Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Population Size Drives the Long-Term Efficiency of Mating Optimization in Traditional and Genomic Selection Breeding Programs

Jose Artur Casimiro
Universidade Federal do Ceará
Liana Arriada Pereira Fritsche
BR Prediction Breeding Solutions
Júlio César DoVale
Universidade Federal do Ceará
Roberto Fritsche-Neto
roberto.neto@ncsu.edu

North Carolina State University

Research Article

Keywords: Genomic Selection, mate selection, breeding size, genetic drift, stochastic simulation, Cucumis sativus

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11040960/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Population Size Drives the Long-Term Efficiency of Mating Optimization in Traditional and Genomic Selection Breeding Programs

José Artur de Oliveira Casimiro – Federal University of Ceará (UFC), Fortaleza, Ceará, Brazil. ORCID Number: 0000-0003-3683-435X

Liana Arriada Pereira Fritsche - BR Prediction Breeding Solutions, Raleigh, North Carolina, USA. ORCID Number: 0009-0008-3646-4487.

Júlio César DoVale – Federal University of Ceará (UFC), Fortaleza, Ceará, Brazil. ORCID Number: 0000-0002-3497-9793.

Roberto Fritsche-Neto (Corresponding Author)- North Carolina State University (NCSU). ORCID Number: 0000-0003-4310-0047. E-mail: roberto.neto@ncsu.edu

Abstract

Recurrent genomic selection accelerates genetic gain but rapidly depletes additive genetic variance, especially in small populations. It remains unclear whether complex mating optimization algorithms can overcome the limits imposed by reduced breeding sizes, common in the public sector. Therefore, we used stochastic simulations to evaluate the long-term (40-year) efficiency of six mating strategies across three breeding population sizes (10, 20, or 40 parents per cycle), focusing on response to selection, accuracy, and genetic variability under traditional and genomic selection (GS)-based breeding programs. We found that breeding size primarily determines long-term genetic response, explaining over 90% of the variation in accumulated gain. Furthermore, mating optimization strategies perform differently depending on population size. Under severe population bottlenecks, genetic drift overpowered the theoretical benefits of all mating algorithms, leading to rapid variance depletion and early response plateaus. Conversely, as population size increased, the true potential of mating optimization emerged. Finally, among the evaluated mating-allocation approaches, GS_Het, a naïve method based on expected progeny heterozygosity, is particularly efficient and outperforms the other tested methods.

Keywords: Genomic Selection; mate selection; breeding size; genetic drift; stochastic simulation; *Cucumis sativus*.

Running title: Population size constrains mating optimization.

Key Message: Extremely small breeding populations limit the benefits of mating optimization algorithms, whereas larger populations allow even simple strategies based on expected progeny heterozygosity to efficiently enhance long-term genetic gain

Introduction

Genomic selection (GS) has transformed modern breeding by enabling early prediction of genetic merit from genome-wide molecular markers, substantially accelerating genetic gain compared to conventional phenotypic selection (Meuwissen et al. 2001; Jannink et al. 2010; Hayes et al. 2013; Crossa et al. 2017). By shortening breeding cycles and increasing selection intensity, GS has become a cornerstone of recurrent breeding programs across crops and livestock. However, the same mechanisms that drive rapid short-term improvement may also compromise the long-term sustainability of genetic progress.

A central challenge of recurrent genomic selection is the progressive erosion of genetic variability. Because selection decisions typically rely on truncated genomic estimated breeding values (GEBVs), individuals with superior predicted performance are repeatedly selected as parents, often co-selecting closely related genotypes (Li et al. 2022; Ahadi et al. 2024). This process accelerates the accumulation of coancestry and reduces effective population size (Ne), increasing the influence of genetic drift and limiting the pool of segregating alleles available for future selection (Makanjuola et al. 2020). Consequently, breeding programs may plateau early despite substantial initial gains (Viana et al. 2026a).

The Bulmer effect further exacerbates the depletion of additive genetic variance, as directional selection induces negative linkage disequilibrium among favorable alleles and reduces the additive variance available to selection before alleles become fixed (Bulmer 1971; Wientjes et al. 2022). Under sustained recurrent selection, this reduction in genetic variance may compromise both the persistence of genetic gain and the long-term accuracy of genomic prediction models. Maintaining

genetic variability, therefore, is a fundamental prerequisite for sustaining response to selection over multiple breeding cycles.

To address this challenge, researchers have developed several mate allocation strategies to balance short-term genetic gain with the preservation of genetic variability. These approaches range from structured mating schemes that equalize parental contributions to optimization algorithms that explicitly account for genomic relationships, expected progeny merit, or inbreeding constraints (Woolliams et al. 2015; Vanavermaete et al. 2020). Recent developments include exact optimization frameworks such as COMA (Endelman 2025) and Simple Mating (Peixoto et al. 2025), as well as heuristic and selection approaches such as STPGA (Akdemir and Sánchez 2016). Although these methods differ in their mathematical formulation, they share a common objective: maximizing long-term genetic improvement while slowing the loss of variability caused by recurrent selection.

Despite substantial advances in mating selection, an important question remains unresolved. Most optimization frameworks have been evaluated under conditions with sufficient genetic diversity and a relatively large search space. In practice, however, many breeding populations operate under population size constraints imposed by limited numbers of parents, restricted crossing capacity, or prolonged cycles of intense selection. Under these conditions, reductions in effective population size may amplify genetic drift and constrain optimization algorithms' ability to identify mating configurations that maximize gain while maintaining variability (Makanjuola et al. 2020; Xia and Dong 2022; Waples 2022).

Consequently, from a population-genetic perspective, it remains unclear whether increasingly sophisticated mating algorithms can overcome the bottleneck's fundamental limitations. As effective population size declines, stochastic changes in allele frequencies become increasingly important, potentially overwhelming the benefits expected from optimization procedures designed to control inbreeding and preserve additive variance. If population size constraints become sufficiently severe, genetic drift may dominate the breeding population’s trajectory, reducing mate allocation effectiveness regardless of algorithmic complexity. In this context, we used stochastic simulations to evaluate the long-term (40-year) efficiency of six mating strategies across three breeding population sizes (10, 20, or 40 parents per

cycle), assessing response to selection, selection accuracy, and genetic variability in both traditional and genomic selection (GS)-based breeding programs.

Materials and methods

Simulation of the founder population

All simulations were conducted in R using the *AlphaSimR* package (Gaynor et al. 2021) implemented in R software environment (*version* 4.5.1; R Core Team, 2023). To provide a biologically realistic framework while maintaining control over selection processes, we parameterized the simulations according to the genomic architecture of cucumber (*Cucumis sativus*). Cucumber was selected as the biological model system because of its well-characterized, compact diploid genome with seven chromosomes (n = 7), rapid generation cycles, and its representativeness of closed vegetable breeding programs, which are often constrained by narrow genetic bases and intense bottlenecks. By using its well-known genomic parameters, such as the number of chromosomes, genome size, the number of expected QTLs controlling the target trait, and the observed heritability (Guan et al. 2024), the simulations capture a realistic linkage and recombination landscape. This enables robust inferences about additive-variance depletion, response to selection, and selection accuracy over long-term recurrent selection cycles.

The founder population was generated using the Markovian Coalescent Simulator (MaCS) (Chen et al. 2009), producing a historical diploid population of 1,234 individuals modeled under high inbreeding. This historical effective population size reflects the standing genetic variability typically observed in cucumber germplasm collections and diversity panels, with seven chromosomes containing 578 segregating sites per chromosome in addition to marker positions. The simulated trait was controlled by 420 quantitative trait loci (QTLs) distributed across the genome (60 QTLs per chromosome), which realistically represents complex polygenic traits while preventing rapid, artificial allele fixation under intense selection (Wientjes et al. 2022). Additive and dominance effects were sampled from a gamma distribution with a standard shape parameter of ($\alpha = 0.4$) (Hayes and Goddard, 2001) and scaled to generate a founder population with a mean genetic value of zero and an additive genetic variance of one (Va = 1.0).

