reciprocal recurrent selection program that selects and recycles parents based on genetic complementation and then moves to GCA (TWO_POOL_COMP_TO_GCA), (4) reciprocal recurrent selection program that selects and recycles parents based on true GCA (TWO_POOL_GCA_TRUE; positive control), (5) recurrent selection program that selects and recycles parents based on breeding value (TWO_POOL_BV; negative control 1), (6) recurrent selection program that selects and recycles parents at random in a single pool (NEGATIVE_CONTROL 2).

control (TWO_POOL_BV) did not increase it (Figure 4 and Figure S2). At all levels of dominance, the complementation approaches were shown to be as effective to increase the dominance values as selecting based on GCA. On the other hand, when looking at the additive gain, we found the GCA-based and breeding value (BV)-based approaches to be the only strategies able to increase this value (Figure 4 and Figure S2). The total-genetic gain was positive for all strategies ranking first the RRS methods followed by complementation and last single-pool RS based on BV when the mean dominance degree was ≥0.5. At lower levels of mean dominance degree single-pool RS based on BV ranked first followed

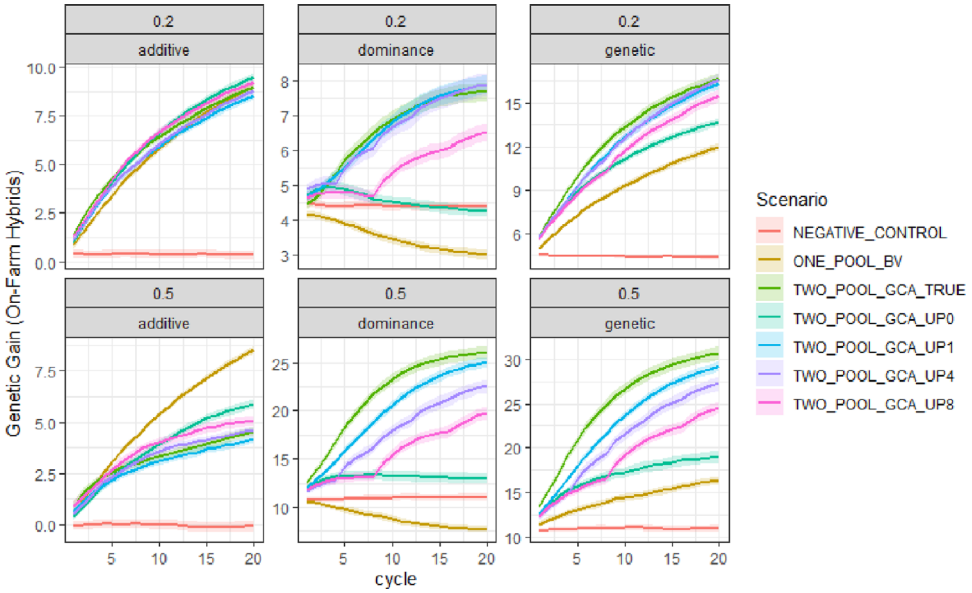

FIGURE 5 Effect of the representativeness of testers in the increase of dominance-based heterosis. The additive, dominance, and total-genetic value increment (columns) based on representativeness of the tester (with respect to its pool) achieved by updating the testers after none, 1, 4, or 8 cycles of selection (treatment lines) at two different levels of dominance degree (rows) is displayed. If the tester(s) are updated often (TWO_POOL_GCA_UP1) they better represent the pool they belong to, and dominance increases at a higher rate compared to programs not updating the testers (TWO_POOL_GCA_UP0) or not that often (rest of the treatments). Testers are updated based on their general combining ability value.

by RRS and complementation last. In addition, we found that trying to increase the dominance value in the breeding populations is not possible when testers are not updated (Figure 5).

3.3 | How to implement RRS by genetic complementation in an active inbred program

To understand the best way to implement the genetic complementation approach in an ongoing program that selects inbred materials (i.e., close to homozygote lines such as rice or wheat), we simulated a breeding program that follows the structure of a line crop. The simulation was initiated with a single population/pool that follows a conventional recurrent selection strategy comprised of a crossing block, a segregation step, a multiplication phase (four generations of single seed descent) and multiple stages of phenotypic evaluation in the target population of environments (TPE) (Figure 6).