Genotyping was simulated using a genome-wide panel of 2,996 single-nucleotide polymorphism (SNP) markers (428 SNPs per chromosome). Phenotypic values were generated by combining true breeding values (TBV) with normally distributed environmental deviations. To mimic the progressive increase in experimental precision typically observed across breeding stages, we dynamically rescaled residual variance throughout the breeding pipeline. Consequently, narrow-sense and broad-sense heritabilities were adjusted according to each generation as follows: 0.03/0.06 (F_2), 0.15/0.20 (F_3), 0.40/0.45 (F_4), 0.60/0.63 (F_5), 0.70/0.72 (F_6), and 0.80/0.81 (F_7). This framework reflects the gradual increase in phenotypic accuracy associated with greater replication, larger experimental units, and more advanced testing stages within each breeding cycle.

To maintain focus on the segregation of standing genetic variation, we did not simulate mutations during the selection cycles. Because mutation is a rare biological event whose effects predominantly manifest over evolutionary timeframes, it does not significantly influence genetic variance within the breeding term horizons. Under these conditions, selection pressure and genetic drift almost exclusively govern the dynamics of genetic gain and variance depletion, operating on the standing genetic variation of the founder population.

Burn-in phase and establishment of population size scenarios

A burn-in phase was implemented to establish breeding populations with distinct population histories before the onset of genomic selection. This stage generated populations that differed in breeding size and available genetic variability, creating contrasting conditions for evaluating the long-term performance of mating allocation strategies.

Three population size scenarios were considered, comprising 10, 20, or 40 parents composing the crossing block every cycle. From these parents, 20, 40, and 80 biparental crosses were obtained, respectively (Figure 1). Consequently, assuming 100 progenies per cross, the initial segregating populations consisted of 2,000, 4,000, and 8,000 F_2 individuals. Through varying the number of parents while maintaining a proportional crossing structure, the simulations generated by breeding populations exposed to distinct levels of populational constraint and expected rates

of genetic drift. These breeding sizes (10, 20, and 40 parents per cycle) represent typical crossing-block dimensions in university and public-sector specialty crop pipelines, where physical and budgetary constraints limit breeding population size relative to industrial programs.

Each scenario underwent three consecutive cycles of recurrent phenotypic selection following a pedigree breeding framework. Within each cycle, we generated random biparental crosses and applied selection across successive selfing generations. Selection intensity was adjusted to the operational limits of each population scenario, ensuring comparable breeding structures while preserving differences in population size.

To emulate modern breeding programs that aim to maximize annual genetic gain by reducing generation intervals, we recombined selected F_5 lines and used them as parents in the subsequent cycle. Advanced evaluations in the F_6 and F_7 generations were conducted in parallel and served exclusively to estimate the maximum genetic merit achievable in fully inbred elite lines. After the third phenotypic selection cycle, the selected F_5 lines constituted Cycle 0 (C0), the starting population for all subsequent genomic selection and mating allocation strategies.

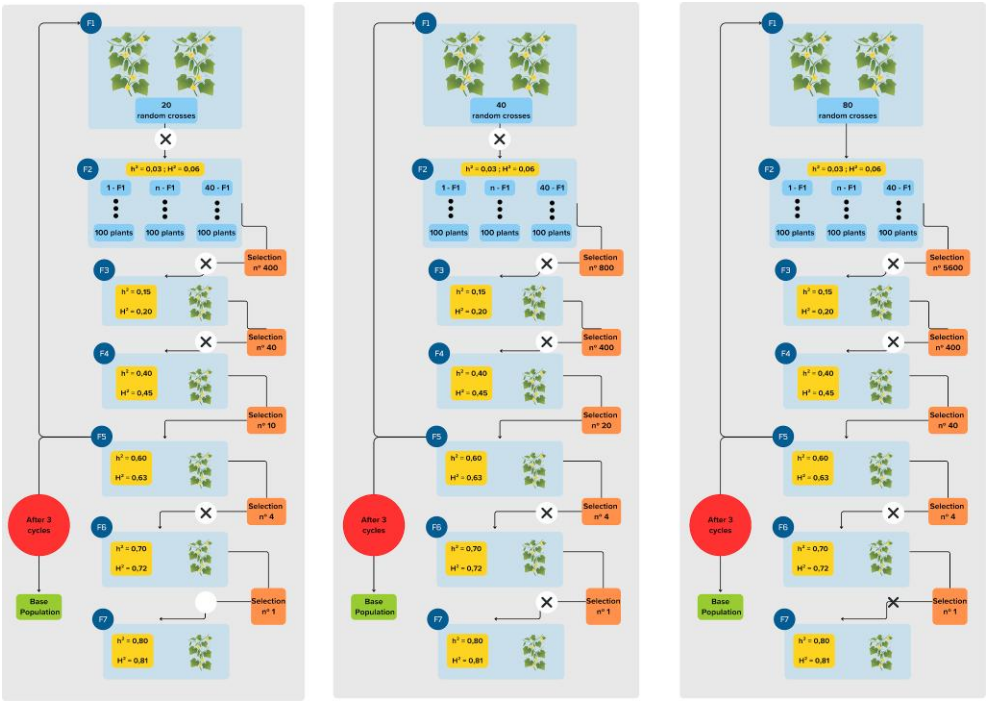


Fig. 1 Schematic representation of the burn-in phase

Importantly, this burn-in procedure established populations that differed not only in parent number but also in their expected trajectories of artificial selection and evolution. Because reductions in breeding size increase the relative influence of genetic drift and accelerate the loss of segregating genetic variation, these population sizes provided a framework for investigating how population size constrains mating optimization algorithms' capacity to sustain long-term genetic improvement.

Genomic prediction and estimation of genomic relationships

Genomic estimated breeding values (GEBVs) were predicted using Ridge Regression Best Linear Unbiased Prediction (RR-BLUP; Endelman 2011), combined with the native genomic prediction functions available in *AlphaSimR* package (Gaynor et al. 2021; Bančič et al. 2024).

Because recurrent selection continuously alters allele frequencies, linkage disequilibrium patterns, and genomic relationships among individuals, we recalibrated prediction models at each breeding cycle to account for temporal changes in population structure. This dynamic updating strategy minimized the loss of predictive ability as marker–QTL associations progressively decayed across successive generations. To do so, we updated the training populations using a sliding-window approach following the framework proposed by Sabadin et al. (2022). At each cycle, the training set incorporated genotypic and phenotypic information from F_4 individuals originating from the three most recent breeding cycles. This procedure maintained a close genetic relationship between training and selection candidates while reflecting the evolving genetic composition of the breeding population.

Prediction accuracy was evaluated throughout the recurrent selection process as the Pearson correlation between genomic estimated breeding values (GEBVs) and true breeding values (TBVs). Monitoring predictive accuracy across cycles provided an indirect measure of how population size, mating selection criteria, and depletion of additive genetic variance influenced the long-term effectiveness of genomic selection.

To support mating allocation algorithms that explicitly incorporate relatedness constraints (GS_STPGA, GS_COMA, and GS_SimpleMating), genomic relationship matrices were computed according to the method of VanRaden (2008) using the *snpReady* package (Granato et al. 2018). Before constructing the genomic relationship matrix, we filtered markers dynamically at each breeding cycle. A minor allele frequency (MAF) threshold of 0.05 and a call rate threshold of 0.95 were applied. Crucially, in our stochastic simulations, genotyping failure does not occur (nominal call rate = 1.0). We included the call rate threshold to maintain a standard quality control pipeline, while the MAF filter dynamically discarded monomorphic markers or those approaching fixation due to intense recurrent selection and genetic drift, preventing numerical instability during genomic relationship matrix computation.