In the case of the positive control or conventional line program (TWO_POOL_BV) the same rate of response to selection with respect to the burn-in strategy across the 40 years was observed as expected, being that this is the continuation of the same burn-in strategy (Figure 6). The negative control showed no increase in additive, dominance, and total-genetic value as would be foreseen under random selection. In the case of the program converted to traditional GCA-based RRS (TWO_POOL_GCA_PHENO) we found a lower rate of additive gain but a higher rate of dominance gain resulting in an overall increased rate of total-genetic value in the hybrid populations.

The RRS strategy using only the complementation surrogate for selection without phenotyping (TWO_POOL_COMP) was able to produce hybrids with similar performance to the conventional GCA-based RRS approach during the first 15 years after the split due to a substantial increase in dominance and consequently total-genetic value (Figure 6 and Figure S3), but the additive value in the hybrids and lines decreased in performance due to inbreeding (Figure 6 and Figure S3) causing a stagnation in hybrid performance increase after year 15. Although not shown, using better the genetic variance (e.g., using maximum avoidance or optimal contribution methods) can increase the window of dominance gain. The reciprocal recurrent genomic selection strategy (TWO_POOL_GCA_GS) showed greater gain than its phenotype-based counterpart (TWO_POOL_GCA_PHENO) but did not increase dominance much faster than the complementation approaches. Both phenotype-based and GS-based RRS strategies can be potentially boosted by the complementation strategy proposed.

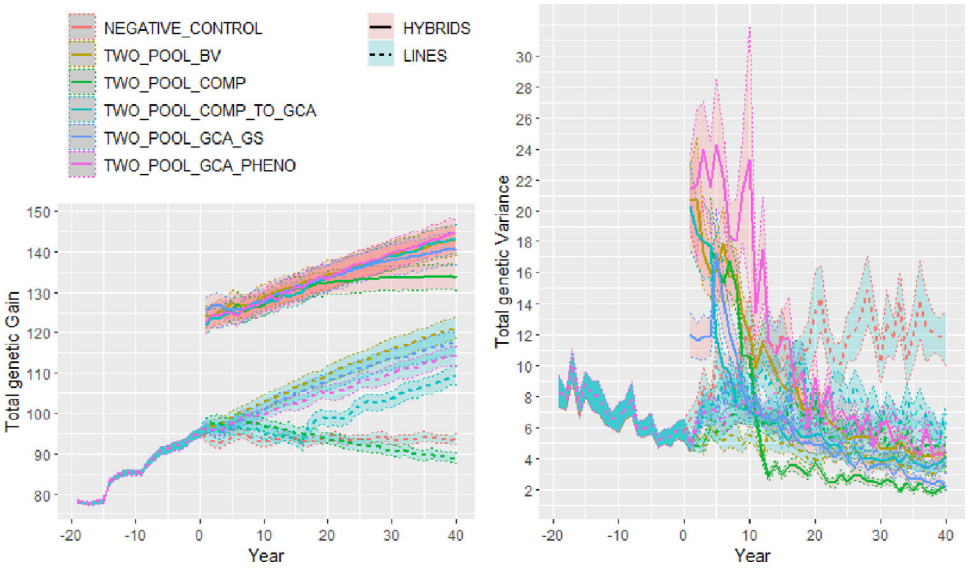

FIGURE 6 Genetic gain for a complex trait (1000 quantitative trait loci behind) with intermediate mean dominance degree (0.5) measured in on-farm hybrids (solid lines) and parental lines (dotted lines) for a simulated line program transitioning to a hybrid program with different breeding strategies. In the *x* axis, the number of years is indicated whereas in the *y* axis the genetic gain for (a) total-genetic, (b) additive, and (c) dominance value are shown. The different lines represent the different selection strategies tested: (1) line program continuing the recurrent selection based on breeding value in two independent pools and making hybrids among the pools (TWO_POOL_BV; positive control; pink lines), (2) line program transitioning to reciprocal recurrent selection (RRS) to select and recycle parents based on general combining ability (GCA) (TWO_POOL_GCA_PHENO; blue lines), (3) line program transitioning to RRS to select and recycle parents based on predicted GCA using GS (TWO_POOL_GCA_GS; blue lines), (4) line program transitioning to RRS to select purely on complementation (χ) (TWO_POOL_COMP), (5) line program transitioning to RRS to select purely on complementation (χ) for 12 years and then moved to formal RRS (TWO_POOL_COMP_TO_GCA), and negative control selecting lines at random (NEGATIVE_CONTROL; red lines). Every year four out of the top five testers are updated. The value of varDD was 0.2 across all scenarios. Shadow lines represent standard errors for treatments across 20 Monte-Carlo replications.