Mating strategies and population size scenarios

Following the establishment of the cycle 0 (C0) populations, recurrent selection was conducted for 40 breeding years under three population sizes differing in parent number and crossing capacity. The scenarios comprised 10, 20, and 40 selected parents, generating 20, 40, and 80 biparental crosses per cycle, respectively. These contrasting population conditions were designed to represent populations exposed to different levels of genetic drift, opportunities for recombination and segregation, and finally, selection.

Baseline mating strategies

To investigate how mating design influences the long-term balance between genetic gain and variability preservation, we evaluated a series of phenotypic and genomic selection strategies. In this context, we used two conventional phenotypic breeding schemes as references. The first corresponded to traditional recurrent phenotypic selection followed by random mating among selected parents (TRAD). The second incorporated a circular mating design (TRAD_BLCD), which promoted a more balanced parental contribution and reduced the probability of repeated mating among closely related individuals.

Similarly, two baseline genomic selection schemes were implemented. The GS strategy combined genomic selection with random mating, whereas GS_BLCD integrated genomic selection with a partial diallel crossing structure. These approaches provided reference scenarios for evaluating whether optimization-based mate-allocation strategies yield advantages beyond those achieved through simple control of parental contributions.

Oriented genomic mating strategies

Four genomic mating approaches were evaluated to explore alternative mechanisms for maintaining genetic variability while maximizing long-term response to selection.

1) The GS_Het strategy explicitly maximizes the retention of segregating genetic variation by prioritizing parental pairs expected to yield highly heterozygous progeny. For a set of active candidates, the algorithm performs in silico crosses for all possible unique parent pairs. Let the expected progeny allele dosage at locus k for the cross between parents i and j be computed as:

$$\hat{g}_{ij,k} = \frac{M_{i,k} + M_{j,k}}{2}$$

where $M_{i,k}$ and $M_{j,k}$ represent the marker genotype vectors (coded as 0, 1, 2) of parents *i* and *j* at locus *k*. Expected progeny heterozygosity is maximized when parents are homozygous for alternative alleles (i.e., $\hat{g}_{ij,k} = 1$). The total expected heterozygosity score ($H_{i,j}$) for the cross is calculated using the indicator function:

$$H_{ij} = \sum_{k=1}^{m} I\left(\hat{g}_{ij,k} = 1\right)$$

All potential crosses are sorted in descending order of expected heterozygosity score. To prevent parental dominance and pedigree imbalance under intense selection, selection of the top crosses is subject to maternal and paternal contribution constraints, capping individual parents to a maximum of 10 crosses as female and 8 crosses as male.

2) The GS_STPGA strategy treats mate allocation as a combinatorial subset selection problem (Akdemir and Sánchez 2016). It searches the

multidimensional space of all possible parental crosses to identify a subset of crosses (x) of size $N_{crosses}$ that maximizes the objective function:

$$f(x) = (1-\lambda)\mu_G(x) - \lambda K(x)$$

where $\mu_G(x)$ is the mean expected progeny genomic estimated breeding value (GEBV), calculated as the average of the mid-parent predicted values for the selected crosses:

$$\mu_G(x) = \frac{1}{N_{crosses}} \sum_{(i,j)\in x} \frac{G_i + G_j}{2}$$

and $K(x)$ is the mean genomic relationship among the selected crosses, extracted from the genomic relationship matrix (***K***) calculated following VanRaden (2008):

$$K(x) = \frac{2}{N_{crosses}(N_{crosses}+1)} \sum_{(i,j),(u,v)\in x\ (i,j)\leq(u,v)} K_{(i,j),(u,v)}$$

Here, λ (set to 0.5) is the weighting parameter balancing short-term response and diversity retention. Heuristic search parameters (population of 100 solutions, 10 elites, mutation of 0.8, and Tabu memory size of 5) ensure a robust exploration of the combinatorial space.

3) The GS_COMA strategy represents the genomic equivalent of Optimal Contribution Selection (OCS) implemented via exact convex optimization (Endelman 2025). It solves for continuous mating weights (w_{ij}) for all candidate parent pairs i,j that maximize expected progeny merit:

$$Maximize \quad \sum_{i<j} w_{ij} \frac{G_i + G_j}{2}$$

subject to normalization and non-negativity constraints: $w_{ij} \geq 0 \quad and \quad \sum_{i<j} w_{ij} = N_{crosses}$, respectively. Inbreeding rate limitation is enforced by constraining the maximum rate of coancestry/inbreeding increment ($\Delta F = 0.005$), quadratically expressed as $c'Kc \leq C$, where c is the vector of parental contributions with elements $c_i = \sum_{j\neq i} \frac{w_{ij}}{2N_{crosses}}$, K is the genomic relationship matrix, and C is the maximum allowable relationship threshold. To prevent single-parent dominance, individual parental contributions are bounded as $c_i \leq 2/N_{parents}$, corresponding to maximum contribution limits of 0.20, 0.10, and 0.05

per parent in the 10-, 20-, and 40-parent scenarios, respectively. The strategy then probabilistically samples continuous allocations based on their weights to generate the final discrete crossing plan.

4) The GS_SimpleMating strategy optimizes crossing configurations to maximize average progeny performance while satisfying predefined linear constraints on parental relationships and contribution limits (Peixoto et al. 2025). The optimization pipeline operates in two sequential stages. First, a kinship-based filtering step reduces genetic redundancy in the candidate pool: parental candidates sharing a genomic relationship greater than 0.5 are grouped into clusters, and a maximum of two representative individuals per cluster are retained. Second, an optimization algorithm selects the set of $N_{crosses}$ crosses that maximizes expected cross performance (mid-parent estimated breeding value). This selection is subject to strict operational constraints: individual parental contributions are bounded between 1 and 8 crosses per parent, and any prospective mating pair exceeding a pairwise genomic relationship threshold of 0.5 is automatically culled to prevent severe immediate inbreeding.

Although these approaches differ in their optimization procedures, all were designed to address the same biological challenge: mitigating the loss of genetic variability associated with recurrent genomic selection while maintaining sustained genetic improvement across cycles.

Evaluation of long-term genetic dynamics

Because breeding populations evolve under the combined influence of selection, recombination, Mendelian sampling, and genetic drift, all simulations were replicated 100 times independently for each combination of population size scenario and mating strategy. Replicated simulations were used to capture stochastic variation among breeding trajectories and to ensure robust inference regarding long-term population responses.

The performance of each strategy was monitored over 40 years of recurrent selection using four complementary indicators describing both genetic improvement and the maintenance of selection potential.

First, population genetic gain was quantified as the mean true breeding value (TBV) of individuals within each breeding cycle, reflecting the realized response to selection independently of environmental effects. Second, maximum genetic gain was measured as the highest TBV observed within each cycle, representing the genetic potential of the best individual available for cultivar development or parental advancement. Third, additive genetic variance was monitored throughout the breeding program to quantify the rate of genetic variability depletion under each mating strategy. Because additive variance determines a population's capacity to respond to future selection, its persistence was considered a key indicator of long-term sustainability. Finally, predictive accuracy was estimated as the Pearson correlation between genomic estimated breeding values (GEBVs) and true breeding values (TBVs), allowing the evaluation of how populational processes and reductions in genetic variation affected the effectiveness of genomic prediction across cycles.

To facilitate comparisons across scenarios, we standardized genetic gain relative to the initial population mean. Specifically, we subtracted the mean TBV of cycle 0 (C0) from all subsequent cycles, expressing genetic progress as accumulated gain relative to the starting population. Similarly, additive genetic variance was expressed as the proportion of variance remaining after the initial cycle, providing a direct measure of diversity retention over time.

Statistical analysis

Statistical inference focused on the accumulated population genetic gain observed after 40 years of recurrent selection. A two-way analysis of variance (ANOVA) evaluated the effects of population size, mating strategy, and their interaction on long-term genetic gain. The relative contribution of each factor was quantified through the partitioning of the total sum of squares.

Because prolonged recurrent selection can generate non-normal distributions due to inbreeding, drift, and stochastic allele fixation, we compared mating strategies within each population-size scenario using the nonparametric Kruskal–Wallis test. For comparisons between structured and random mating within the conventional phenotypic framework, the Mann-Whitney test was applied.

Multiple comparisons were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure at a 5% significance threshold.

Results

Population size is the primary determinant of long-term genetic response

Analysis of variance revealed significant effects of population size, mating strategy, and their interaction on accumulated genetic gain after 40 years of recurrent selection (Table 1). However, population accounted for most of the observed variation, explaining 92.93% of the total sum of squares. In contrast, mating strategy explained only 2.38% of the variation, although it remained highly significant ($P < 0.01$).

Table 1. Analysis of variance (ANOVA) for accumulated genetic gain in cucumber crossing blocks after 40 selection cycles

Source of Variation	Df	Mean Sq	F value	Pr(>F)	Contribution (%)
Methods (M)	7	20.86	177.25	< 2.2e-16	2.33%
Population Size (P)	2	2913.82	24759.28	< 2.2e-16	92.93%
Interaction (P x M)	14	2.92	24.84	< 1.4e-60	0.65%
Residuals	217 8	0.12			4.09%

The significant interaction between population size and mating strategy indicated that mating optimization effectiveness depended strongly on population size. Consequently, the relative benefits of alternative mating strategies varied across population size scenarios, suggesting that population size determines the extent to which optimization procedures can influence long-term breeding outcomes. Therefore, overall, these results identify population size as the principal determinant of long-term genetic improvement, whereas the benefits of mating optimization emerge only within the constraints imposed by population size and the availability of genetic variation.

Breeding bottlenecks accelerate the loss of genetic variability

Under the most restrictive population-size scenario (10 parents and 20 crosses), all genomic selection strategies rapidly depleted additive genetic variance (Figure 2c). Genetic variability declined continuously throughout the recurrent selection process, reaching near-complete exhaustion by the end of the simulation period.

The progressive erosion of additive variance was accompanied by a marked decline in predictive accuracy after approximately 16 selection years of breeding (Figure 2d). As genomic prediction accuracy deteriorated, all genomic selection strategies approached an early response plateau, with population genetic gain stabilizing despite continued selection pressure (Figure 2a).

Although statistically significant differences among genomic mating strategies were detected at year 40 (Kruskal–Wallis, $P < 0.001$; Figure 3a), the magnitude of these differences remained small relative to the overall loss of genetic variability observed across treatments. Standard genomic selection (GS) and genomic selection combined with a balanced crossing design (GS_BLCD) achieved the highest accumulated gains, whereas strategies imposing stricter restrictions on parental relationships showed reduced long-term performance. In contrast, phenotypic selection strategies retained substantially more additive genetic variance throughout the simulation. The TRAD_BLCD strategy preserved more than 40% of the initial additive variance after 40 years (Figure 2c), resulting in greater temporal stability despite lower absolute genetic gain (Figure 2b).

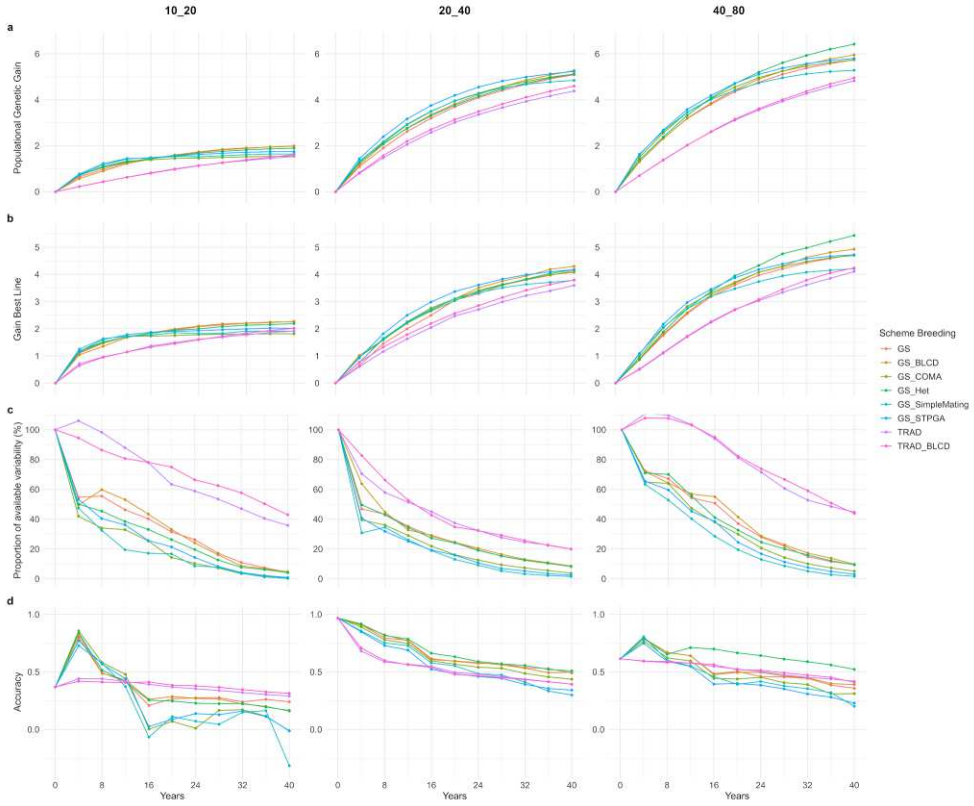


Fig. 2 Temporal dynamics of genetic and predictive parameters over 40 breeding years across three population restriction scenarios (10, 20, and 40 parents). The panels represent: (a) Populational Genetic Gain, (b) Gain of the Best Line, (c) Proportion of available additive variability (%), and (d) Predictive Accuracy. Each data point represents the mean of 100 stochastic simulation replicates for the different mating allocation algorithms evaluated (GS, GS_BLCD, GS_COMA, GS_Het, GS_SimpleMating, GS_STPGA, TRAD, and TRAD_BLCD)

Intermediate population sizes unleash the effectiveness of mating optimization

Increasing the number of parents from 10 to 20 substantially altered long-term population dynamics. Genomic selection strategies achieved about twice the accumulated genetic gain observed under the most restrictive population size scenario (Figure 2a), indicating that a larger breeding population increased the availability of exploitable genetic variability. Despite this improvement, additive genetic variance declined sharply during the first selection and continued to decrease

throughout the breeding program (Figure 2c). By year 40, most genomic selection strategies retained only a small fraction of their original genetic variance, demonstrating that population expansion delayed but did not prevent the progressive depletion of variability.

Prediction accuracy declined more gradually than in the most restrictive scenario, remaining within an intermediate range throughout the simulation (Figure 2d). This increased persistence of prediction accuracy was associated with sustained genetic gain over a longer period and greater differentiation among mating strategies.