4 | DISCUSSION

4.1 | Validating the complementation model in a generic simulation and real datasets

Under complete dominance, the expected correlation between the hybrid performance and the metrics χ_{ip} and χ_{ii} is high since the genetic value is mainly driven by the dominance effects (Table 2), which at the same time are linked to heterozygosity. But as the mean dominance degree decreases the predictive ability of the complementation surrogate decreases (Table 2). The generic simulation model shows that as dominance degree increases the complementation surrogate can be useful for breeding programs to predict the dominance component of the hybrid performance.

The correlation values between the χ_{ip} metric and the hybrid performance found using the dataset from Kadam et al. (2016) and Technow et al. (2014), are closer to the correlation values found in the simulation for intermediate levels of dominance (meanDD ∼ 0.5) which are very similar to the degrees of dominance observed in maize at equilibrium in Nebraska's and North Carolina's experiments in the 1980s and 1990s (Bingham, 1998; Doebley, 2004; Duvick et al., 2004). Since simulations for the adoption of the complementation approach showed an advantage at intermediate and high levels of dominance, programs currently working as line programs are suitable candidates for the adoption of genetic complementation to increase dominance value and build heterosis (and de novo heterotic pools) more efficiently. The methodology does not focus on the initial split of germplasm in pools as the methodology proposed by Zhao et al. (2015), but on the formation of pools through the accumulation of dominance complementations through recurrent selection using molecular markers to build desired complementary haplotypes.

Regarding the low correlation values of χ_{ii} complementation (i.e., individual-to-individual complementation) with the hybrid performance, compared to the χ_{ip} complementation (i.e., the population level complementation), our hypothesis is that since the complementation is defined with respect to a desired haplotype of the opposite population pool, the complementation at the individual level loses accuracy and relevance.

4.2 | The influence of dominance degree and number of QTLs in total-genetic value

The fact that RRS strategies (TWO_POOL_GCA_PHENO and TWO_POOL_GCA_TRUE) strategies were able to build dominance effects efficiently, whereas the negative control (TWO_POOL_BV) did not increase it is because the separation of pools without a complementation surrogate or a GCA metric becomes a random process where complementary alleles are not guaranteed to go to the opposite pools and both pools may end up fixing the same allele. Building genetic distance alone is not expected to increase the dominance efficiently. For example, Butruille et al. (2004) observed that RRS caused changes in genetic distance between two maize heterotic pools formed out of a single population as simulated here. Interestingly, they hypothesized that the genetic distance observed was mainly caused by genetic drift since dominance-based heterosis was not built. This coincides with our results where two pools formed with purely intra-pool selection can generate genetic distance but not necessarily a continuous increase of heterosis. It becomes a random process (drift) where some populations may generate heterosis and others do not. This lack of correlation between genetic distance and heterosis has been observed in several studies in the past (Lamkey et al., 1987; Charcosset et al., 1991; Bernardo, 1992). Although the genetic complementation is based on the same principle, the complementation approach can be thought of as a controlled genetic distance that guarantees to maximize the heterotic response.