Under these conditions, differences among genomic mating strategies became more evident (Kruskal–Wallis, $P < 0.001$; Figure 3c). The heuristic optimization approach (GS_STPGA) and the structured genomic diallel strategy (GS_BLCD) achieved the highest accumulated gains, whereas the exact optimization framework implemented in GS_SimpleMating exhibited comparatively lower long-term performance.

Phenotypic selection strategies continued to retain greater levels of additive genetic variance than genomic selection approaches, preserving approximately 20% of their initial variability by the end of the simulation period (Figure 2c). Nevertheless, their rates of genetic improvement remained lower than those achieved under genomic selection.

Larger populations reveal the benefits of variability-preserving mating strategies

The largest population scenario (40 parents and 80 crosses) generated the greatest divergence among mating strategies and provided the clearest evidence that breeding population size influences the effectiveness of genomic optimization.

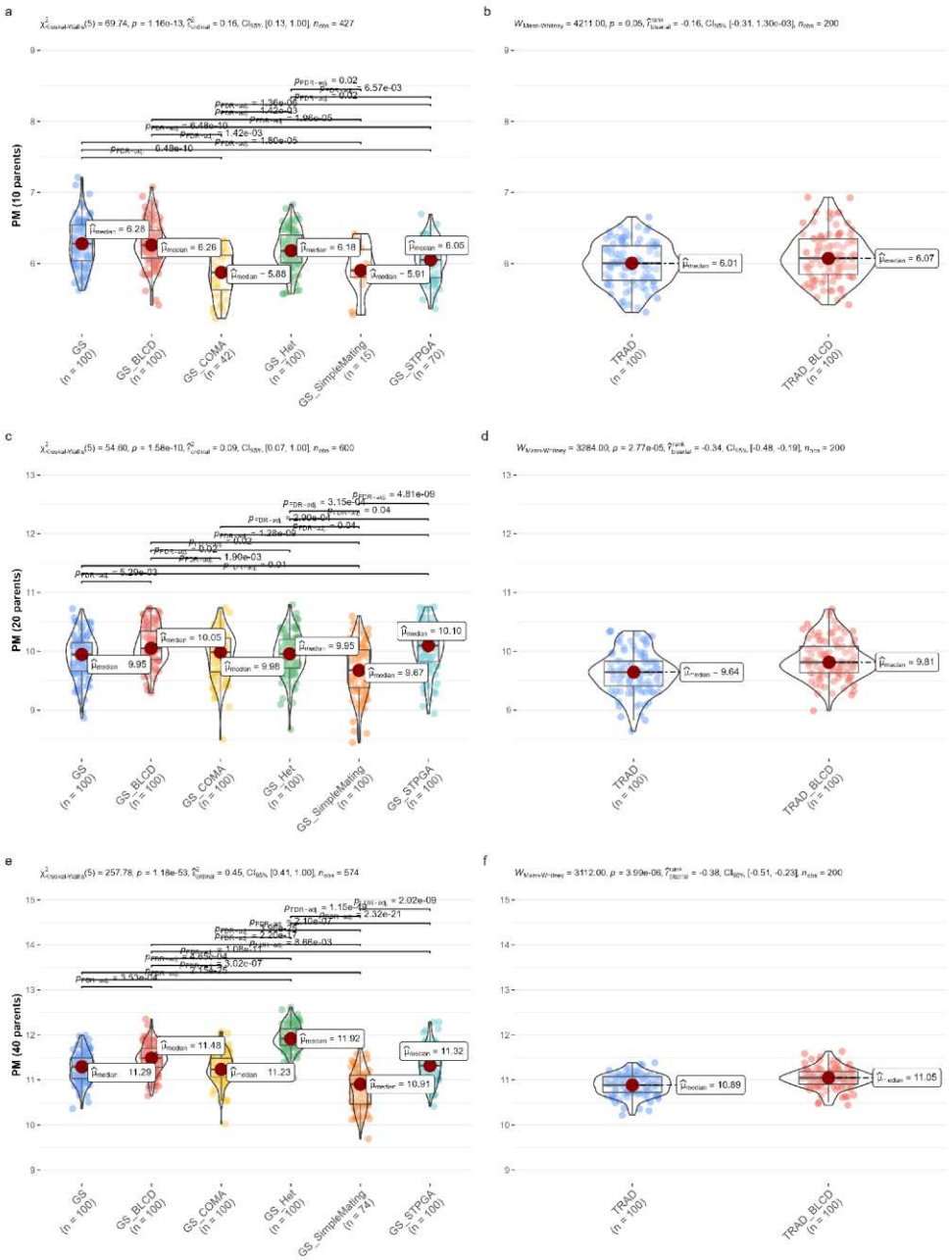
Although all genomic selection strategies showed progressive reductions in additive genetic variance (Figure 2c), the larger initial reservoir of genetic variability enabled substantial differences in maintaining predictive accuracy throughout the breeding program (Figure 2d). These differences translated directly into contrasting long-term genetic-gain trajectories.

Among the evaluated strategies, GS_Het consistently maintained the highest predictive accuracy over time, with values close to 0.50 at the end of the simulation

period. In contrast, GS_SimpleMating showed the steepest decline in predictive performance, approaching 0.20 by cycle 40.

The superior persistence of predictive accuracy under GS_Het was associated with the highest accumulated population genetic gain among all evaluated treatments (Kruskal–Wallis, $P < 0.001$; Figure 3e). This result suggests that preserving heterozygosity helped maintain the genetic variation necessary for continued response to selection across cycles.

Phenotypic selection strategies also benefited from the larger breeding population size. Additive genetic variance remained largely stable during the initial years and persisted at more than 40% of the original variance after 40 years (Figure 2c). Despite substantial retention of variability, phenotypic selection produced lower accumulated gains than genomic selection, placing it in the lowest performance tier for population genetic gain (Figure 3f).



451

452 **Fig. 3** Distribution of the accumulated population genetic gain after 40 selection
453 years across 100 simulation replicates under three breeding population sizes: 10
454 parents (a, b), 20 parents (c, d), and 40 parents (e, f). Violin plots contrast the long-
455 term performance of optimization algorithms under Genomic Selection (left panels:

a, c, e) and standard Phenotypic Selection (right panels: b, d, f). Red dots indicate the mean of the replicates

Structured mating improves long-term performance under phenotypic selection

Across all population size scenarios, structured mating through partial diallel designs consistently outperformed random mating within the phenotypic selection framework. Mann–Whitney tests confirmed significant differences between TRAD and TRAD_BLCD across breeding population sizes (Figures 3b, 3d, and 3f). These results indicate that, even without genomic information, mating designs that promote more balanced parental contribution can improve long-term breeding outcomes. The benefits of structured mating were particularly evident in greater retention of additive genetic variance, reinforcing the importance of controlling the distribution of genetic contributions across selection.

Discussion

Breeding population size governs the long-term sustainability of genomic selection

The simulations demonstrated that breeding population size is the dominant factor determining long-term genetic response under recurrent genomic selection. Although mating strategy significantly influenced breeding outcomes, population size consistently conditioned its effects. The ANOVA results support this interpretation: breeding population size explained more than 90% of the observed variation in accumulated genetic gain. Consequently, the effectiveness of mating optimization should be viewed not as an intrinsic property of a particular algorithm, but as a function of the amount of genetic variation available within the breeding population.

This result is consistent with theoretical expectations from quantitative and population genetics. The response to selection ultimately depends on the availability of additive genetic variance, which itself is strongly influenced by effective population size and the rate at which variability is lost through selection and genetic drift. When populations are small, stochastic changes in allele frequencies become increasingly important, reducing the efficiency with which favorable alleles can be

retained and recombined across cycles (Waples 2022; Johnson et al. 2024). Under such conditions, even highly efficient selection strategies may be constrained by the progressive depletion of genetic variation.