Previously, genetic distance has been proposed as a potential predictor of hybrid performance, with correlation between hybrid performance and genetic distance ranging between 0 and 0.3 (Bernardo, 1992; Frei et al., 1986; Zhang et al., 1996). Bernardo (1992) for example, found that for different values of allele frequencies between set A and set B, the correlation between μ_{ij} (performance) and D_{ij} (distance) was $r\mu_{ij}D_{ij} = 0.25$, whereas with partial dominance the correlation between μ_{ij} and D_{ij} decreased to $r\mu_{ij}D_{ij} = 0.13$. In other empirical studies, correlations of 0.09, 0.14, 0.32, and 0.46 were obtained by Godshalk et al. (1990), Dudley et al. (1991), Melchinger et al. (1990), and Lee et al. (1989), respectively. This is unsurprising under the directional-dominance theory where heterosis is the accumulation of the dominance effects, but not the total performance that includes the additive effects. What builds and maximizes heterosis (dominance value) is the divergence of two pools (for a diploid species) in a controlled manner to put complementary/opposite alleles (maybe even indels) or haplotypes in each pool. Splitting two population/pools without a control builds genetic distance and some random heterosis, but permits the same and different alleles (and small-effect mutations) to be increased or fixed in both populations by genetic drift (Charlsworth, 2009). In this case, it is most likely that populations which diverged for opposite alleles are the ones that will display greater heterosis (dominance effects) when crossed back (e.g., flint and dent population in maize), but there should be many others that have diverged in average but which have fixed more similar alleles and therefore show less heterosis (Allendorf, 1986; Lande, 1976; Lynch & Walsh, 1998; Lynch et al., 2016). One of the reasons why we found a clear correlation between the χ metric (a controlled genetic distance method) with the dominance effects could be because we started from a single pool with clear LD patterns as opposed to using genetic distance in natural populations where other subpopulation structure may affect the performance of genetic distance to predict dominance interactions. In summary, a big component of the total-genetic value is the additive component, and the creation of controlled genetic distance through complementation should only be able to predict the dominance component of the equation but not the total performance as expected by previous studies in the 1980s and 1990s. Still, there is value in the genetic complementation as a way to create de novo heterotic pools quickly without phenotyping to lower the costs prior to the start of the GCA-based approaches which increases both the additive and dominance value of populations.

The observation that at all levels of dominance, the complementation approach was shown to be almost as effective to increase the dominance values as selecting based on GCA is a promising result given the reduced complexity and resources (the only cost incurred is in genotyping a sample of the population in a nursery) involved in enabling the complementation approach for a couple of cycles before a GCA program is implemented. The observation that the complementation approach did not increase the additive value was not a surprise since no phenotyping is used in this method and there is no way to select for the positive allele. In a practical program, the complementation approach would only be used to build dominance-based heterosis and drive the allele frequencies apart and then the GCA approach would need to come into play.

4.3 | Implementing genetic complementation in an active inbred program to start hybrid breeding

We recorded the additive, dominance, and total-genetic value of the parental lines and the advanced hybrids (in the case of the hybrid strategies) across 40 years of applying the different breeding strategies (Figure 6). The increase of performance of hybrids compared to lines when the simulated program moves from one pool to two pools but recycles parents based on BV (TWO_POOL_BV) is only due to the recovery of

baseline heterosis rather than an increase in dominance or panmictic heterosis (based on the simulated level of dominance, meanDD = 0.5) (Cowling et al., 2020; Labroo et al., 2021; Lamkey & Edwards, 1999). Increase in genetic gain in hybrids following a two pool BV method is based on the increase of additive gain.

Interestingly, the lower rate of increase in additive gain of RRS based on GCA compared to the TWO_POOL_BV programs is due to the additional time taken for the additional crossing and evaluation generation of testcrosses and the way the GCA estimate is constructed (Figure 6) (Hallauer et al., 2010). The lower rate of additive gain using GCA gets compensated by higher dominance gain if and only if the testers are updated on a recurrent basis (based on their GCA value) (Figure 4). The decrease in additive value in lines and hybrids observed in the complementation approach (TWO_POOL_COMP) (Figure 6) is expected since both populations are trying to fix specific alleles without purging the negative alleles giving place to the average decrease in additive value due to inbreeding (Bernardo, 1992, 2002). With incomplete dominance, the complementation approach, the harnessing of heterosis comes at a "cost" to additive gain as Ne decreases and inbreeding depression increases. In hybrids, such inbreeding is surpassed by the increase of dominance value when the mean dominance degree >0.5, but as dominance variance gets depleted (in our simulations around year 15) inbreeding surpasses the effect of dominance value leading to stagnation of genetic gain in hybrids (Figure 6 and Figure S3). It is important to point out that it is expected that low values of dominance degree would not give any potential advantage to the complementation approach.