The marked decline in additive genetic variance observed across genomic selection strategies reflects the combined action of directional selection and the Bulmer effect (Fritsche-Neto et al. 2024). Intense selection generates negative linkage disequilibrium among favorable alleles, reducing the additive variance available to future generations before fixation is achieved (Bulmer 1971; Wientjes et al. 2022; Lara et al. 2022; Fritsche-Neto et al. 2024). As recurrent selection proceeds, this reduction in variance is further amplified by the accumulation of relatedness among selected individuals, accelerating the loss of selection potential within the breeding population.

The temporal decline in predictive accuracy observed throughout the simulations appears to stem directly from these processes. As additive genetic variance was depleted and allele frequencies shifted across cycles, the genetic architecture in the training population progressively diverged from that of selection candidates. This divergence weakens marker–QTL associations and reduces the predictive ability of genomic models, a phenomenon widely documented in long-term genomic selection studies (DoVale et al. 2022; Pocrnic et al. 2023).

Taken together, these results reinforce the view that maintaining genetic diversity is not simply a desirable characteristic of breeding programs, but rather a fundamental prerequisite for sustaining long-term response to selection. The benefits of genomic prediction can be realized over extended time horizons only when sufficient additive genetic variance remains to support continued genetic improvement.

Population bottlenecks constrain the benefits of mating optimization

The most restrictive population size scenario revealed a fundamental limitation of genomic mating optimization under recurrent selection. Despite substantial differences in their optimization procedures, all genomic mating strategies converged toward similar long-term outcomes when only ten parents were

available. This convergence suggests that severe bottlenecks can reduce mating algorithms' capacity to influence selection-pipeline trends in breeding populations.

From a population-genetic perspective, this pattern is consistent with the increasing importance of genetic drift as effective population size declines. In small populations, stochastic fluctuations in allele frequencies become a dominant evolutionary force, reducing selection efficiency and increasing the probability that favorable alleles are lost by chance before they can contribute to long-term genetic improvement (Waples 2022; Gargiulo et al. 2024; Yadav et al. 2025). Under such conditions, the scope for optimization becomes inherently limited because the pool of available genetic variation is already severely constrained.

The rapid depletion of additive genetic variance observed in all genomic selection strategies supports this interpretation. As variability declined, opportunities for generating novel favorable allele combinations through recombination became progressively restricted. Consequently, breeding populations approached a state in which alternative mating decisions produced increasingly similar genetic outcomes (Gorjanc and Hickey 2018). This phenomenon may explain why optimization procedures designed to balance gain and variability yielded only modest advantages relative to simpler mating strategies under severe population restriction.

The decline in predictive accuracy further reinforces the central role of population processes in shaping long-term breeding outcomes. As genetic diversity was lost, changes in allele frequencies and linkage disequilibrium progressively altered the relationship between training and selection populations. The resulting reduction in genomic prediction accuracy limited selection effectiveness regardless of the mating strategy employed (Lorenz 2013; Liu et al. 2018; Alemu et al. 2024), contributing to the early response plateau observed across genomic selection treatments.

Importantly, these findings should not be interpreted as evidence that mating optimization becomes ineffective in small populations. Rather, they suggest that biological constraints imposed by reduced effective population size limit the magnitude of the benefits optimization can achieve. When genetic diversity becomes scarce, maintaining variation itself may become a more critical determinant of long-term success than the specific optimization framework used to allocate matings.

This interpretation aligns with previous studies showing that long-term genetic gain depends not only on selection intensity but also on breeding programs' ability to preserve sufficient genetic variability to sustain future response (Li et al. 2022; Wientjes et al. 2022; Pook et al. 2025). Together, these results indicate that population management should be viewed as a prerequisite for effective mating optimization rather than a secondary consideration in breeding program design.

Preserving genetic diversity enhances the persistence of genomic prediction

One of the most consistent patterns across population size scenarios was the close association between retaining additive genetic variance and the persistence of genomic prediction accuracy. Populations that maintained higher levels of genetic variability showed a slower decline in predictive accuracy and sustained genetic gain for more selection years. Conversely, rapid depletion of additive genetic variance was invariably accompanied by an accelerated deterioration of genomic prediction performance.

This relationship reflects a fundamental property of recurrent genomic selection. Genomic prediction models rely on the persistence of linkage disequilibrium between markers and quantitative trait loci (QTL). As recurrent selection proceeds, recombination, directional selection, and stochastic changes in allele frequencies progressively alter the population's genetic architecture. Consequently, marker effects estimated in previous generations become increasingly less representative of the genetic composition of future breeding candidates (DoVale et al. 2022; Pocrnic et al. 2023).

The results obtained here suggest that the loss of predictive accuracy should be attributed primarily to genetic erosion rather than to a limitation of the prediction model itself. Although we recalibrated all genomic selection strategies using updated training populations, the progressive reduction in additive genetic variance limited the information available to retrain the models. As diversity declined, genomic prediction's ability to discriminate among superior individuals decreased, contributing to the plateau in response observed in later selection years.

This interpretation is consistent with recent studies demonstrating that the long-term success of genomic selection depends on maintaining sufficient genetic

variability to support both continued response to selection and stable prediction accuracy (Wientjes et al. 2022; Pocrnic et al. 2023). In this context, additive genetic variance serves a dual role: it provides the raw material for future genetic improvement while simultaneously sustaining the predictive framework on which genomic selection depends.

The contrasting trajectories across population-size scenarios further reinforce this interpretation. Larger populations maintained predictive accuracy for substantially longer periods than populations founded by fewer individuals, indicating that population expansion indirectly improves genomic prediction by slowing the loss of genetic variability. Thus, the persistence of prediction accuracy appears to emerge as an important population-level consequence of diversity management rather than merely a statistical property of genomic models.

Taken together, these findings highlight that strategies aimed at preserving genetic variability may provide benefits beyond maintaining breeding potential. By slowing the depletion of additive genetic variance, such strategies also help preserve the reliability of genomic prediction, thereby enhancing the long-term efficiency of genomic selection programs.

Heterozygosity-based mating strategies are simple and effective

The superior performance of the heterozygosity-oriented mating strategy under the largest population size scenario provides important insight into the mechanisms that sustain long-term response to genomic selection. Unlike strategies that primarily focus on maximizing immediate genetic merit, GS_Het explicitly favored parental combinations expected to preserve heterozygosity in subsequent generations. The benefits of this approach became increasingly evident as breeding population size expanded and larger reservoirs of genetic variation became available for selection.

Notably, the advantage of GS_Het was not associated with greater retention of additive genetic variance alone, but also with its ability to maintain predictive accuracy over an extended period. Under the largest population size scenario, GS_Het consistently showed the highest accuracy among genomic selection strategies, suggesting that preserving heterozygosity helped maintain a more stable

relationship between marker information and the population's underlying genetic architecture.

From a population-genetic perspective, this result is consistent with the expectation that heterozygosity indicates the genetic variability available for future recombination. By reducing the rate at which alleles become fixed, heterozygosity-preserving mating strategies maintain a broader pool of segregating variants that can contribute to future selection response (Niehoff et al. 2024; Yadav et al. 2023). This mechanism may be particularly important under recurrent genomic selection, where strong directional selection continuously drives populations toward increased genetic uniformity.

The benefits associated with heterozygosity preservation became most apparent when population size constraints were relaxed. Under severe bottlenecks, limited available genetic variation restricted any mating strategy's ability to substantially alter long-term outcomes. In contrast, larger populations provided enough variability for heterozygosity-based mating decisions to influence selection trends in the breeding population. These results suggest that the effectiveness of diversity-preserving strategies depends not only on their design but also on the amount of genetic variation available for management.