4.4 | Changes in allelic diversity across cycles

After several cycles of applying the complementation principle (TWO_POOL_COMP) we found the majority of the desired alleles for the causal QTLs simulated to be closed to fixation in each pool as strategized in a relatively short period of time. Selection intensity and low Ne likely played a major role in the quick depletion of genetic variance under the aggressive schemes tested (Walsh, 2004). Interestingly, in the case of the GCA-based approaches (TWO_POOL_GCA_PHENO and TWO_POOL_GCA_GS) the alleles fixed in each pool were in many cases the same, the positive allele of the causal QTL and a part were different between pools complementing each other and maintaining the heterozygous state in the interpopulation hybrids. Solomon et al. (2010), for example, reported the fixation of complementary alleles in two tropical maize populations that went under RRS for 35 years (13 cycles). One of the possible explanations for the fixation of complementary alleles in the pools is the presence of directional (positive) dominance. The estimates of expected versus observed heterozygosity after RRS shows that drift was not the only force behind the allele fixation. Several analyses of the changes in genetic variation and allelic diversity in maize RRS programs have been summarized by Labate et al. (1999) which confirms that complementation (maintenance of separate gene pools that allows different alleles to be fixed within each population) is the mechanism by which hybrid breeding leverages dominance-based heterosis. Of particular interest are the diversity studies based on molecular markers that confirm that selection and not drift is the main force of creating the complementary maize pools since allele frequencies in many alleles went beyond the expectation under drift (Hinze et al., 2005; Labate et al., 1999).

4.5 | Similar performance with lower cost

As mentioned throughout this study the major potential of the complementation methodology is the reduction of cost to produce high-performance complementary pools in the long term through recurrent selection for dominance at a low cost when the species displays an intermediate or high level of mean dominance degree. The lowers cost of the complementation + GCA comes from the fact that the number of complementation cycles that the breeding program decides to run prior to start formal RRS (GCA based) only incurs in genotyping costs (~$10 per sample) and a growing nursery (~$5K) where the material can be grown (without replication and experimental design) while the genotyping occurs and recombined once the complementation surrogate has been computed. Assuming the same size as the simulated RRS program with 60 crosses and 50 F2s (3000 F2s) the complementation approach would give a cost of ~$35K per cycle. This cost per cycle is notable lower than a formal RRS cycle where a nursery (~$5K), DH formation (~$100K assuming ~$2.2K per cross producing ~100 DHs per cross and 50 crosses), one per se evaluation of 6000 lines ($30K), a crossing block for testcross formation ($5K), and three seasons of testcross evaluation with 500, 200, 45 hybrids (~$30K per evaluation), which add up to a total of $230K (only focused on the variables costs and excluding fixed costs to simplify). This is approximately a 1:6 cost relationship.

A new complementation cycle can be immediately started since phenotyping for traits is not required, which allows to complete as many cycles in a year as the biological limit and speed breeding methods allow (Watson et al., 2018). In this paper we assumed a complementation cycle to last ~1 year, but nothing impedes the use of speed breeding to run 2 or more cycles per year and use 2–3 years of complementation prior to starting formal RRS. Is important to highlight that the complementation cycles are just a preamble to start the formal RRS since the increase of additive gain and production of varieties requires the phenotyping and multi-environment

testing that any breeding programs does, and the complementation approach can be seen as a enhancing methodology of GCA approaches to increase initially the dominance-based heterosis.