Our findings align with theoretical and empirical evidence indicating that maintaining diversity is essential for sustaining long-term genetic improvement (Woolliams et al. 2015; Vanavermaete et al. 2020; Wientjes et al. 2022; Tchounke et al. 2023). However, the present study further suggests that heterozygosity itself may serve as a practical target for mating allocation in recurrent breeding programs. By delaying the loss of genetic variation and extending the persistence of genomic prediction accuracy, heterozygosity-oriented strategies may help reconcile the longstanding trade-off between short-term genetic gain and long-term sustainability.

Importantly, the advantage observed for GS_Het should not be interpreted as evidence that heterozygosity is universally the optimal criterion for mate allocation. Rather, the results indicate that strategies explicitly designed to preserve diversity become increasingly valuable as breeding population size increases and opportunities to manage genetic variation expand. Under such conditions,

maintaining heterozygosity appears to provide a biologically effective pathway for sustaining both genomic prediction and long-term response to selection.

Another important consideration is that the breeding framework evaluated in this study assumes a closed population, with no migration or introgression from external germplasm sources. Under this scenario, when a novel parent from another breeding program or germplasm pool is introduced into the crossing block, the optimization algorithm is likely to prioritize crosses involving this parent almost exclusively. In practice, the new parent acts as a tester, maximizing the expected gain from introducing previously untapped genetic variability. However, this behavior can generate a highly unbalanced crossing scheme that may not be operationally desirable in breeding programs. Therefore, exercise caution when applying this method in situations where external germplasm is regularly incorporated, as additional constraints may be required to maintain a balanced and sustainable mating allocation.

Implications for breeding and active germplasm management

Although this study was motivated by challenges associated with recurrent genomic selection in plant breeding, the underlying processes identified here extend beyond agricultural applications. The interaction between selection, genetic drift, and variability preservation represents a fundamental feature of recurrent selection programs and influences the long-term selection response of both managed and closed elite breeding populations.

A central finding of this study is that breeding population size determines how much selection and mate allocation can shape population trajectories. When effective population size declines, stochastic processes increasingly influence selection outcomes, limiting selection's ability to consistently accumulate favorable alleles. Conversely, populations with larger reservoirs of genetic variability retain greater adaptive potential and provide more opportunities for selection to generate sustained responses across cycles.

These principles are directly relevant to germplasm management. Many active germplasm collections and public breeding programs are characterized by small effective population sizes, restricted gene flow, and high relatedness among

breeding individuals. Under such conditions, sustainable germplasm management strategies frequently face the same challenge observed in elite lines: balancing immediate genetic gain against the long-term preservation of genetic diversity. Our results suggest that populational management may represent a prerequisite for maximizing the effectiveness of genomic-informed mating decisions in both contexts (Kristensen et al. 2015).

The findings also contribute to the broader discussion regarding the sustainability of genomic selection. While genomic technologies have substantially increased selection efficiency, they do not eliminate the biological constraints imposed by population genetics. Long-term genetic improvement remains fundamentally dependent on maintaining sufficient genetic variability to support continued response to selection. Consequently, breeding programs that focus exclusively on maximizing short-term gain may inadvertently compromise their future adaptive capacity.

More broadly, the results highlight that population processes and diversity management should be considered integral components of genomic breeding strategies rather than secondary constraints. As genomic tools become increasingly accessible across agricultural, forestry, and elite germplasm enhancement programs, understanding how population factors influence the persistence of genetic variation will be essential for designing populations capable of sustaining long-term breeding success.

Possible limitations of this study

Despite some limitations, this study provides a controlled framework for isolating the effects of breeding population size on the performance of genomic mating strategies, thereby contributing to a mechanistic understanding of how population-genetic processes influence the sustainability of genomic selection. However, we acknowledge that i) our simulations were parameterized using Cucumber and a specific genomic architecture. While the underlying mechanisms are expected to be general, responses may differ across species and breeding systems; ii) Although traits included additive and dominance effects, analyses focused on additive genetic variance and prediction. Epistatic effects, which can

contribute to long-term genetic gain through conversion into additive variance, were not explicitly considered; iii) the study assumed closed breeding populations. Therefore, results may not directly apply to programs that routinely introduce external germplasm, where optimization algorithms may overemphasize newly introduced parents, and iv) other factors common in breeding programs, including genotype-by-environment interactions, migration, and changing selection objectives, were not considered and may influence long-term outcomes.

Conclusions

Breeding size primarily determines long-term genetic response, explaining over 90% of the variation in accumulated gain. Mating optimization strategies also perform differently depending on population size. Under severe population bottlenecks, genetic drift overpowered the theoretical benefits of all mating algorithms, leading to rapid variance depletion and early response plateaus. Conversely, as population size increased, the true potential of mating optimization emerged. Finally, among the evaluated mating-allocation approaches, GS_Het, a naïve method based on expected progeny heterozygosity, is particularly efficient and outperforms the other tested methods. Collectively, these findings indicate that mating optimization success depends not only on algorithmic sophistication but also on the population context in which it is implemented. Effective management of genetic variability should therefore be regarded as a fundamental component of sustainable breeding programs.

References

Ahadi P, Balasundaram B, Borrero JS, Chen C (2024) Development and optimization of expected cross value for mate selection problems. Heredity (Edinb) 133:113–125. https://doi.org/10.1038/s41437-024-00697-y

Akdemir D, Sánchez JI (2016) Efficient breeding by genomic mating. Front Genet 7:. https://doi.org/10.3389/fgene.2016.00210

731 Alemu A, Åstrand J, Montesinos-López OA, et al (2024) Genomic selection in
732 plant breeding: Key factors shaping two decades of progress. Mol Plant 17:552–
733 578

734 Bančič J, Greenspoon P, Gaynor CR, Gorjanc G (2024) Plant breeding simulations
735 with AlphaSimR. Crop Sci 64:714–722. https://doi.org/10.1002/csc2.21312

736 Bulmer MG (1971) The Effect of Selection on Genetic Variability. Am Nat
737 105:201–211. https://doi.org/10.1086/282718

738 Chen GK, Marjoram P, Wall JD (2009) Fast and flexible simulation of DNA
739 sequence data. Genome Res 19:136–142. https://doi.org/10.1101/gr.083634.108

740 Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant
741 Breeding: Methods, Models, and Perspectives. Trends Plant Sci 22:961–975

742 da Silva Viana J, DoVale JC, Fritsche-Neto R (2026) Optimizing progeny size and
743 number of crosses under genomic selection: insights into additive and epistatic
744 contributions to long-term genetic gain. Theoretical and Applied Genetics 139:.
745 https://doi.org/10.1007/s00122-026-05164-2

746 DoVale JC, Carvalho HF, Sabadin F, Fritsche-Neto R (2022) Genotyping marker
747 density and prediction models effects in long-term breeding schemes of cross-
748 pollinated crops. Theoretical and Applied Genetics 135:4523–4539.
749 https://doi.org/10.1007/s00122-022-04236-3

750 Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection
751 with R Package rrBLUP. Plant Genome 4:250–255.
752 https://doi.org/10.3835/plantgenome2011.08.0024

753 Endelman JB (2025) Genomic prediction of heterosis, inbreeding control, and mate
754 allocation in outbred diploid and tetraploid populations. Genetics 229:.
755 https://doi.org/10.1093/genetics/iyae193

756 Fritsche-Neto R, Ali J, de Asis EJ, et al (2024) Improving hybrid rice breeding
757 programs via stochastic simulations: number of parents, number of hybrids, tester

758 update, and genomic prediction of hybrid performance. Theoretical and Applied
759 Genetics 137:. https://doi.org/10.1007/s00122-023-04508-6