5 | CONCLUSION

The potential for hybrid breeding to fulfill the nutritional needs of a growing world population is based on the exploitation of heterosis, which is currently best explained by the theory of directional dominance. Hybrid breeding is approached through GCA using testcross-based or diallel-based approaches aiming to increase additive and dominance effects through recurrent selection, although this approach is highly efficient in diploids, it also requires substantial investment in phenotypic evaluation while a genomic selection variant can make gains faster. Here, we proposed that genetic marker data can be used to compute a surrogate of complementation between pools or subpopulations in order to accumulate dominance interactions through recurrent selection quickly and create de novo heterotic pools prior to the start of the GCA-based approaches (potentially in parallel). We found that the proposed genetic-complementation-based approach outperforms conventional approaches in terms of dominance gain per unit cost when dominance ranges from intermediate to high values (meanDD ≥ 0.5) but as expected not for additive gain. The complementation approach can be thought of as a coordinated genetic distance method or recurrent selection for dominance. The results in real datasets from dent by flint maize hybrids seem to validate the complementation theory and surrogates of complementation. The complementation surrogate is an alternative for breeding programs attempting to transition from a non-hybrid to a hybrid system (in a diploid species) to increase and maximize the dominance-based heterosis per unit cost and complement the GCA-based approaches. In addition, this simulation study shows the consequences of the allelic complementation (under the directional-dominance heterosis model) created by RRS programs and how quickly the heterotic potential of a population can be exposed. Interestingly, if the marker-based complementation approach does not produce the heterotic response in some traits as expected, it could also provide insight on whether dominance is the main driver of heterosis observed in different species.

AUTHOR CONTRIBUTIONS

Giovanny Covarrubias-Pazaran: Conceptualization; formal analysis; funding acquisition; methodology; writing – original draft. **Christian Werner**: Conceptualization; validation; writing – review & editing. **Dorcus Gemenet**: Conceptualization; validation; writing – review & editing.

ACKNOWLEDGMENTS

We thank the different breeding teams from the CGIAR that helped through their questions to conceive these ideas. We thank the two anonymous reviewers for the feedback and ideas that helped to improve considerably the quality of this manuscript. The CGIAR Excellence in Breeding Platform received funding from the Bill and Melinda Gates Foundation (BMGF) grant number OPP1177070.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The real hybrid datasets used in this study are available as supplementary files in the publications from Technow et al. (2014) and Kadam et al. (2016) cited in this publication.

ORCID

Giovanny Covarrubias-Pazaran https://orcid.org/0000-0002-7194-3837

REFERENCES

Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, *5*(2), 181–190. https://doi.org/10.1002/zoo.1430050212

Bernardo, R. (1992). Relationship between single-cross performance and molecular marker heterozygosity. *Theoretical and Applied Genetics*, *83*(5), 628–634. https://doi.org/10.1007/BF00226908

Bernardo, R. (2002). *Breeding for quantitative traits in plants* (Vol. 1, p. 369). Stemma Press.

Bingham, E. T. (1998). Role of chromosome blocks in heterosis and estimates of dominance and overdominance. *Concepts and Breeding of Heterosis in Crop Plants*, *25*, 71–87.

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., & Veitia, R. A. (2010). Heterosis. *The Plant Cell*, *22*(7), 2105–2112. https://doi.org/10.1105/tpc.110.076133

Butruille, D. V., Silva, H. D., Kaeppler, S. M., & Coors, J. G. (2004). Response to selection and genetic drift in three populations derived from the Golden Glow maize population. *Crop Science*, *44*(5), 1527–1534. https://doi.org/10.2135/cropsci2004.1527

Charcosset, A., Lefort-Buson, M., & Gallais, A. (1991). Relationship between heterosis and heterozygosity at marker loci: A theoretical computation. *Theoretical and Applied Genetics*, *81*(5), 571–575. https://doi.org/10.1007/BF00226720

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, *10*(11), 783–796.

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*(1), 136–142.

Comstock, R. E., & Robinson, H. F. (1952). Estimation of average dominance of genes. *Heterosis*, *2*, 494–516.

Comstock, R. E., Robinson, H. F., & Harvey, P. H. (1949). Breeding procedure designed to make maximum use of both general and specific

combining ability. *Agronomy Journal*, *41*, 360–367. https://doi.org/10.2134/agronj1949.00021962004100080006x

Cowling, W. A., Gaynor, R. C., Antolín, R., Gorjanc, G., Edwards, S. M., Powell, O., & Hickey, J. M. (2020). In silico simulation of future hybrid performance to evaluate heterotic pool formation in a self-pollinating crop. *Scientific Reports*, *10*(1), 4037. https://doi.org/10.1038/s41598-020-61031-0

Crow, J. F. (1999). Dominance and overdominance. In J. G. Coors & S. Pandey (Eds.), *Genetics and exploitation of heterosis in crops* (pp. 49–58). ASA, CSSA.