760 Gargiulo R, Decroocq V, González-Martínez SC, et al (2024) Estimation of
761 contemporary effective population size in plant populations: Limitations of
762 genomic datasets. Evol Appl 17:. https://doi.org/10.1111/eva.13691

763 Gorjanc G, Hickey JM (2018) AlphaMate: A program for optimizing selection,
764 maintenance of diversity and mate allocation in breeding programs. Bioinformatics
765 34:3408–3411. https://doi.org/10.1093/bioinformatics/bty375

766 Granato ISC, Galli G, de Oliveira Couto EG, et al (2018) snpReady: a tool to assist
767 breeders in genomic analysis. Molecular Breeding 38:.
768 https://doi.org/10.1007/s11032-018-0844-8

769 Guan J, Miao H, Zhang Z, et al (2024) A near-complete cucumber reference
770 genome assembly and Cucumber-DB, a multi-omics database. Mol Plant 17:1178–
771 1182

772 Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting
773 quantitative traits in livestock. Genetics Selection Evolution 2001 33:3 33:209-.
774 https://doi.org/10.1186/1297-9686-33-3-209

775 Hayes BJ, Cogan NOI, Pembleton LW, et al (2013) Prospects for genomic
776 selection in forage plant species. Plant Breeding 132:133–143

777 Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from
778 theory to practice. Brief Funct Genomics 9:166–177.
779 https://doi.org/10.1093/BFGP/ELQ001

780 Johnson JP, Piche L, Worral H, et al (2024) Effective population size in field pea.
781 BMC Genomics 25:. https://doi.org/10.1186/s12864-024-10587-6

782 Kristensen TN, Hoffmann AA, Pertoldi C, Stronen A v. (2015) What can livestock
783 breeders learn from conservation genetics and vice versa? Front Genet 5

784 Lara LA de C, Pocrnic I, Oliveira T de P, et al (2022) Temporal and genomic
785 analysis of additive genetic variance in breeding programmes. Heredity (Edinb)
786 128:21–32. https://doi.org/10.1038/s41437-021-00485-y

787 Li Y, Kaur S, Pembleton LW, et al (2022) Strategies of preserving genetic diversity
788 while maximizing genetic response from implementing genomic selection in pulse
789 breeding programs. Theoretical and Applied Genetics 2022 135:6 135:1813–1828.
790 https://doi.org/10.1007/S00122-022-04071-6

791 Liu X, Wang H, Wang H, et al (2018) Factors affecting genomic selection revealed
792 by empirical evidence in maize. Crop Journal 6:341–352.
793 https://doi.org/10.1016/j.cj.2018.03.005

794 Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and
795 genetic gain of genomic selection in plant breeding: A simulation experiment. G3:
796 Genes, Genomes, Genetics 3:481–491. https://doi.org/10.1534/g3.112.004911

797 Makanjuola BO, Miglior F, Abdalla EA, et al (2020) Effect of genomic selection
798 on rate of inbreeding and coancestry and effective population size of Holstein and
799 Jersey cattle populations. J Dairy Sci 103:5183–5199.
800 https://doi.org/10.3168/jds.2019-18013

801 Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value
802 Using Genome-Wide Dense Marker Maps

803 Niehoff TAM, ten Napel J, Bijma P, et al (2024) Improving selection decisions
804 with mating information by accounting for Mendelian sampling variances looking
805 two generations ahead. Genetics Selection Evolution 56:.
806 https://doi.org/10.1186/s12711-024-00899-2

807 Peixoto MA, Amadeu RR, Bhering LL, et al (2025) SimpleMating: R-package for
808 prediction and optimization of breeding crosses using genomic selection. Plant
809 Genome 18:. https://doi.org/10.1002/tpg2.20533

Pocrnic I, Obšteter J, Gaynor RC, et al (2023) Assessment of long-term trends in genetic mean and variance after the introduction of genomic selection in layers: a simulation study. Front Genet 14:. https://doi.org/10.3389/fgene.2023.1168212

Pook T, Tost ML, Simianer H (2025) Optimization of recurrent rapid cycle breeding in maize for sustained long-term genetic improvement via stochastic simulations. G3: Genes, Genomes, Genetics 15:. https://doi.org/10.1093/g3journal/jkaf100

Sabadin F, DoVale JC, Platten JD, Fritsche-Neto R (2022) Optimizing self-pollinated crop breeding employing genomic selection: From schemes to updating training sets. Front Plant Sci 13:. https://doi.org/10.3389/fpls.2022.935885

Tchounke B, Sanchez L, Bell JM, Cros D (2023) Mate selection: A useful approach to maximize genetic gain and control inbreeding in genomic and conventional oil palm (Elaeis guineensis Jacq.) hybrid breeding. PLoS Comput Biol 19:. https://doi.org/10.1371/journal.pcbi.1010290

Vanavermaete D, Fostier J, Maenhout S, de Baets B (2020) Preservation of genetic variation in a breeding population for long-term genetic gain. G3: Genes, Genomes, Genetics 10:2753–2762. https://doi.org/10.1534/g3.120.401354

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Viana J da S, DoVale JC, Fritsche-Neto R (2026) Using Stochastic Simulations to Shed Light on How to Deploy Speed Breeding and Genomic Selection in Self-Pollinated Recurrent Breeding Programs. Plant Breeding 145:625–636. https://doi.org/10.1111/pbr.70068

Waples RS (2022) What Is Ne, Anyway? Journal of Heredity 113:371–379. https://doi.org/10.1093/jhered/esac023

Wientjes YCJ, Bijma P, Calus MPL, et al (2022) The long-term effects of genomic selection: 1. Response to selection, additive genetic variance, and genetic

837 architecture. Genetics Selection Evolution 54:. https://doi.org/10.1186/s12711-022-
838 00709-7

839 Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE (2015) Genetic
840 contributions and their optimization. Journal of Animal Breeding and Genetics
841 132:89–99

842 Xia M, Dong M (2022) A novel two-archive evolutionary algorithm for
843 constrained multi-objective optimization with small feasible regions. Knowl Based
844 Syst 237:107693. https://doi.org/10.1016/J.KNOSYS.2021.107693

845 Yadav S, Dillon S, McNeil M, et al (2025) Optimising parent selection in plant
846 breeding: comparing metaheuristic algorithms for genotype building. Theoretical
847 and Applied Genetics 138:. https://doi.org/10.1007/s00122-025-05028-1

848 Yadav S, Ross EM, Wei X, et al (2023) Optimising clonal performance in
849 sugarcane: leveraging non-additive effects via mate-allocation strategies. Front
850 Plant Sci 14:. https://doi.org/10.3389/fpls.2023.1260517

851 **Statements and Declarations**

852 ***Funding***

853 This work was supported by the National Council for Scientific and Technological
854 Development (CNPq).

855 ***Competing Interests***

856 The authors have no relevant financial or non-financial interests to disclose.

857 ***Author Contributions***

858 JAOC expanded the initial simulation code to include the different breeding
859 population sizes, modified the optimization algorithms, and wrote the original draft
860 of the manuscript. RFN conceived the original idea, defined the population size
861 scenarios, and performed the computer simulations. RFN and LAPF developed the
862 initial simulation code. RFN and JCD critically reviewed and edited the manuscript,
863 providing significant intellectual considerations. All authors read and approved the
864 final manuscript

865 ***Data Availability***

The R scripts and simulation datasets generated during the current study are available from the corresponding author upon request and will be deposited in a public repository upon acceptance of the manuscript.

Ethics Approval

This study was based on stochastic computer simulations; therefore, it did not involve human subjects, live vertebrates, or higher invertebrates, and ethical approval was not required.