Davenport, C. B. (1908). Degeneration, albinism and inbreeding. *Science*, *28*(718), 454–455. https://doi.org/10.1126/science.28.718.454.c

de Boer, I. J., & Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theoretical and Applied Genetics*, *86*(2), 245–258. https://doi.org/10.1007/BF00222086

Doebley, J. (2004). The genetics of maize evolution. *Annual Review of Genetics*, *38*, 37–59. https://doi.org/10.1146/annurev.genet.38.072902.092425

Dudley, J. W., Maroof, M. S., & Rufener, G. K. (1991). Molecular markers and grouping of parents in maize breeding programs. *Crop Science*, *31*(3), 718–723. https://doi.org/10.2135/cropsci1991.0011183X003100030036x

Duvick, D. N., Smith, J. S. C., & Cooper, M. (2004). Long-term selection in a commercial hybrid maize breeding program. *Plant Breeding Reviews*, *24*(2), 109–152.

East, E. M. (1936). Heterosis. *Genetics*, *21*(4), 375. https://doi.org/10.1093/genetics/21.4.375

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, *300*(5620), 758–762. https://doi.org/10.1126/science.1078710

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Longman.

Frei, O. M., Stuber, C. W., & Goodman, M. M. (1986). Use of allozymes as genetic markers for predicting performance in maize single cross hybrids 1. *Crop Science*, *26*(1), 37–42. https://doi.org/10.2135/cropsci1986.0011183X002600010008x

Gaynor, R. C., Gorjanc, G., & Hickey, J. M. (2021). AlphaSimR: An R package for breeding program simulations. *G3: Genes, Genomes, Genetics*, *11*(2), jkaa017. https://doi.org/10.1093/g3journal/jkaa017

Godshalk, E. B., Lee, M., & Lamkey, K. R. (1990). Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. *Theoretical and Applied Genetics*, *80*(2), 273–280. https://doi.org/10.1007/BF00224398

Hallauer, A. R., Carena, M. J., & Filho, J. D. M. (2010). *Quantitative genetics in maize breeding* (Vol. 6). Springer Science & Business Media.

Hedden, P. (2003). The genes of the Green Revolution. *TRENDS in Genetics*, *19*(1), 5–9. https://doi.org/10.1016/S0168-9525(02)00009-4

Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 73–85. https://doi.org/10.1098/rstb.2009.0203

Hinze, L. L., Kresovich, S., Nason, J. D., & Lamkey, K. R. (2005). Population genetic diversity in a maize reciprocal recurrent selection program. *Crop Science*, *45*(6), 2435–2442. https://doi.org/10.2135/cropsci2004.0662

Jiang, Y., Schmidt, R. H., Zhao, Y., & Reif, J. C. (2017). A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nature Genetics*, *49*(12), 1741–1746. https://doi.org/10.1038/ng.3974

Jinks, J. L., & Jones, R. M. (1957). Estimation of the components of heterosis. *Genetics*, *43*(2), 223. https://doi.org/10.1093/genetics/43.2.223

Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., Jackson, A. U., Schurmann, C., Smith, A. V., Zhang, W., Okada, Y., Stančáková, A., Faul, J. D., Zhao, W., Bartz, T. M., Concas, M. P., Franceschini, N., Enroth, S., Vitart, V., … Kubo, M. (2015). Directional dominance on stature and cognition in diverse human populations. *Nature*, *523*(7561), 459–462. https://doi.org/10.1038/nature14618

Kadam, D. C., Potts, S. M., Bohn, M. O., Lipka, A. E., & Lorenz, A. J. (2016). Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. *G3: Genes, Genomes, Genetics*, *6*(11), 3443–3453. https://doi.org/10.1534/g3.116.031286

Labate, J. A., Lamkey, K. R., Lee, M., & Woodman, W. L. (1999). Population genetics of increased hybrid performance between two maize populations under reciprocal recurrent selection. In J. G. Coors & S. Pandey (Eds.), *Proceedings of the International Symposium on the genetics and exploitation of heterosis in crops* (pp. 127–137). ASA, CSSA, and SSSA.

Labroo, M. R., Studer, A. J., & Rutkoski, J. E. (2021). Heterosis and hybrid crop breeding: A multidisciplinary review. *Frontiers in Genetics*, *12*, 643761.

Lamkey, K. R., & Edwards, J. W. (1999). Quantitative genetics of heterosis. In J. G. Coors & S. Pandey (Eds.), *Genetics and exploitation of heterosis in crops* (pp. 31–48). ASA, CSSA.

Lamkey, K. R., Hallauer, A. R., & Kahler, A. L. (1987). Allelic differences at enzyme loci and hybrid performance in maize. *Journal of Heredity*, *78*(4), 231–234. https://doi.org/10.1093/oxfordjournals.jhered.a110372

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution; International Journal of Organic Evolution*, *30*, 314–334. https://doi.org/10.2307/2407703

Lee, M., Godshalk, E. B., Lamkey, K. R., & Woodman, W. W. (1989). Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. *Crop Science*, *29*(4), 1067. https://doi.org/10.2135/cropsci1989.0011183X002900040050x

Lippman, Z. B., & Zamir, D. (2007). Heterosis: Revisiting the magic. *Trends in Genetics*, *23*(2), 60–66. https://doi.org/10.1016/j.tig.2006.12.006

Lynch, M., Ackerman, M. S., Gout, J. F., Long, H., Sung, W., Thomas, W. K., & Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, *17*(11), 704–714. https://doi.org/10.1038/nrg.2016.104

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates.

Melchinger, A. E., Lee, M., Lamkey, K. R., & Woodman, W. L. (1990). Genetic diversity for restriction fragment length polymorphisms: Relation to estimated genetic effects in maize inbreds. *Crop Science*, *30*(5), 1033. https://doi.org/10.2135/cropsci1990.0011183X003000050016x

Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values*. CABI.

Nishio, M., & Satoh, M. (2014). Including dominance effects in the genomic BLUP method for genomic evaluation. *PLoS ONE*, *9*(1), e85792. https://doi.org/10.1371/journal.pone.0085792

Shull, G. H. (1952). Beginnings of the heterosis concept. In J. W. Gowen (Ed.), *Heterosis* (pp. 14–48). Iowa State College Press.

Solomon, K. F., Martin, I., & Zeppa, A. (2010). Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. *Euphytica*, *176*(2), 239–249. https://doi.org/10.1007/s10681-010-0238-4

Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., & Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics*, *197*(4), 1343–1355.

Varona, L., Legarra, A., Toro, M. A., & Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Frontiers in Genetics*, *9*, 78.

Walsh, B. (2004). Population-and quantitative-genetic models of selection limits. *Plant Breeding Reviews*, *24*(1), 177–226.

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., Md Hatta, M. A., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N. M., Breakspear, A., Korolev, A., Rayner, T., Dixon, L. E., Riaz, A., Martin, W., Ryan, M., Edwards, D., … Hickey, L. T. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants*, *4*(1), 23–29. https://doi.org/10.1038/s41477-017-0083-8

Werner, C. R., Gaynor, R. C., Sargent, D. J., Lillo, A., Gorjanc, G., & Hickey, J. M. (2023). Genomic selection strategies for clonally propagated crops. *Theoretical and Applied Genetics*, *136*(4), 74.

Zhang, Q., Zhou, Z. Q., Yang, G. P., Xu, C. G., Liu, K. D., & Maroof, M. S. (1996). Molecular marker heterozygosity and hybrid performance in indica and japonica rice. *Theoretical and Applied Genetics*, *93*(8), 1218–1224. https://doi.org/10.1007/BF00223453

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., Mock, H.-P., Matros, A., Ebmeyer, E., Schachschneider, R., Kazman, E., Schacht, J., Gowda, M., Longin, C. F. H., & Reif, J. C. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proceedings of the National Academy of Sciences*, *112*(51), 15624–15629. https://doi.org/10.1073/pnas.1514547112

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Covarrubias-Pazaran, G., Werner, C., & Gemenet, D. (2023). Reciprocal recurrent selection based on genetic complementation: An efficient way to build heterosis in diploids due to directional dominance. *Crop Science*, *63,* 2205–2219. https://doi.org/10.1002/csc2.21018